University of South Bohemia in České Budějovice

Faculty of Science

Cell segmentation from wide-field light microscopy images using CNNs

Ph.D. Thesis

Msc. Ali Ghaznavi

Supervisor
Prof. RNDr. Dalibor Štys, CSc.
Institute of Complex Systems
Faculty of Science
University of South Bohemia in České Budějovice

Co-advisor
Ing. et Bc. Renata Rychtáriková, Ph.D.
Institute of Complex Systems
Faculty of Fisheries and Protection of Waters
University of South Bohemia in České Budějovice

České Budějovice
2023

This thesis should be cited as:

Ghaznavi A., 2023: Cell segmentation from wide-field light microscopy images using CNNs. Ph.D. Thesis Series, No. 7. University of South Bohemia in České Budějovice, Faculty of Science, České Budějovice, Czech Republic, 157 pp.

Annotation

Image object segmentation allows localising the region of interest in the image (ROI) and separating the foreground from the background. Cell detection and segmentation are the primary and critical steps in microscopy image analysis. Analysing microscopy images allows us to extract vital information about the cells, including their morphology, size, and life cycle. On the other hand, living cell segmentation is challenging due to the complexity of these datasets. This research focused on developing Artificial Intelligence/Machine Learning methods of single- and multi-class segmentation of living cells. For this study, the Negroid cervical epithelioid carcinoma HeLa line was chosen as the oldest, immortal, and most widely used model cell line. Several time-lapse image series of living HeLa cells were captured using a high-resolved wide-field transmitted/reflected light microscope (custom-made for the Institute of Complex System, Nové Hrady, Czech Republic) to observe micro-objects and cells. Employing a telecentric objective with a high-resolution camera with a large sensor size allows us to achieve a high level of detail and sharper borders in large microscopy images. The collected time-lapse images were calibrated and denoised in the pre-processing step. The data sets collected under the transmission microscope setup were analyzed using a simple U-Net, Attention U-Net, and Residual Attention U-Net to achieve the best single-class semantic segmentation result. The data sets collected under the reflection microscope setup were analyzed using hybrid U-Net methods, including Vgg19-Unet, Inception-Unet, and ResNet34-Unet, to achieve the most precise multi-class segmentation result.

Declaration

I hereby declare that I am the author of this dissertation and that I have used only those sources and literature detailed in the list of references.

České Budějovice, 09.05.2023 Ali Ghaznavi

This thesis originated from Faculty of Science, University of South Bohemia and Institute of Complex System, Faculty of Fisheries and Protection of Waters, University of South Bohemia.

Financial support

This work was supported by the European Regional Development Fund in the frame of the project ImageHeadstart (ATCZ215) in the Interreg V-A Austria–Czech Republic programme and by the project GAJU 017/2016/Z.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof Dalibor Štys for his continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me throughout the time of my related research and writing of this thesis. Thank you so much!
Besides my advisor, I would like to thank my co-supervisor, Dr. Renata Rychtáriková. I appreciate all her help and support during my PhD research and the writing of my thesis and publications. Thank you so much!
I would like to thank my supervisors during my internships, Dr. Mohammadmehdi Saberioon and Dr. Sibylle Itzerott, both from the Helmholtz-Zentrum Potsdam (GFZ). I am thankful for all their guidance, helps and supports during my PhD research and projects. Thank you so much!
I also want to thank my dear colleagues: Dr. Petr Císař, Dr. Jan Urban, Dr. Štěpán Papáček, Dr. Jiří Jablonský, and Mohammad Mehdi Ziaei for all the discussions and collaboration on the projects and for all the fun and nice time we have had in the last five years.
I must express my gratitude to my lab colleagues, Šárka Beranová and Pavlína Tláskalová, for all the help and contribution to my PhD projects and research!
I want to give special thanks to Soňa Vodková, our institute assistant, who helped me to solve all my issues in the Czech Republic and Petra Korcová, who helped me to manage all my work in the study department.

My friends, who provided a much-needed form of escape from my studies and created a nice and enjoyable time: Jan Procházka, Guillaume Dillenseger, Meysam Aryafard, Mahyar Zare, Hassan Nazari, Vladyslav Bozhynov, Ganna Platonova, Oleksandr Movchan, Dinara Bekkozhayeva and all other friends who deserve thanks for making more enjoyable life with perfect and fantastic memory in Budweis.
Last but not least, I would like to thank my family, especially my mother and brothers, for their support throughout my life. I love all of you; thanks a lot! Also, I would like to thank my father, who is no longer with us. I will never forget you and all your memories…

List of papers and author's contribution included in the thesis:

- **Ghaznavi, A.**, Rychtáriková, R., Saberioon, M., and Štys, D.: Cell segmentation from telecentric bright-field transmitted light microscopic images using a Residual Attention U-Net: a case study on HeLa line. *Computers in Biology and Medicine* **147**, 105805, 2022, **IF = 6.69**. DOI: 10.1016/j.compbiomed.2022.105805

  Ali Ghaznavi developed the methods, analysed the data to obtain the results, and wrote the first draft of the manuscript. Percentage of contribution around 75%.

- **Ghaznavi, A.**, Rychtáriková, R., Císař, P., Ziaei, M., and Štys, D.: Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images. *Under Review*.

  Ali Ghaznavi collected the data, developed the methods, analysed the data to obtain the results, and wrote the first draft of the manuscript. Percentage of contribution around 75%.

- **Ghaznavi, A.**, Saberioon, M., Brom, J., and Itzerott, S.: Comparative performance analysis of simple U-Net, Residual Attention U-Net, and VGG16-U-Net for inventory inland water bodies. *Under Review*.

  Ali Ghaznavi collected the data, developed the methods, analysed the data to obtain the results, and wrote the first draft of the manuscript (Internship project). Percentage of contribution around 70%.

- Lonhus, K., Rychtáriková, R., **Ghaznavi, A.**, and Štys, D.: Estimation of rheological parameters for unstained living cells. *The European Physical Journal – Special Topics* **230**, 1105–1112, 2021, **IF = 2.8**. DOI: 10.1140/epjs/s11734-021-00084-2.

  Ali Ghaznavi contributed by data collection phase. Percentage of contribution around 25%.

Co-author agreement

Dalibor Štys, the supervisor of this Ph.D. thesis and co-author of papers 1,2 and 4 fully acknowledges the stated contribution of Ali Ghaznavi to these manuscripts.

České Budějovice, Prof. RNDr. Dalibor Štys, CSc.

Mohammadmehdi Saberioon, the correspondence author of paper 3 fully acknowledges the stated contribution of Ali Ghaznavi to these manuscripts.

Potsdam, Germany, Msc. Mohammadmehdi Saberioon, Ph.D.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	**A**rtificial **I**ntelligence
ANN	**A**rtificial **N**eural **N**etworks
AUC-PR	**A**rea **U**nder **C**urve **P**recision **R**ecall
CISS	**C**ompensatory **I**terative **S**ample **S**election
CNN	**C**onvolutional **N**eural **N**etwork
DCNN	**D**eep **C**onvolutional **N**eural **N**etwork
DIC	**D**ifferential **I**nterference **C**ontrast
DL	**D**eep **L**earning
DSC	**D**ice **S**imilarity **C**oefficient
FCN	**F**ully **C**onvolutional **N**etwork
FRF	**F**ast **R**andom **F**orest
GMM	**G**aussian **M**ixture **M**odel
GT	**G**round **T**ruth
H&E	**H**ematoxylin & **E**osin
HeLa	**He**nrietta **La**cks
HOG	**H**istogram of **O**riented **G**radient
HT	**H**ough **T**ransform
IoU	**I**ntersection **o**ver **U**nion
LFANet	**L**ightweight **F**eature **A**ttention **N**etwork
LoG	**L**aplacian of **G**aussian
LSTM	**L**ong **S**hort **T**erm **M**emory
ML	**M**achine **L**earning
MSCN	**M**ulti **S**cale **C**onvolutional **N**etwork
MSER	**M**aximally **S**table **E**xtremal **R**egions
RBC	**R**ed **B**lood **C**ells
ReLU	**R**ectified **L**inear **U**nit
ROI	**R**egion **O**f **I**nterest
RPE	**R**etinal **P**igment **E**pithelium
RRU-Unet	**R**ecurrent **R**esidual **Unet**
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
SVM	**S**upport **V**ector **M**achine
ssTEM	**s**erial **s**ection **T**ransmission **E**lectron **M**icroscopy
WBC	**W**hite **B**lood **C**ell

CHAPTER 1

Introduction

1.1 OVERVIEW

In this thesis, the artificial intelligence (AI)-based segmentation of living cells over wide-field light microscopy images is proposed and developed. Chapter 1 describes the human HeLa living cells and the structure of the custom-made wide-field microscope with light transmission and reflection setup used for data collection. The last part of Introduction reviews the AI methods and their usage in object detection and segmentation, namely, machine learning (ML) and deep learning (DL) methods in cell segmentation. The knowledge gap between these methods is highlighted. Chapter 2 introduces the newly developed methods. Different variants of DL methods based on convolutional neural network (CNN) were tested to achieve the best precise segmentation result in our datasets. Chapter 3 contains all results in the form of published papers. The last Chapter 4 summarises and concludes the results presented in Chapter 3.

1.2 HeLa cell line

The HeLa cell line is the human epithelial cancer cell line derived from cervical epithelial carcinoma of an African-American woman, Henrietta Lacks, on February 8, 1951 [11]. The cells were propagated by a famous cell biologist George Otto Gey shortly before Lacks died of her cancer in 1951.

HeLa is the first human cell line that can be cultured rapidly. It is used in medical (cancer, AIDS, toxicological, or gene mapping) research as a gold standard. As the HeLa cells originate from aggressive cancer cells, they can proliferate rapidly with a replication rate of up to two times in 24 h [12]. The replication rate and the ubiquity in cell culture laboratories make HeLa an efficient and appropriate living cell line for research, industrial, and medical applications.

1.3 Wide-field microscopy

A wide-field microscope is a type of optical (light) microscope with the simplest optical path and fast acquisition speed. The microscope principle predominantly utilizes visible light originating from a light source (lamp or diode) and illuminating a large field of view of the sample to produce (Fig. 1.2)

1. a dark image with a bright background (in the transmission mode when the light source is located opposite to the microscope objective and light is passing through the specimen) or
2. a bright image with a dark background (in the reflection mode when the light source and the microscope objective are located on the same side and the light refracted or emitted from the specimen is analysed).

The interaction of light with the specimen under leads to a combination of absorptive, diffractive, refractive, or fluorescence contrast in the image. An image is seen through the digital camera or eyepiece. It is possible to modify

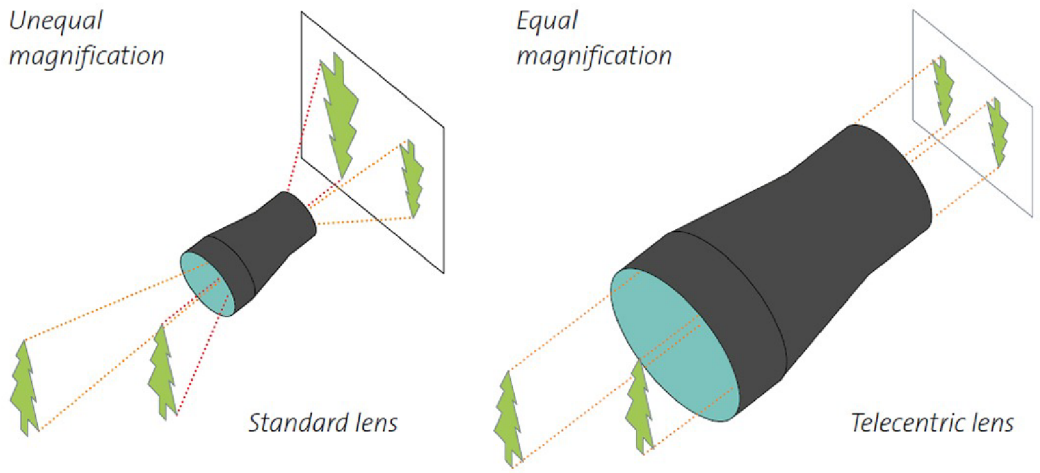

FIGURE 1.1: Telecentric and standard objective mechanism [1].

the microscope objective and digital camera easily to achieve better observation with the naked eye or capturing high-detail digital images, depending on the type of specimen.

The wide-field microscopes, mainly in the transmission mode, are helpful in education and many research fields from biology and medicine up to material engineering. In biology, these microscopes can be used in the simplest up to most advanced research, e.g., [13, 14] to understand intracellular structures in animal and plant cells, to visualise prokaryotic and eukaryotic microorganisms and parasitic organisms.

The specimens must be mostly stained to enable visualisation by negative, Gram, or Papanicolaou staining [15]. These microscopes are appropriate for observing fixed as well as living specimens.

During the measurement, the telecentric objective accepted the light rays parallel to the optical axis. This makes telecentric lenses perfectly suited for measurement applications, where perspective errors and changes in magnification can lead to inconsistent measurements. During time-lapse experiments, the telecentric measurement objective has no angular field or perspective. This objective resolves magnification changes due to object displacement, image distortion, and uncertain object localisation problems. Combining the telecentric lens with a bigger camera chip sensor allows us to obtain sharper images with a high level of detail around the cell borders. Figure 1.1 represents the mechanism of the telecentric and standard objective.

1.4 Cell segmentation methods

Digital image processing means applying computer algorithms to manipulate, enhance, or extract useful information from those images [16]. Detecting and segmenting the objects over digital images into different classes provide

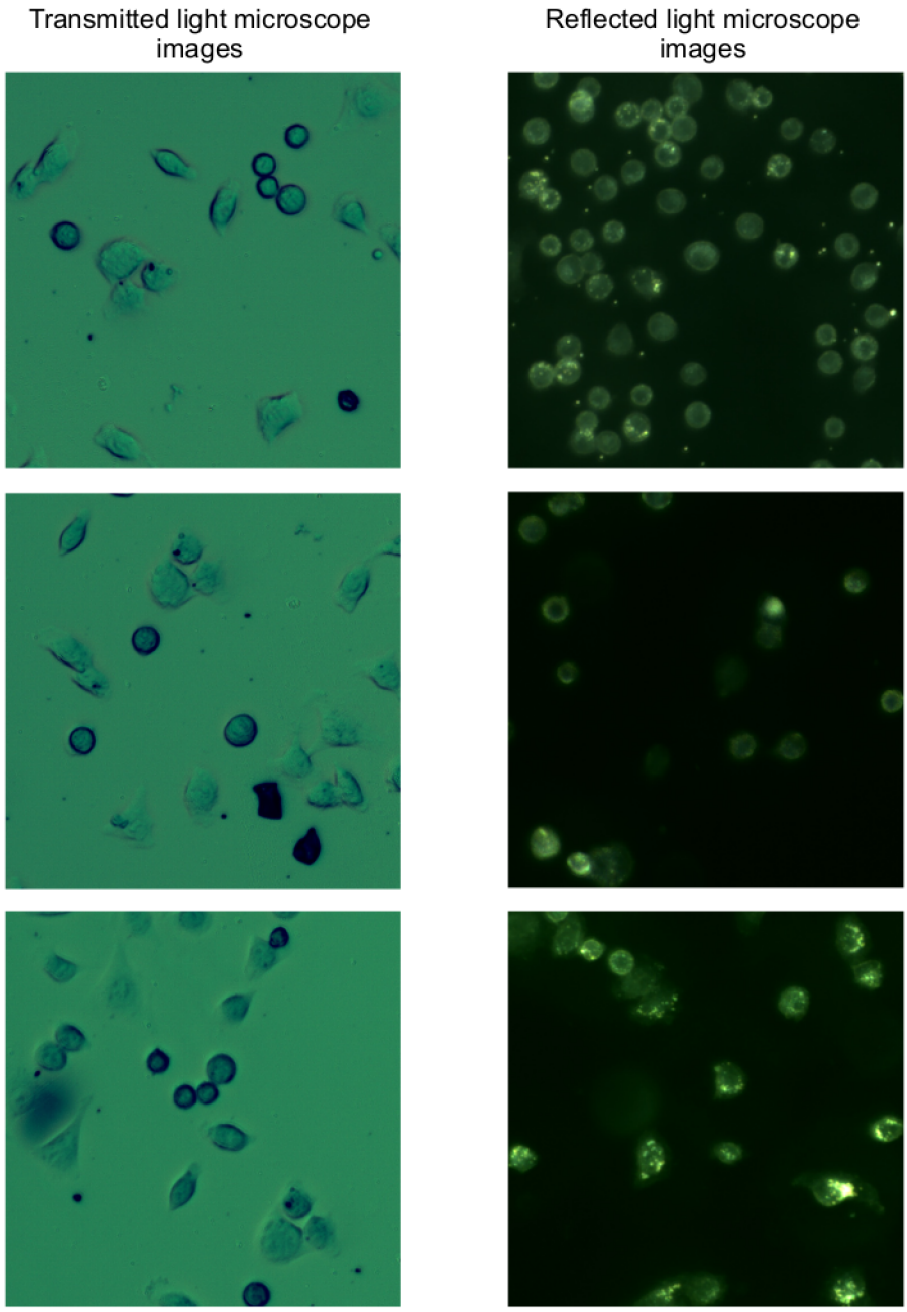

FIGURE 1.2: Examples of unstained living cell data collected by transmitted/reflected microscope with telecentric optics (ICS Nové Hrady). An 8-bit visualisation of the 10-bit primary signal by LIL algorithm [2].

vital information about the target object. The primary purpose of the segmentation is to localise the target objects and their boundaries inside digital images.

Living cell segmentation over time-lapse experiments is essential in analysing microscopy images and provides crucial information about cell behaviour, number, life cycle and dimensions. However, such image analysis is hard due to the changing behaviour and morphology of each cell as well as the whole cell population over time, challenging illumination conditions and optical path inhomogeneities projected in the image.

In general, the segmentation methods can be categorised into three main groups:

1. *traditional*, simplest methods applied in research during the last two decades,
2. more advanced *machine learning* methods dealing with challenges and difficulties, and
3. the most recent, advanced and accurate *deep learning* methods.

To fulfil the task of cell segmentation in image data sets, AI-based detection and segmentation methods, including machine learning and deep learning methods, have been rapidly developed (Fig. 1.3).

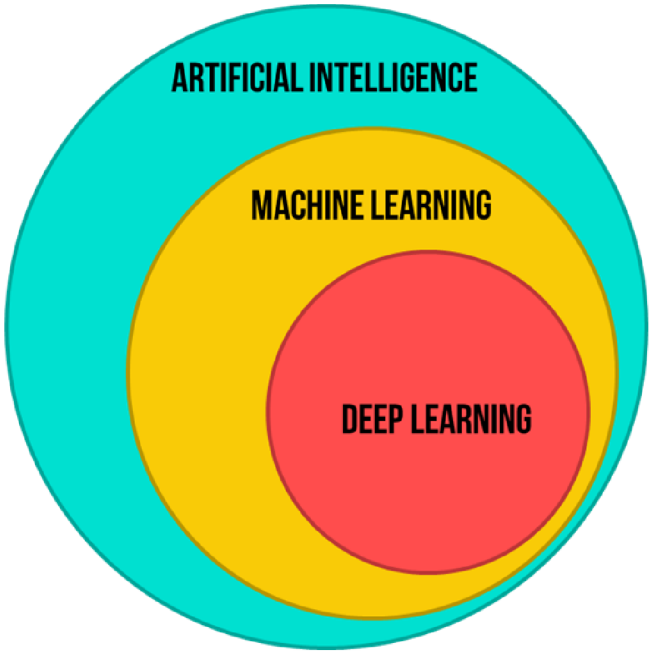

FIGURE 1.3: Visualization of the relationship between AI, ML, and DL methods.

1.4.1 Traditional cell segmentation methods

Over the last two decades, traditional image segmentation methods have been applied in research and often combined to achieve the best possible output. Thus the classification of the relevant literature is not unambiguous.

The number of papers dealing with traditional image processing techniques in light microscopy reaches a few thousand. Here only a few of them is selected.

Intensity thresholding Thresholding techniques are one of the oldest and simplest foreground-background segmentation methods [17]. The thresholding methods convert an image into a binary image by considering a level of threshold (image intensity) that depends on the image condition.

Callau et al. [18] proposed a two-step, fast and simple, intensity-based method to segment the breast cancer epithelial cell over microscopy grayscale images. However, the output is not accurate as more advanced automated methods.

Zhou et al. [19] applied adaptive thresholding with a watershed algorithm for HeLa cell nuclei segmentation from time-lapse fluorescence image series. In the next step, a method of fragment merging that combines two scoring models based on trend and no trend features was applied. In the final step, a Markov model identified phases of cell nuclei.

Morphological erosion-dilation Morphological dilation adds pixels to the boundaries of imaged objects. In contrast, morphological erosion removes pixels on the boundaries of objects. The number of pixels added or removed depends on the size and shape of the structuring element in the image processing.

Using iterative erosion, Schmitt and Hasse [20] separated the cell clumps over bright-field grayscale images into different parts. Firstly, the enhanced erosion operators detected specific cell markers within the eroded scales. Next, an iterative dilation operation expands the markers and regenerates the cell shape, avoiding merging markers. This method is independent of the cell shape and fast but suffers from mis- and under-segmentation of dense cell clumps.

Wang et al. [21] proposed precise single-cell segmentation combining iterative morphological erosion and dilation for fluorescent images of three types of bacteria, budding yeast, and human cells. The method suffered from over-segmentation.

Watershed transform The watershed algorithm is the most well-known morphological method for extracting the foreground from the background. The exact boundary of the target object is extracted using any thresholding or morphological operations as a marker with the watershed method. The image is considered a topographic map where the intensity of each pixel represents its height, and the algorithm finds the lines that run along the tops of ridges. This algorithm efficiently detects and segments touching and overlapping image objects and can be applied in post-processing [22].

Adiga et al. [23] presented a method to detect and segment breast cancer cells over fluorescence images. The authors applied pre-processing steps of image smoothing and thresholding to enhance cell nuclei's edge or boundary features for further watershed-based region-growing segmentation. This

method delivers a more efficient segmentation result than thresholding methods but not ML methods.

Li et al. [24] proposed an automated detection, segmentation and tracking method to analyse the HeLa cell cycle. The authors first binarised the images using adaptive thresholding in the detection and segmentation step. Then, they detected the centre of nuclei using intensity and shape information to achieve seed points. The extracted seed points were used in the watershed algorithm to reach the final segmentation result. The reported results showed 0.995 segmentation accuracy and 0.90 tracking accuracy.

Cheng and Rajapakse [25] introduced a segmentation method over fluorescence images mostly focused on cells and nuclei overlapped in the migration phase. They first applied the active contours method to segment the cells without clear borders and outer distance transform to generate markers. Then, a marker-controlled watershed algorithm with a marking function was applied and achieved 0.95 accuracies of segmentation from the clusters. However, the method suffered from over- and under-segmentation.

Zhou et al. [26] proposed a method to identify and segment the cell phenotypes of the RNAi genome. Firstly, the rough boundary of each cell was extracted. Then, the centre and polygon of each cell were identified. Next, a fuzzy C-means and a marker-controlled watershed extracted each cell. The Voronoi diagrams were applied in the last step to enhance the overlapping cell segmentation. The authors achieved an accuracy of 0.62–0.75 according to the cell phenotype.

Hough transform The Hough transform (HT) is a widespread detection and segmentation method for microscopy images due to the morphological shapes of cells. This method is helpful to find features of any shape, especially straight lines, circles, or curves, in a target image by exploiting the duality between the points on the curve and parameters of this curve [27].

Zhang et al. [28] segmented yeast cells in bright-field in-focus and out-of-focus microscopy images. They first employed the "ilastik" pixel-based classifier to detect the cell boundaries. Cell centre candidates were detected using a Hough transform, and cell edge points were clustered using Integer Linear Programming. Finally, the seeded watershed method was applied to achieve the segmentation result. This method is robust to diverse imaging conditions and out-of-focus images but sensitive to parameter tuning.

Filipczuk et al. [29] developed a method to segment breast cancer cells. The Otsu thresholding was used to detect and extract nuclei masks. The circular HT was applied to determine the nuclei. Afterwards, the circles were filtered out and recognised as nuclei using the support vector machine (SVM) learning method based on the texture features and size of the nuclei masks. This method is robust to high noise levels and object irregularity but sensitive to parameter values to optimise the SVM and the base thresholding step.

Laplacian of Gaussian filter The Laplacian of Gaussian (LoG) filter is a morphological method suitable for identifying small blob objects such as nuclei, or cells [30].

Peng et al. [31] proposed a method to segment the stem cells over microscopy images under different perturbations and conditions. The multiscale blob and curvilinear LoG filter were applied to detect stem cells' structure and skeleton. Then, the extracted cell skeletons were refined using multilevel sets methods to achieve complete and accurate segmentation of the cell buddies. However, this method suffered from high under-detection and under-segmentation.

Li et al. [32] described a segmentation method for cancer cell migration studies from phase contrast images. The original images were filtered with a series of LoG filters of different scales to separate the bright and dark regions of cell bodies. Both detected regions were then concluded, and the cell bodies were segmented by summarising these two regions. This method did not deliver efficient performance for microscopy images with changing illumination. The segmentation accuracy was not comparable with advanced techniques.

Maximally stable extremal regions The maximally stable extremal region (MSER) detector is a method to detect image blobs as areas characterised by bright uniform intensities and their outer boundaries [33].

Zhi et al. [34] proposed the segmentation of nuclei and cells from clumps of overlapping cervical cells. The MSER algorithm was applied to detect and segment the not overlapped nuclei. The output images missed the cytoplasm boundaries on some overlapping cells in poorly contrasted regions.

Arteta et al. [35] described a method to detect and segment H&E stained cells over fluorescence and phase-contrast images. The MSER detector was applied to find a broad number of candidate regions. Then, the SVM classifier classified the extracted regions and scored each region for the detection task. A subset of non-overlapping regions that match the model was selected by maximising the total scores using dynamic programming. The authors annotated a few images with a simple dot to train the model using the SVM classifier. This method achieved a precision of 0.86 and an F1-score of 0.88.

Buggenthin et al. [36] proposed an automatic method for cell detection in bright-field microscopy images. The cell borders were extracted using the active contours method. Then, the MSER algorithm identified and separated nearly all cell bodies. Eventually, a two-step marker-based watershed approach was applied to splitting multiple cells segmented as single foreground objects. The method achieved 0.82 cell detection accuracy (but was insufficient for out-of-focus images) and efficient computation cost.

Thresholding methods [18, 19] are the easiest to separate the foreground and background in the target image. On the other hand, they did not achieve good segmentation results for images with complex intensity distributions, such as microscopy and medical images. *Edge-based methods* [31, 32] deliver efficient segmentation results for objects with sharp and prominent edges but face the problem of multiple, smooth, and vanishing edges of overlapped living cells in microscopy images. *Region-based methods* [25, 26, 35, 36] deal more efficiently with the noisy images and vanishing borders of the target

objects, especially in microscopy images. However, these methods require specifying the seed points and suffer from over- and under-segmentation.

Due to the low performance of the traditional methods on microscopy and medical images, machine learning methods have rapidly grown and expanded in microscopy and medical research region.

1.4.2 Machine Learning methods

Machine learning is a subset of artificial intelligence (AI) in computer science. It allows computers to learn from experience like humans using data and algorithms and gradually improve their accuracy [37]. The ML methods deliver higher performance facing complex and challenging data sets such as microscopy and medical images. Generally, The ML methods could be classified into two main categories:

1. supervised machine learning methods and
2. unsupervised machine learning methods.

Supervised methods

The supervised machine learning techniques use the target data sets and related corrected replies to teach the algorithm and generate the model [38].

Support vector machine One of the well-known supervised and kernel-based learning methods is a support vector machine (SVM). The SVM analyses data to achieve the optimal hyperplane for separation of the high dimensional data with minimum errors in classification and regression tasks [39] (Fig. 1.4).

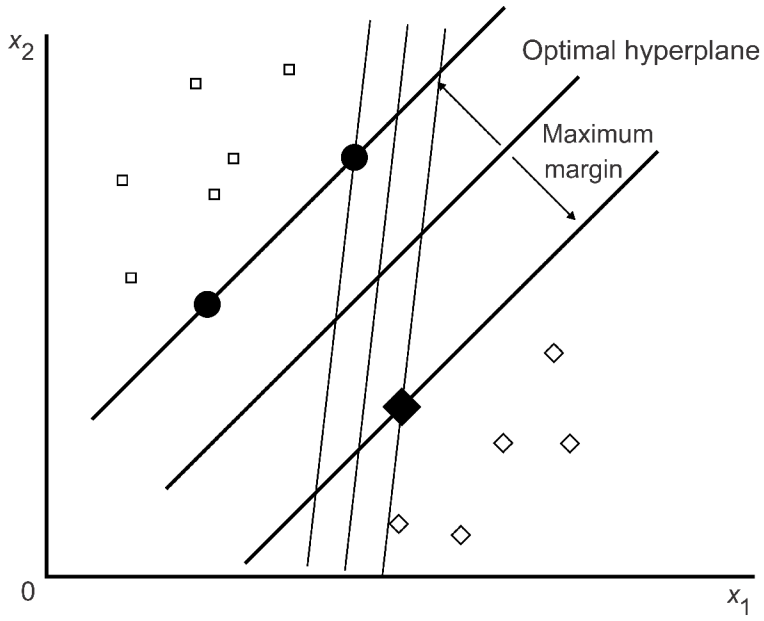

FIGURE 1.4: The structure of SVM classifier [3].

Janssens et al. [40] used a multi-class SVM classifier to separate cells from segmented clumps and connective tissue in H&E stained skeletal muscle cell

images. The clumps were segmented using thresholding of the bright regions. Afterwards, the SVM classified the segments into individual cells, cell clumps, or remnant connective tissues. The method achieved a 0.62 F1 score and suffered from over-segmentation of overlapping cells.

Cheng et al. [41] proposed an SVM classifier for microscopic cellular segmentation. The image pixels were characterised according to their shape, appearance, and context feature descriptors. Then, extracted features pooled to form one vector for a superpixel. Finally, the SVM classifier achieved a segmentation prediction for the input images and delivered a 0.75 pixel accuracy based on the serial section Transmission Electron Microscopy (ssTEM) data set. The method was sensitive to hyper-parameter tuning and showed a low accuracy in detecting and segmenting the vanished mitochondria objects.

Tikkanen et al. [42] applied a histogram of oriented gradient (HOG) feature extractor and SVM classifiers to classify pixels into cell or non-cell regions over bright-field images. This method was sensitive to parameter tuning in the training step to eliminate false positive detections.

Sommer et al. [43] developed a hierarchical supervised classification using an SVM with a Gaussian kernel for automated mitosis detection and segmentation of breast cancer cells over microscopy images. They further optimised cost and gamma hyper-parameters in the classification process by the grid-search parameters. This method suffered from extracting exact localisation properties for small cells and objects and achieved a 0.70 area-under precision-recall curve accuracy.

Lupica et al. [44] applied an SVM-based method to detect and segment cells over bright-field microscope images. The edge boundaries of the target objects were identified using a Canny edge detector. Then, morphological filters filled small gaps and holes to achieve morphological information about the size and shape of the nuclei and cells. The compensatory iterative sample selection algorithm (CISS) trained binary SVM classifiers with radial basis function kernel. The trained model classified the trainset images with a relatively high accuracy rate.

Random forest The random forest (Fig. 1.5) is a supervised classification method that contains a large number of decision trees [45] operating as an ensemble during the training phase. Each tree in the random forest spits out a class prediction. The class with the highest number of votes (trees) is considered the model prediction [46].

Mualla et al. [47] proposed a cell detection and segmentation method based on the random forest over bright-field microscopy images. The representative features were extracted using a scale-invariant feature transform (SIFT). Then, the balanced random forest was applied as a classifier to calculate and classify the descriptive cell key points according to their similarity. Eventually, the key points were clustered with the agglomerative hierarchical algorithm. The weighted mean of the key points was calculated to determine the exact cell region. The SIFT descriptors were invariant to illumination conditions, cell size, and orientation.

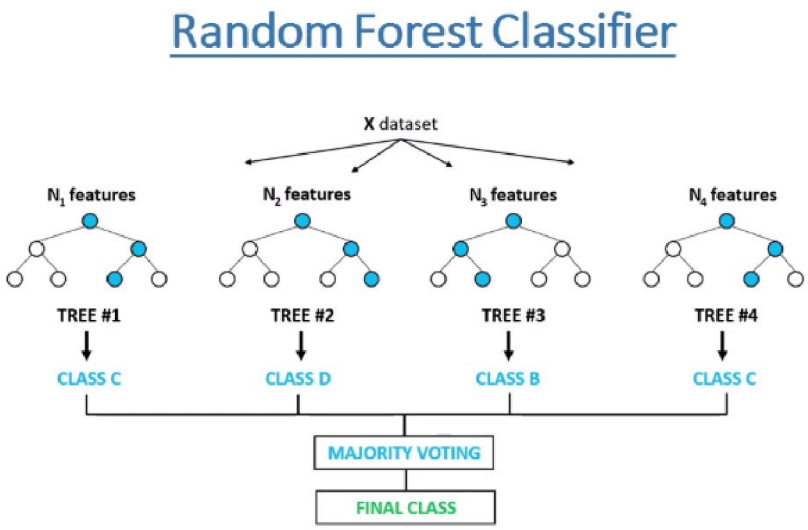

FIGURE 1.5: The structure of Random Forest classifier [4].

Mah et al. [48] described a supervised ML technique to extract the interstitial cells of Cajal networks from 3D confocal microscopy images. The fast random forest classification using trainable Weka segmentation outperformed the decision table and naïve Bayes classification methods in sensitivity, accuracy, and F-measure. However, the process had a higher computational cost due to the structure of the fast random forest method.

Gall et al. [49] constructed random forests-based discriminative class codebooks to cast probabilistic votes within the Hough transform. This approach was called the Hough forests object detection. Yao et al. [50] used the Hough forests to detect and segment the mitotic cells in DIC images. This method has a structure similar to the random forest generating discriminative class-specific parts and achieving the probabilistic votes within the Hough transform framework.

Other supervised methods Liimatainen et al. [51] proposed a supervised method for cell counting in bright-field images using a logistic regression classification with intensity values of 25 focal planes as features. The binary erosion with a large circular structuring element was applied as a post-processing step. However, the method suffered from miss-segmentation and a low recall rate.

Yin et al. [52] proposed pixel-wise segmentation over phase-contrast and DIC images. The segmentation step was completed by classifying individual pixels with an ensemble of Bayesian classifiers. Then, accurate cell boundaries were achieved by assigning each pixel with a posterior probability to the cell or background pixel classes. This method showed a segmentation problem with overlapped cells and might need further processing to split touching cells or nuclei.

Fatakdawala et al. [53] proposed a method to detect and segment H&E breast cancer cells over RGB medical images. They applied the Gaussian

mixture model (GMM) to classify image regions into four pre-defined classes: different cell regions and the background. The method did not need training data sets that are difficult to define owing to variability across images. Due to the absence of prior knowledge of nucleus shape, this method cannot guarantee accurate boundary delineation.

Unsupervised methods

The unsupervised ML methods work without supervision or training. The unsupervised methods are trained with data that is neither labelled, classified, nor scored for training [54].

The best-known unsupervised methods are clustering methods. Clustering expresses grouping data points or objects into clusters according to their similarities. Calculating this similarity is crucial in selecting the appropriate similarity measure and achieving the best clustering result [55]. One such algorithm is K-means (Fig. 1.6) [56].

Xin et al. [57] applied a self-supervised method together with an unsupervised initial segmentation to segment white blood cells. Firstly, the K-means clustering was applied to extract the overall foreground of coarse white blood cells. The second module used the coarse segmentation results as automatic labels to train an SVM classifier. The trained SVM classifier then classified each image pixel and achieved a more accurate segmentation result. However, the unsupervised part of the method generates a rough segmentation result. In the case of complex data sets, the supervised part of the method cannot work efficiently due to fuzzy boundaries.

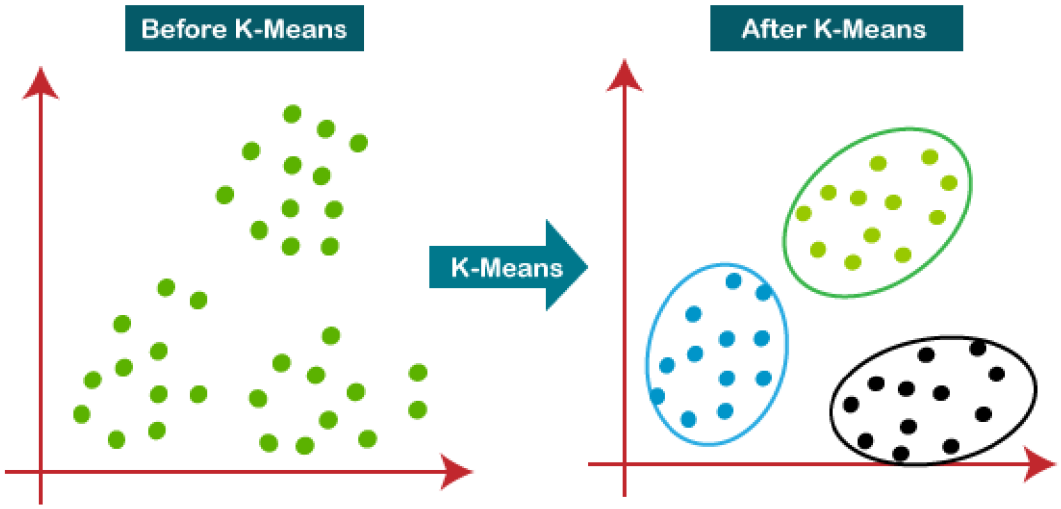

FIGURE 1.6: The scheme of K-means clustering [5].

Antal et al. [58] described unsupervised segmentation over microscope cell images using the Markov Random Field. This method considers an image a series of planes based on Bit Plane Slicing. The planes were used as initial labelling for an ensemble of segmentations. The robust cell segmentation was achieved with pixel-wise voting. However, this method was too sensitive to the confidence threshold and unable to manage huge data sets.

Mualla et al. [47] applied supervised and unsupervised methods together and combined a SIFT to extract key points, a self-labelling, and two clustering methods to segment unstained cells in bright-field micrographs. The computational cost and the achieved accuracy were acceptable, but the technique was sensitive to the feature selection to eliminate the overfitting.

The machine learning methods rapidly expanded due to the low performance of simple image processing methods to detect and segment cells in complex medical and microscopy images. The ML methods have received more attention than traditional methods [40, 42, 47, 49, 51], since they brought more accurate detection and segmentation outputs. Nevertheless, the ML methods are also problematic in aspects as follows:

1. sensitivity of the hyper-parameter tuning to achieve a high-performance trained model [25, 42]
2. over- and under-segmentation in case of complex images of overlapped cells and unstable lighting conditions [40, 43],
3. the high computational cost for model training and the disability to analyse time series and huge data sets [48].

Deep learning (DL) methods have been developed to resolve these problems and achieve higher accuracy and performance.

1.4.3 Deep learning methods

Deep learning is a subset of machine learning methods that allow computers to learn from experience and examples like the structure of the human brain's neural network. Neural networks try to learn and find a correlation pattern between a set of data using a process that the human brain operates on [59]. Deep learning methods are widely used in many application fields, such as speech recognition, visual object recognition, object detection and segmentation and achieved results previously impossible with traditional and ML methods. Many DL methods have been developed for image segmentation tasks, especially for analysing complex microscopy and medical image.

Convolutional neural network Convolutional neural network (CNN) is an artificial neural network (ANN) applied in various computer vision tasks, including radiology and microscopy research. The CNN learns the spatial features during the automatic and adaptive procedure through the back-propagation mechanism. This mechanism is built with convolution layers, including convolution filters, pooling layers for decreasing the extracted feature vector's dimensions, and fully connected layers to merge the extracted features in previous layers for classification [60].

According to the CNN structure, Sermanet et al. [61] developed and proposed a new concept of CNN known as a fully convolutional network (FCN). One of the most popular models for semantic segmentation is a fully convolutional network (FCN) architecture [6]. The FCN methods merge deep semantic information with a shallow appearance to achieve satisfactory segmentation results. The FCN involves the arbitrary size of input images in the

training phase and produces an output of the corresponding size with efficient inference and learning giving a semantic segmentation mask. The most significant difference between CNNs and FCNs is in the last layers. The CNN base methods use fully connected layers for mostly binary and multi-class classification tasks. On the other hand, FCN methods use convolutional layers to generate and predict a segmentation result according to the extracted features at the feature extraction step of the network.

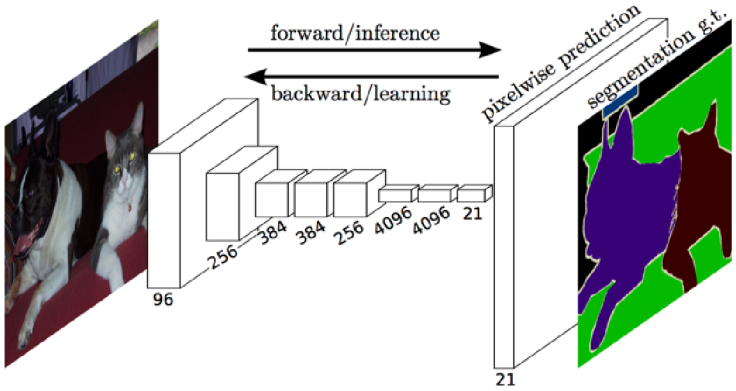

FIGURE 1.7: The FCN architecture [6].

Sadafi et al. [62] proposed a deep learning method to segment red blood cells. The technique used the manual labelled ground truth masks to train the neural network based on FCN structure. The network was trained on small images to decrease the computational cost. The method achieved an accuracy of 0.9 and showed false negative predictions due to the out-of-focus cells.

Lin et al. [63] combined a mask RCNN with a shape-aware loss to achieve HeLa segmentation over DIC and phase-contrast images with a 0.91 IoU accuracy.

Ciresan et al. [64] proposed a DCNN to detect and segment breast cancer cells over histology images. The max-pooling CNN network provided a probability map by classifying each image pixel. The achieved probability map was smoothed with a disk kernel in post-processing. The final centroid was detected with non-maxima suppression.

Song et al. [65] applied the multiscale convolutional network (MSCN) to extract scale-invariant features and segment regions centred at each pixel. Coarse segmentation was completed by an automated graph partitioning method based on the pre-trained features. The Dice metric and standard deviation were significantly improved compared with similar methods.

Liu and Yang [66] combined ML and DL algorithms. The LoG, MSER, and iterative voting learning methods were used to find the candidates for the cell regions. Then, a seven-layer DCNN was used to train the model, assign a score for each extracted candidate, and find the best candidate region. The method achieved 0.90 Dice metric accuracy but is sensitive to parameter optimisation in the supervised ML step to achieve the best detection result using DCNN.

Xie et al. [67] proposed a method to detect and segment the nucleus centroids over bright-field images. The DCNN was applied to learn the voting offset vectors and voting confidence jointly achieved by the Hough voting. Then, the nucleus centroids were localised and detected using heavy clustering and morphological variations. The method reached 0.85 and 0.81 precision and Dice accuracy, respectively. However, the computational cost was high, and the outputs were less satisfying than in other algorithms.

Chang et al. [68] proposed a CNN to detect and segment induced pluripotent human stem cells over bright-field images. The regions of various cell differentiation phases were represented as probability images. The CNN classifier trained the multi-class classification model with multiple types of image patches, including individual types of cells. The five-layer CNN classifier included max-poling and activation function steps and three fully connected layers. The method showed misclassification when the classes were very similar.

Thi et al. [69] introduced a convolutional blur attention (CBA) network containing down- and upsampling procedures for nuclei segmentation in standard challenge datasets [70, 71]. The network assigns deterministic labels to the pixels through the features of input images. The authors achieved a 0.92 F1 score accuracy. The number of trainable parameters lower than in other DCNNs decreased the computational cost.

Jingru et al. [72] developed a CNN for an attentive instance cell detection and segmentation. The algorithm accurately predicts the bounding box and segmentation mask of each cell. The authors first employed a single shot multi-box detector (SSD) [73] to detect neural cells in the input image. Various FCNs that shared the backbone layers with SSD were employed in the segmentation phase. The skip connections in the FCN generate semantics from the deep into the shallow layers. The attention mechanism suppressed noise and highlighted regions with a 0.775–0.779 mean-IoU accuracy.

Wan et al. [74] proposed a DCNN detection-segmentation framework for overlapping cells in digital cytological images. The ROIs identified in the first – cell detection – phase were used as training samples for the subsequent cytoplasm segmentation phase. The TernausNet model was trained and used as a modified FCN as a segmentation neural network. The method could deal with low-quality (poor-contrast, ambiguous foreground/background regions) images.

The U-Net is a convolutional network architecture for fast and precise image segmentation. For the first time, the U-Net was introduced for biomedical image segmentation [7]. The name of this network comes from its shape, which is similar to the letter "U". This network was designed as an extended FCN working with fewer training images but with more precise output.

The U-Net architecture is symmetric (Fig. 1.8). Its left part – the encoder section – extracts the representative features from image regions at different levels of the network convolution operations and hidden layers to reach the network's bottom. The right part – the decoder section – uses the feature representation extracted in the encoder to generate a semantic segmentation

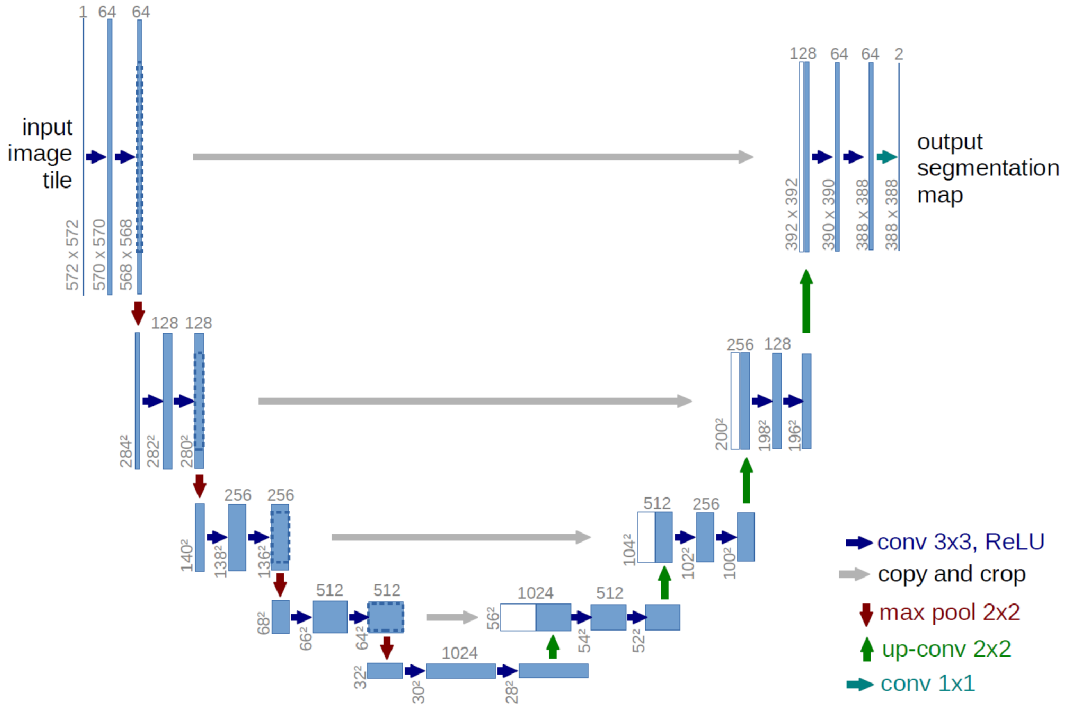

FIGURE 1.8: The default U-Net by [7].

map. The U-Net benefits the concatenation step from the encoder to the decoder merging shallow and deep feature maps and achieving more precise localisation information.

Long et al. [75] modified the U-Net to a light-weighted U-Net (U-Net+) with a customised encoded section to reduce the computational cost for limited computational resources. Due to a weaker feature extraction structure, the method did not deliver higher mean-IoU accuracy in nuclei segmentation over bright-field, dark-field, and fluorescence images.

Bagyaraj et al. [76] proposed two automatic deep learning networks: U-Net-based deep convolution network and U-Net with a dense convolutional network (DenseNet) for detection and segmentation of brain tumour cells. The authors achieved remarkable results with the DenseNet.

Shibuya et al. [77] proposed a Feedback U-Net using the convolutional Long Short-Term Memory (LSTM) network, working on *Drosophila* and mouse cell image data sets. This method showed a low level of accuracy, depending on the segmented class (cytoplasm, cell membrane, mitochondria, and synapses).

Chen et al. [8] proposed a Bridged U-Net (Fig. 1.9) with two different U-Nets to segment prostate cancer over medical images. The method objective was to use the skip connection bridging two U-Net networks as a feature fusion step. The Bridged U-Net was used for feedforward processing from the lower to the upper layer. Using two U-Net architectures leads to more trainable parameters and higher computational costs. The method achieved a 0.881 Dice accuracy which was no significant improvement compared to similar works.

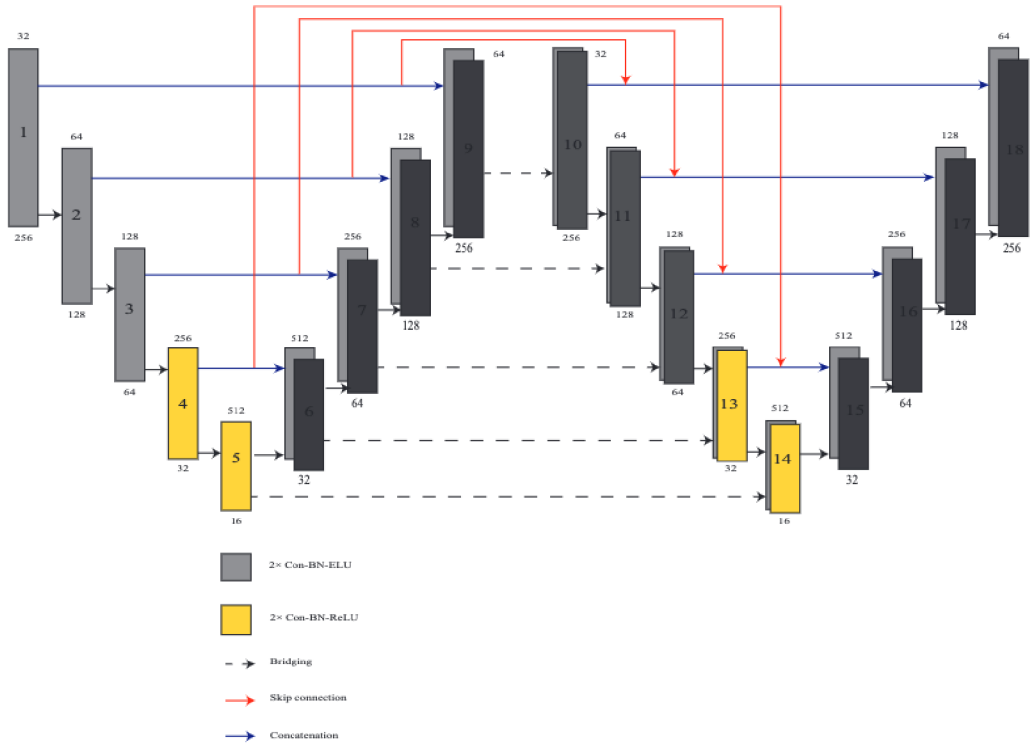

FIGURE 1.9: The bridge U-Net architecture by [8].

Alom et al. [9] proposed a Recurrent Residual CNN (R2U-Net, Fig. 1.10) based on the U-Net for medical image segmentation. The method objective was to improve the performance of the reference U-Net by implementing the recurrent and residual mechanism into each convolutional layer. The method successfully overcame the gradient vanishing problem by continuously updating the gradient values in this very deep neural network architecture. The R2U-Net achieved 0.87, 0.81, and 0.79 F1 scores for DRIVE, STARE, and CHASE medical data sets. Applying recurrent and residual mechanisms together increased the number of trainable parameters and computational costs.

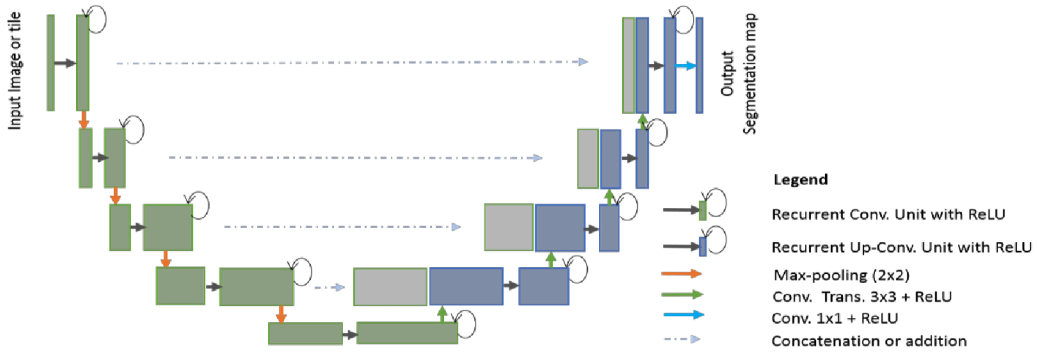

FIGURE 1.10: The R2U-Net architecture by [9].

Pereira et al. [78] proposed a CNN with the 3×3 kernel size to segment the

brain tumour over MRI images. The small kernel made the CNN deeper and mitigated the overfitting by assigning a lower weight value. The data was augmented and normalised in the pre-processing phase. The method performance evaluated on the BRATS 2013 dataset reached 0.78, 0.65, and 0.75 Dice coefficients for the complete, core, and enhancing regions, respectively.

Stawiaski et al. [79] proposed semantic segmentation based on a DenseNet to segment brain tumour regions over medical images. The method used the U-Net as a backbone, utilising dense connections between the layers through dense blocks. The method reached the Dice metric values of 0.79 and 0.85.

Sunny et al. [80] proposed a multi-class cell segmentation in fluorescence images using a hybrid DL method. The authors combined a modified U-Net with the ResNet34 deep encoder network as a feature extraction part to enhance the multi-class segmentation result. Applying the ResNet34 with residual mechanism overcame the gradient vanishing (often occurring in deep neural networks) and gave more representative features to generate the segmentation masks. The ResNet34-Unet achieved a 0.79 IoU accuracy on the SNA-1 SEC data set.

Bakir and Yalim Keles [81] developed a two-step U-Net segmentation over a DIC-C2DH-HeLa data set. The first U-Net was responsible for localising the HeLa cells. The output of the first U-Net served as prior information for the second U-Net to train the model and obtain the exact cell boundaries. The method showed a 0.85 segmentation accuracy. However, the number of trainable parameters and computational costs increased dramatically.

Piotrowski et al. [82] developed a fully automated DL-based multi-class cell state recognition and segmentation over phase-contrast images. The method was based on a U-Net and segmented different classes (colonies, single, differentiated, and dead) of human induced pluripotent stem cells from each other. This method obtained an overall 0.777 IoU metric accuracy, and 0.918 and 0.653 IoU values for the class of colonies and the class of dead cells, respectively, as the best and worst results.

Yu et al. [83] proposed a semi-supervised DL algorithm – MultiHeadGAN – with an encoder and two separate decoders to segment low-contrast retinal pigment epithelium cells over fluorescent microscopy images. The designed Multi-Head structure could train the model with a small scale of annotated data. The method showed segmentation accuracy of 0.873 and 0.801 as the precision and recall metric respectively.

Zhao et al. [84] developed a semantic segmentation for abnormal cells in cervical cytology images. This lightweight feature attention network (LFANet) method combines a feature extraction approach with the attention module to extract abundant representative features from different parts of images of various image resolutions for the training phase. The trained model segmented the nucleus and cytoplasm regions over the cervical images. The method achieved a 0.8760 Jaccard metric value.

Khamene et al. [10] proposed a modified U-Net-based method (Fig. 1.11) to segment membranes over microscopy images to evaluate human epidermal growth factor receptor 2 (HER2) proteins. The method consists of three

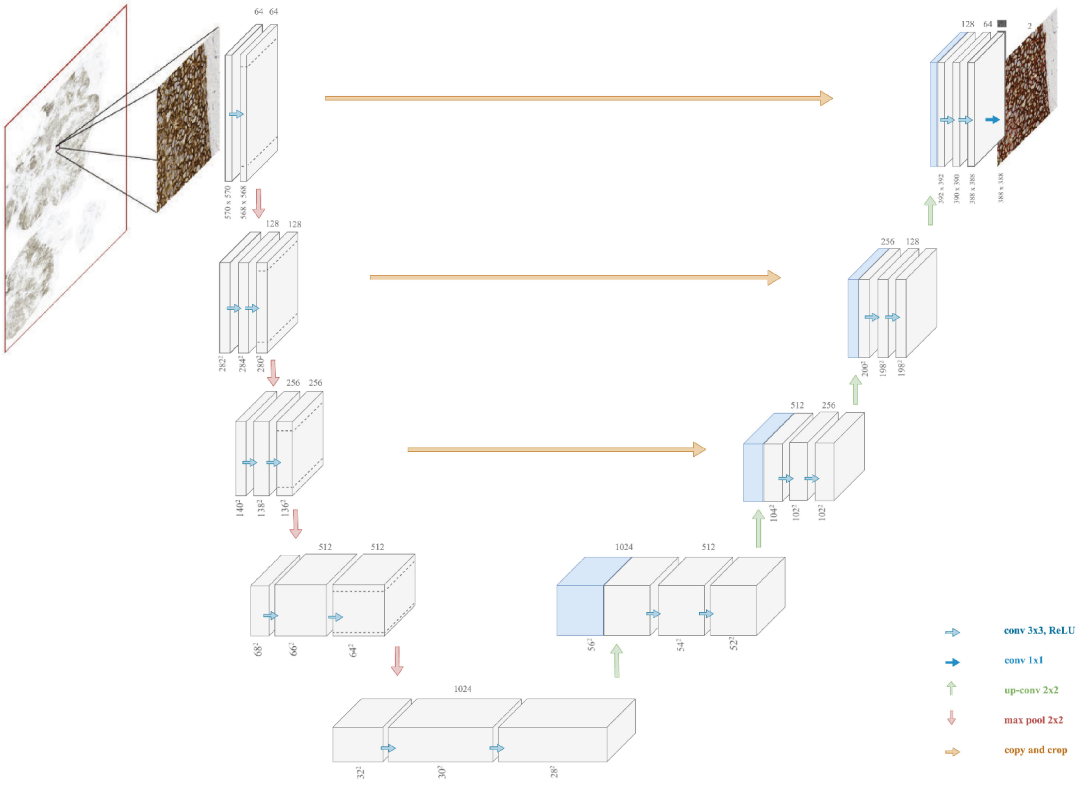

FIGURE 1.11: The modified U-Net-based architecture by [10].

main phases. Firstly, a superpixel SVM feature classifier was used to classify epithelial and stromal regions from the slide image. In the second step, the CNN segmented the membrane regions from the classified epithelial regions. In the last step, the overall score of each slide was obtained by merging and evaluating the divided tiles. The method showed a 0.93 accuracy metric value.

Eschweiler et al. [85] developed a CNN-based multi-class instance cell segmentation method for 3D confocal images. This method integrated the U-Net method with watershed segmentation to benefit both techniques. The proposed CNNs achieved accurate performance in segmentation tasks, even in deeper tissue layers with vanishing fluorosphore responses. The method reached a 0.870 Jaccard index accuracy.

Khan and Mir [86] segmented white blood cells (WBC) from red blood cells and platelets over microscopy images using a U-Net variant with a bigger input image size to obtain the segmentation masks with a 0.687 overall Jaccard metric accuracy. The segmented WBCs regions were then classified into five categories according to the extracted shape and texture features by applying an SVM classifier.

Tran et al. [87] segmented and identified red and white blood cells over microscopy peripheral blood cells images using DL SegNet encoder-decoder architecture with a 0.89 IoU metric value.

1.5 Our research objectives

As described above, traditional image processing [19, 25, 31, 35] and ML methods [40, 42, 43, 48] did not deliver sufficient detection and segmentation outcomes facing difficulties (e.g., background complexity, cell overlapping and vanishing cell borders or large time-lapse and 3D datasets) in biological and medical micrographs. However, compared with ML methods, some CNN methods demand huge computational costs and many manually labelled data to achieve accurate training and high-performance models [6, 88].

The main objective of this PhD thesis is to develop and propose the most accurate and computationally reasonable optimisable AI approaches based on deep learning methods to segment the HeLa cells over transmitted and reflected wide-field microscopy images.

The U-Net-based architecture has been chosen and applied to the transmitted wide-field microscopy images to obtain the single-class semantic segmentation in the first project. The U-Net has been selected since it is a well-known semantic segmentation method with a promising outcome and the ability to work with a reasonable amount of trainable data [7]. Variants of the U-Net architecture – an Attention and a Residual Attention U-Net – have been assembled and examined to find the best architecture for our telecentric bright-field microscopy dataset.

The main objective of the second project was to develop a hybrid deep-learning method for multi-class cell segmentation to classify living cells according to the life cycle phases over unique telecentric wide-field reflected light microscopy images. We replaced the encoder part of the U-Net with VGG19, Inception, and ResNet34 encoder architecture. These CNN variants were examined to enhance the feature extraction step and find the most efficient multi-class segmentation architecture to classify living HeLa cells according to morphological shape in their lifetime.

In this research, a microscope in two light source arrangements (transmission vs reflection) was used to collect time-lapse series of HeLa cells (Fig. 1.2) as raw data with a theoretical pixel size (size of the object projected onto the camera pixel) of 113 nm. This microscope was designed by the Institute of Complex Systems (ICS, Nové Hrady, Czech Republic) and built by Optax (Prague, Czech Republic) and ImageCode (Brloh, Czech Republic) in 2021. The microscope was equipped with the telecentric measurement objective TO4.5/43.4-48-F-WN (Vision & Control GmbH, Shul, Germany) [89] and an AR1820HS 1/2.3-inch 10-bit RGB digital camera (ArduCam Technology CO., Ltd., Kowloon, Hong Kong) with a chip of 4912×3684 pixel resolution. The custom-made software controlled capturing the primary signal with a camera exposure of 2.75 and 998 ms for transmission and reflection, respectively. (Jena, Germany). In the first project of single-class semantic segmentation, we used two light-emitting diodes CL-41 (Optika Microscopes, Ponteranica, Italy) [90] in the transmission arrangement. In the second project on the multi-class living cell segmentation, a light source Schott VisiLED S80-25 LED Brightfield Ringlight [91] in the reflection position was used.

CHAPTER 2

Data collection and methodology

2.1 Overview

Deep learning methods were widely used in many research fields, including medicine and microscopy, for object detection and segmentation. Due to the promising outcome in living cell segmentation, we developed and applied different variants of DL methods to our transmitted and reflected wide-field microscopy image datasets.

We will first describe sample preparation and data collection steps in Section 2.2. Section 2.3 describes the data acquisition and pre-processing steps for both projects. Section 2.4 describes the single-class cell segmentation methods based on transmitted wide-field light microscopy images. The last Section 2.5 describes the hybrid DL methods for multi-class living cell segmentation in detail.

2.2 Sample preparation and data collection

The cell line chosen for both single and multi-class segmentation was HeLa line (Section 1.2). This cell line was provided by (European Collection of Cell Cultures, Cat. No. 93021013) in frozen shape with dry ice. The cells were cultivated to low optical density at 37°C, 5% CO_2, and 90% relative humidity overnight. The nutrient solution consisted of Dulbecco's modified Eagle medium (87.7%) with high glucose (>1 g L^{-1}), fetal bovine serum (10%), antibiotics and antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all purchased from Biowest, Nuaille, France). The HeLa cells were maintained in a Petri dish with a cover glass bottom and lid at room temperature of 37°C.

2.3 Data acquisition and pre-processing

Time-lapse experiments with different time intervals were performed to capture raw data series of living HeLa cells on the glass Petri dishes using the custom-made microscope in a transmitted and reflected setup. The complete description of both transmitted and reflected wide-field light microscope was written in Section 1.3. The obtained raw image series were calibrated by the algorithm proposed in [92] implemented in the microscope control software to minimize the noise and image background inhomogeneities.

After the image calibration, the raw 16-bit time-lapse data were transferred into the quarter-resolved 8-bit colour (RGB) images by the method introduced in [93]. Each pair of green camera filter pixels' intensities were averaged to the green image channel. The red and blue camera filter pixels were assigned to the relevant image channel. Then, images were rescaled to 8 bits after creating the image series intensity histogram and omitting unoccupied intensity levels. This bit reduction ensured the maximal information preservation and mutual comparability of the images through the time-lapse series.

All 8-bit RGB images were denoised by the method proposed in [94] to decrease the background noise to the minimum level and keep the maximum

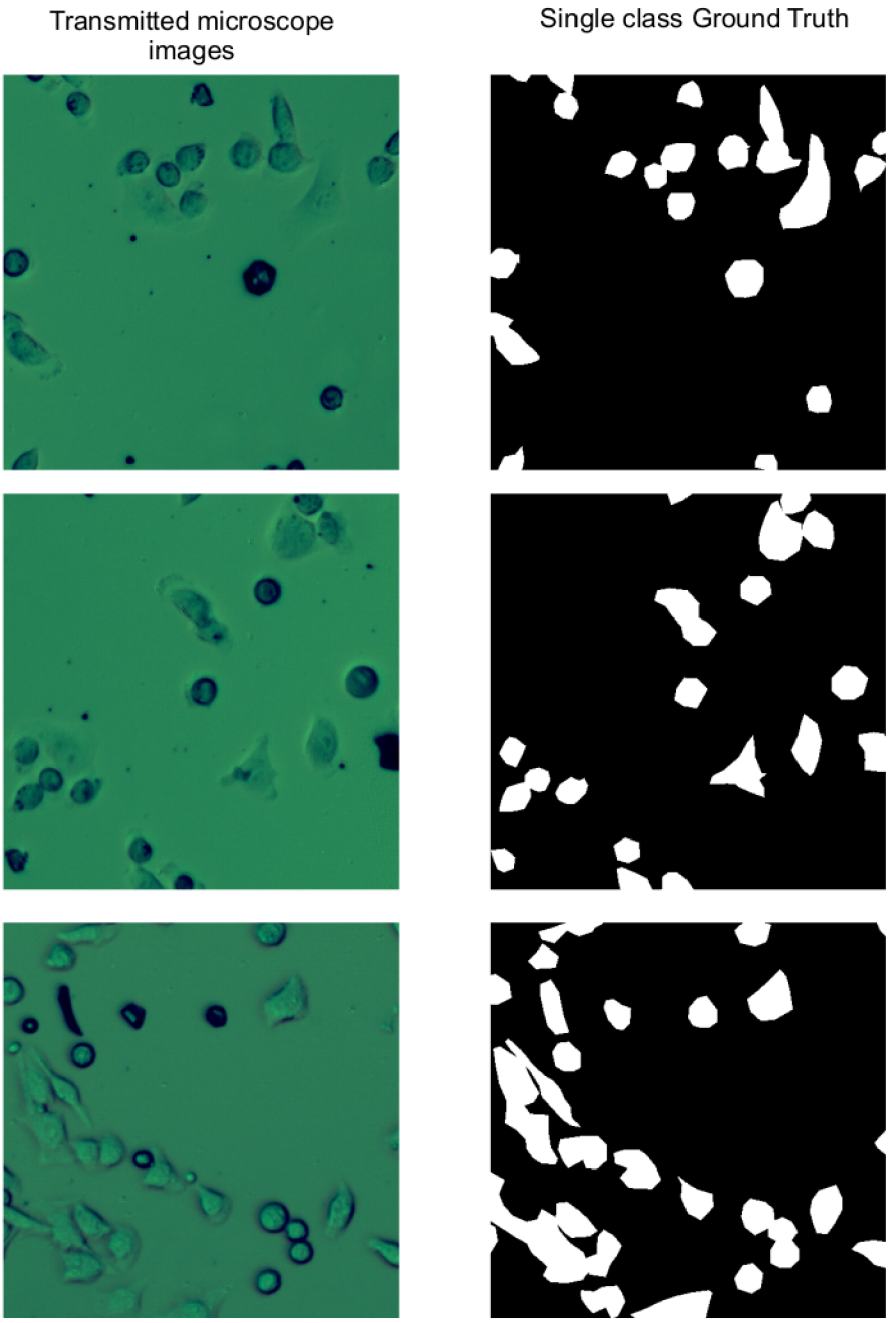

FIGURE 2.1: Examples of collected and manually labelled data in light transmission telecentric microscope.

texture details. Then, the image series were cropped to the 1024 × 1024 pixel size for further analysis.

In the way described above, we obtained 500 light transmission images for training the single-class cell segmentation model and 650 light reflection images for the multi-class cell segmentation model.

In the single-class segmentation project, the images of living cells have been marked manually with human eyes in MATLAB (MathWorks Inc., Natick, Massachusetts, USA) as the Ground-Truth (GT) single-class masks. Figure 2.4 represents a sample of the single-class segmentation data with the corresponding GT.

In the multi-class segmentation project, each cell was manually labelled in the Apper platform and assigned to the cell class according to its morphological shape and life cycle. We distinguished three image region classes:

1. a cell-free background class,
2. a class with cells of larger morphological shapes without cell borders, where the cells are migrating or dividing,
3. a class with roundish cells with sharper borders, where the cells are in their early life-cycle state without division state yet.

Figure 2.5 shows the sample of the multi-class images and ground-truth mask classes.

For both single and multi-class projects, 80% of the labelled images (512 × 512 pixels) were used for model training and remained 20% of the data sets were used for testing and model evaluation. 20% of the training sets were used for the model validation during the training of the neural network architectures.

2.4 Single-class cell segmentation

Three different U-Net architectures were implemented to examine single-class cell segmentation of light transmission microscopy data set to achieve the most accurate semantic segmentation result.

2.4.1 Simple U-Net Model

The U-Net is one of the promising neural network architectures for semantic segmentation [7]. The U-Net was based on the FCN architecture consisting of encoder-decoder layers. This architecture includes various feature channels to merge shallow and deep features. The extracted deep features are utilised for positioning and the shallow features are used for precise segmentation. The architecture of the U-Net chosen for single-class segmentation is represented in Fig. 2.3.

The input layer accepts the RGB colour images as a training set. Each level of the U-Net structure includes two 3×3 convolutions. Batch normalization follows each convolution, and "LeakyReLU" activation functions follow a rectified linear unit. In the encoder part of the network (Fig. 2.3, left part), each "level" consists of a 2 × 2 max pooling operation with the stride of two

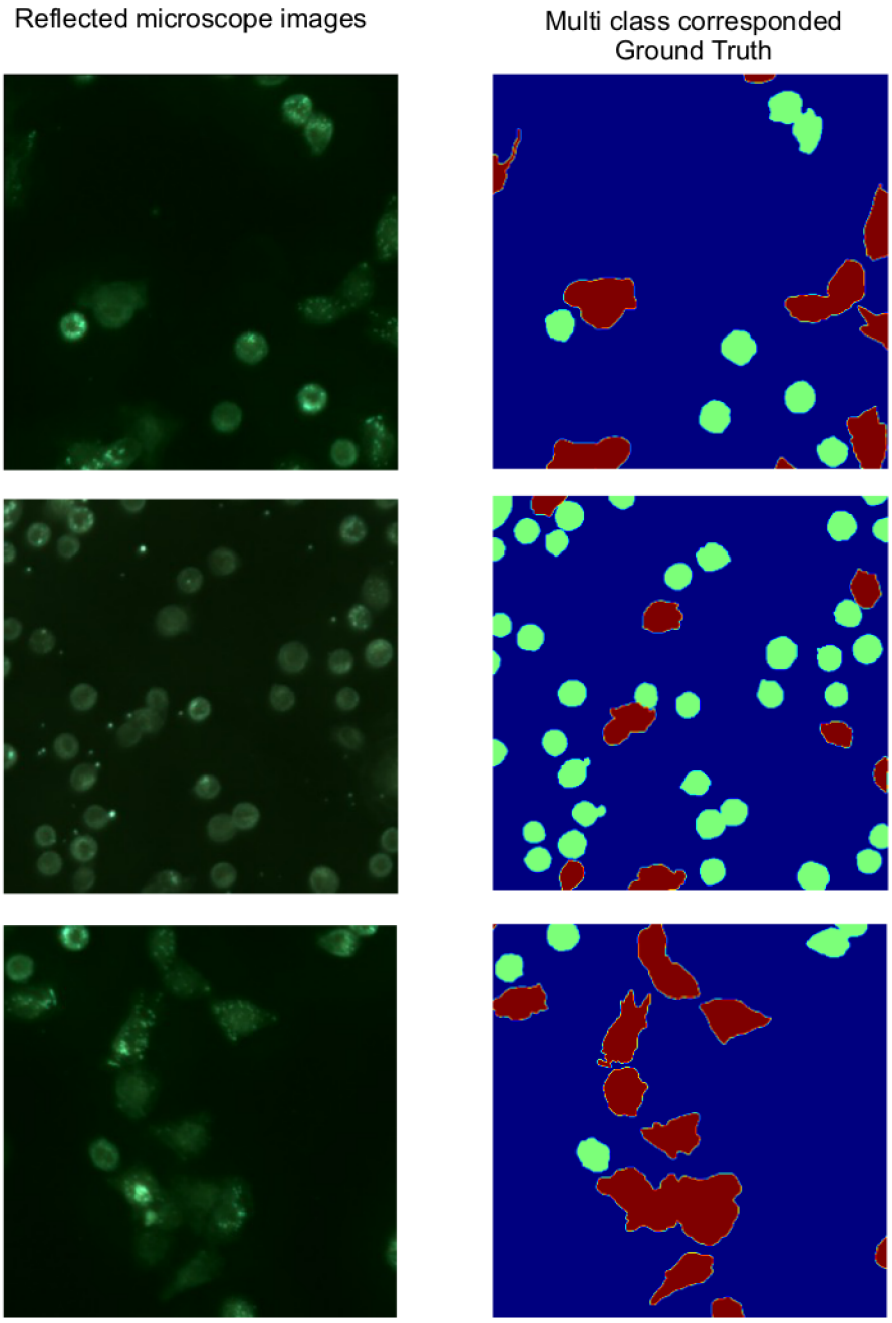

FIGURE 2.2: Examples of light reflection telecentric data and corresponding GT. The green and red class represents the roundish sharp cells and the migrating vanish cells, respectively.

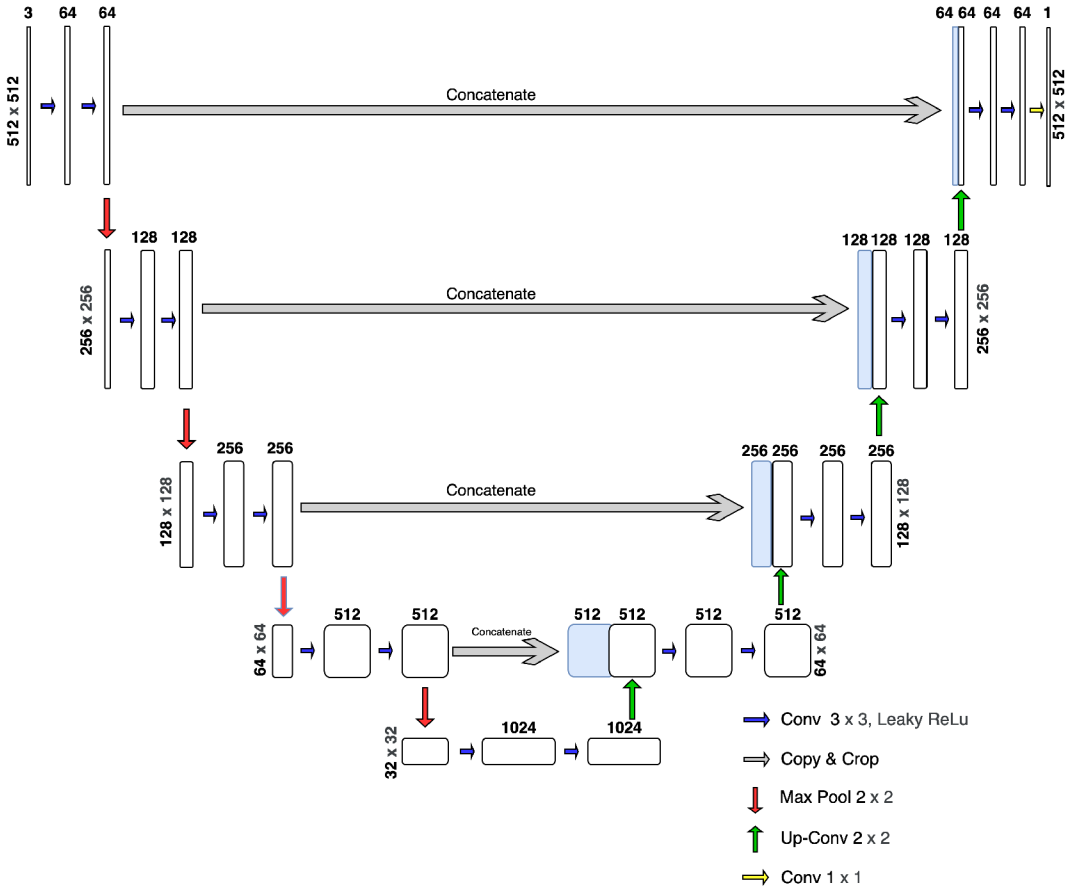

FIGURE 2.3: Architecture of the simple U-Net architecture.

to find the maximal value in the 2×2 area. By completing down-sampling in each level of the encoder part, convolutions will double the number of feature channels.

The height and width of the existing feature map were doubled in each level of the decoder section (Fig. 2.3, right part) from bottom to top. In the next phase, the deep semantic and shallow extracted features were combined and concatenated with the feature maps from the encoder section. After concatenation, the output feature maps have channels twice the size of the input feature maps. The output of the last decoder layer at the top was achieved by 1×1 convolution size and predicts the probability of each pixel. The padding in the convolution process allowed us to obtain the same sizes of input and output layers.

2.4.2 Attention U-Net Model

In the U-Net architecture, the encoder and decoder sections were connected to each other using bridge connections to combine the down-sampling path with the up-sampling path and achieve spatial information. However, this concatenation process brings many irrelevant feature representations from the initial layers. The Attention U-Net architecture [95] showing improvement in medical imaging performance was implemented (Fig. 2.4 A) to avoid transferring irrelevant feature representations and improve segmentation results achieved by a standard U-Net.

The attention gate at the skip connections between the encoder and decoder layers highlights the remarkable features and suppresses activations in the irrelevant regions. In conclusion, the attention gate improves model sensitivity and performance without any complicated computational costs and requirements.

The proposed attention gate (Fig. 2.4B) accept two inputs – x and g. Input x is achieved by the skip connection from the encoder layers. Coming from the early layers, this input contains better spatial information. A gating signal input g comes from the deeper network layer and includes a better feature representation. The attention part weights different parts of the images. This process adds the weights to the pixels based on their relevance in the training step. The relevant parts of the image get large weights than the less relevant parts. The achieved weights are also trained in the training process and make the trained model more attentive to the relevant regions.

2.4.3 Residual attention U-Net Model

The residual mechanism was initially implemented into the U-Net architecture for nuclei segmentation [9]). The architecture was named the Residual U-Net. The simple U-Net architecture was built of repetitive convolutional blocks at each level (Fig. 2.5B). On the other hand, very deep convolutional networks suffer from vanishing gradients at deeper levels. The residual step was developed to continuously and incrementally update the weights in each

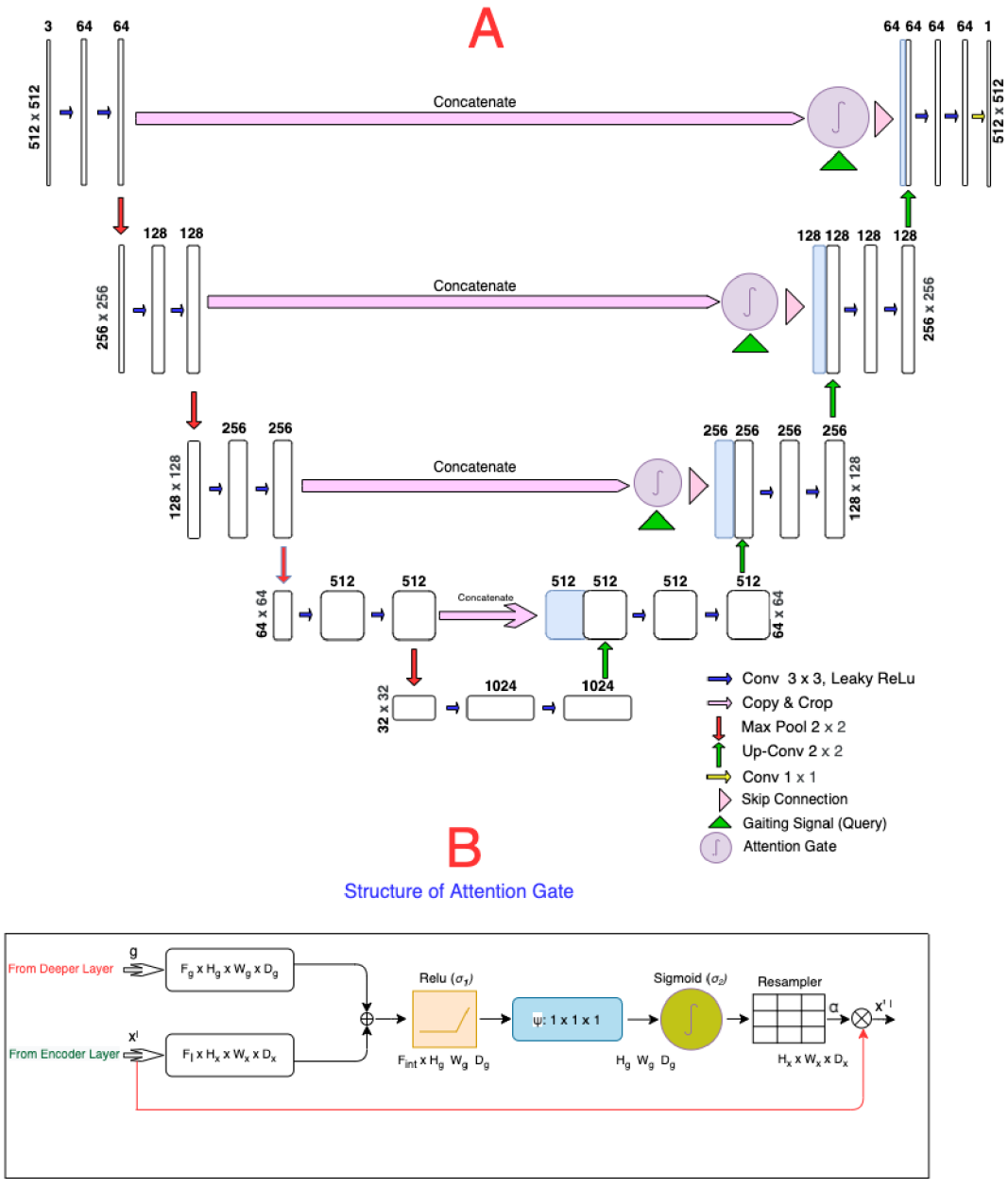

FIGURE 2.4: A) The Attention U-Net architecture, B) the attentive module mechanism. The size of each feature map is $H \times W \times D$, where H, W, and D indicate height, width, and number of channels, respectively.

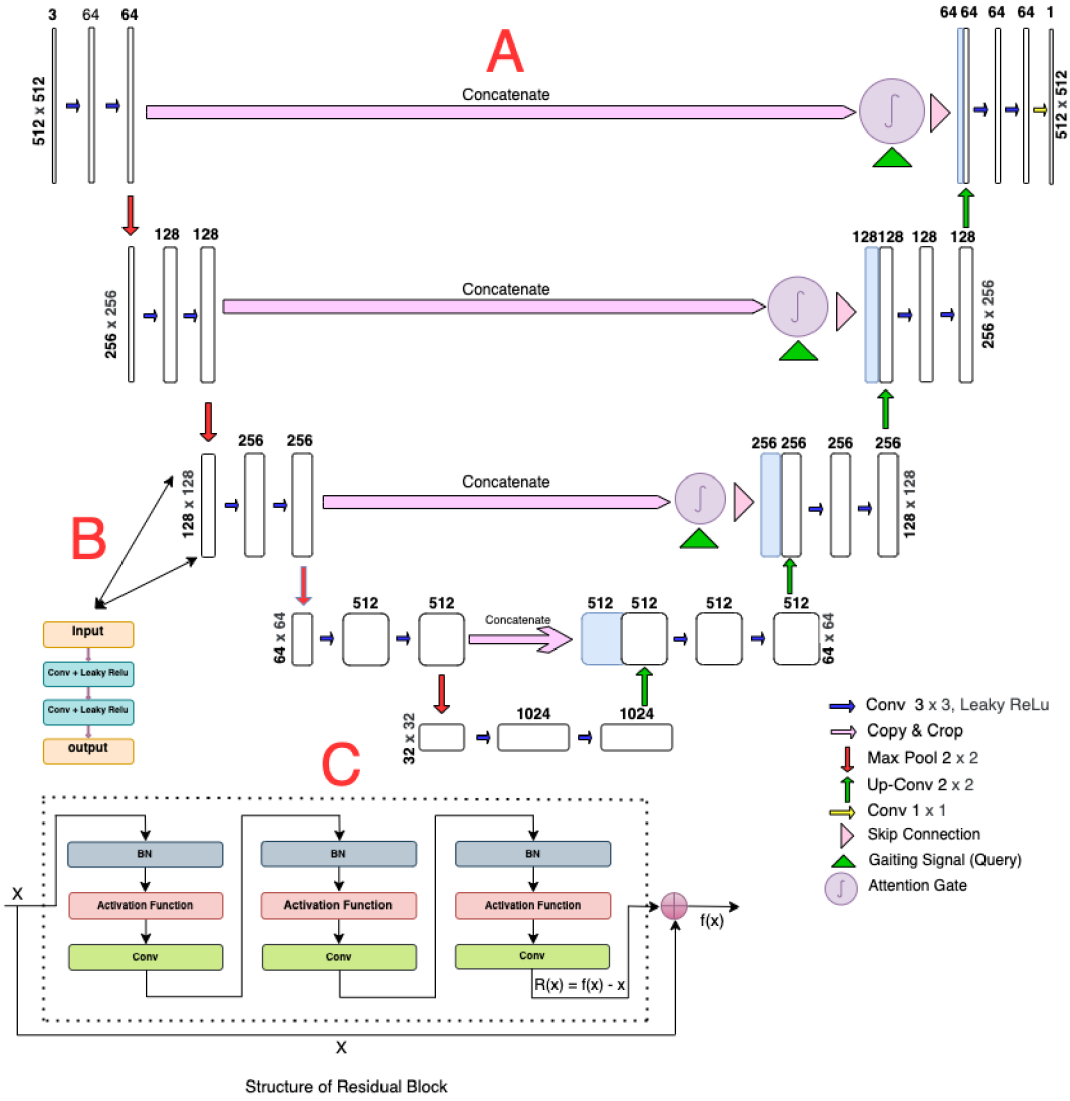

FIGURE 2.5: A) The Residual Attention U-Net architecture. B) A U-Net layer structure. C) The sample of residual block progress. *BN* refers to Batch Normalization.

convolutional block (Fig. 1.6C) to improve the network performance and resolve the vanishing gradient problems.

The mechanism of neural networks is a continuous process in which each convolutional block feeds the next block. A problem in deep convolutional neural networks (DCNN) when stacking convolutional layers is that the generalisation ability of the trained model can be affected by the deeper network's structure. The skip connections–the residual blocks–resolve this problem and improve the network performance, with each layer feeding the next layer and layers about two or three steps apart (Fig. 1.6C). The Residual and Attention U-Net architectures were connected to model our data sets more effectively and further improve segmentation results.

The computational results combined with the Binary Focal Loss function become the energy function of the proposed U-Net-based methods.

After obtaining the most accurate semantic segmentation result in the Residual Attention U-Net, the morphological reconstruction by the watershed algorithm [96] was applied to achieve instance segmentation of each cell. The watershed segmentation further helped us solve the over- and under-segmented regions and specify each separated cell by, e.g., cell diameters, solidity, or mean intensity.

2.5 Multi-class cell segmentation

The simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet32-U-Net architectures were developed and implemented to achieve the most accurate multi-class semantic segmentation result in reflected wide-field light microscopy image series.

2.5.1 Simple U-Net Model

The U-Net [7] is a well-known deep neural network architecture for semantic segmentation based on encode-decoder layers. In this research, a simple – five-"level" – U-Net neural network architecture was implemented as the first method for multi-class segmentation purposes. The architecture of this U-Net (Fig. 2.6) is similar to the simple U-Net proposed in Section 2.4.1. The main difference relies on the last – output – decoder layer.

The top output decoder layer with a 1×1 convolution size predicts the probability of each pixel that the pixel belongs to one of three classes using the "softmax" activation function. Padding in the convolution process allowed us to achieve the same sizes of the input and output layers. Each pixel was assigned to one certain class according to the highest probability values achieved among different classes using the "argmax" operation in the final step.

2.5.2 The VGG19-U-Net

The U-Net is a famous architecture for semantic segmentation tasks. However, the complexity of the U-Net in terms of the number of trainable parameters and weaker feature extraction structures in multi-class segmentation over complex microscopy images affect the trained model's performance. The VGG-Net architecture replaced the U-Net encoder path. In this way, we combined two powerful architectures and improved the categorical segmentation of our unique microscopy data set. The VGG-Net was introduced by Simonian and Zisserman from Oxford's Visual Geometry Group (VGG) in 2015 [97].

The VGG is a popular image recognition architecture, designed to reduce the number of parameters in the convolutional layers and improve training time. The VGG-19 comprises a network with a deeper topology and smaller convolution kernels to simulate a perceptual field of view. Figure 2.7 represents the VGG19-U-Net proposed in this study. The left side of the network (Fig. 2.7A) shows the architecture of the VGG-19 encoder section with 16

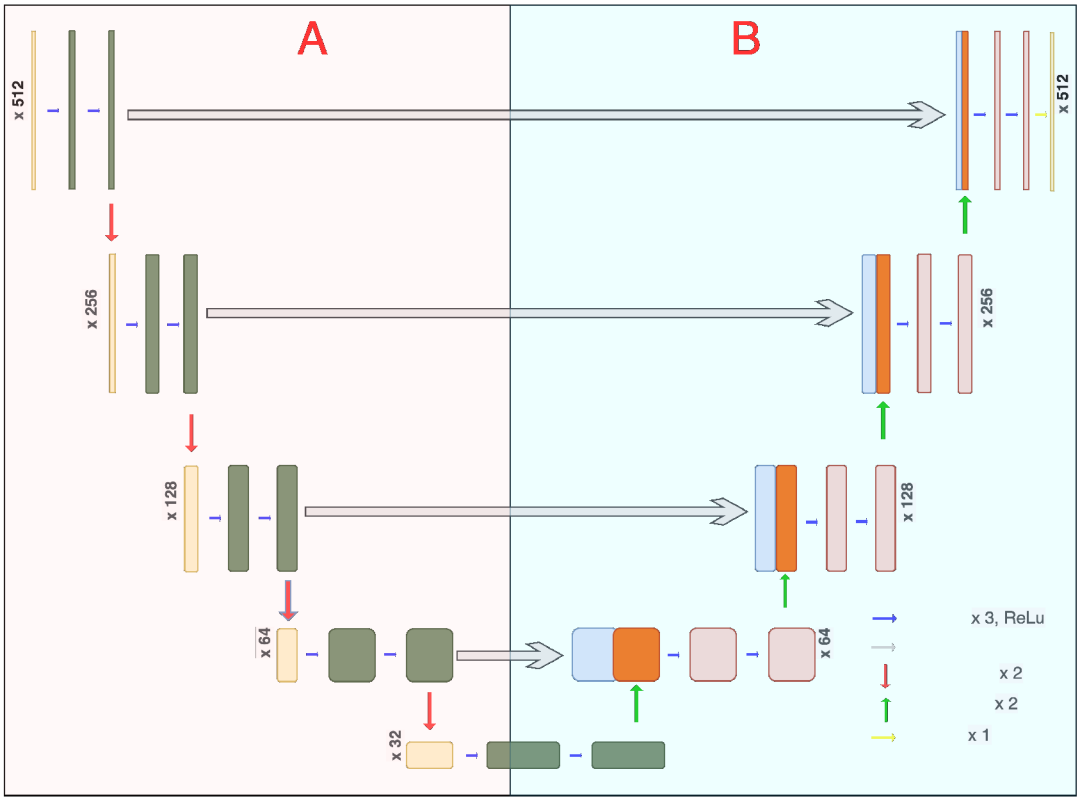

FIGURE 2.6: The simple U-Net model architecture. A) The encoder section. B) The decoder section.

convolution layers, three fully connected layers, and 5 MaxPool layers in five blocks.

The right side of the network (Fig. 2.7B) represents the decoder part with five blocks. The concatenation step between each VGG-19 encoder layer and U-Net decoder layer (Fig. 2.7) combines the feature maps from the encoder part with the high-resolution deep semantic and shallow features from the decoder part. The last decoder layer has a convolution size of 1×1 and predicts the probability values for each pixel and each of the three classes using the "softmax" activation function.

2.5.3 The Inception-U-Net

Analysing microscopy images with fixed kernel size in all convolution layers can make extracting the feature descriptors of different sizes difficult. The bigger kernel can extract a global feature representation over a large image area, and the smaller kernel is suitable for detecting area-specific features. Google's inception deep learning method [98], known as the Inception architecture, was selected to build a hybrid Inception-U-Net architecture (Fig. 2.8) further to improve multi-class segmentation in our data sets.

The inception modules were developed to reduce computational costs by integrating different sizes of convolutions. The inception module applies kernels of various sizes within the same architecture layer and becomes wider (instead of deeper) with the layers (Fig. 1.6A).

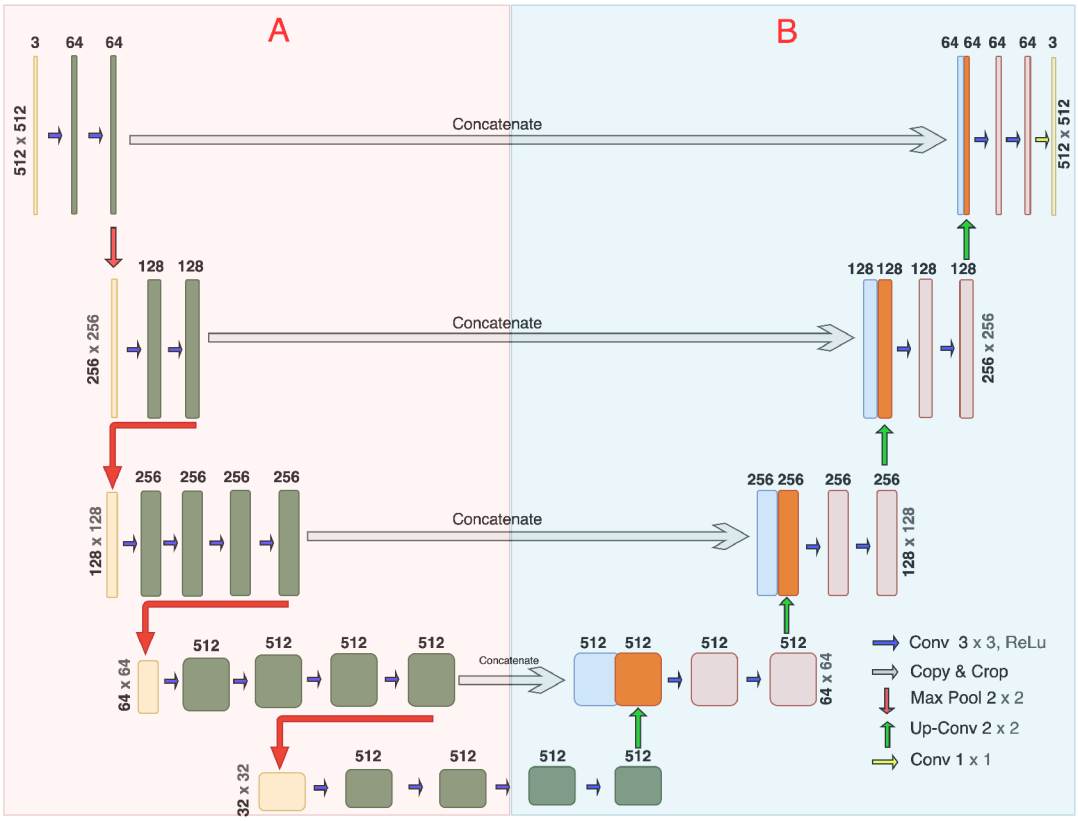

FIGURE 2.7: The hybrid VGG19-U-Net architecture. A) The VGG-19 encoder part. B) The U-Net decoder part

The convolution layers were replaced with an inception module (Fig. 1.6B) in all five levels of the encoder and decoder sections of the original U-Net structure. Each inception module is built of multiple sets of 3×3 and 1×1 convolutions, 3×3 max-pooling, and cascaded 3×3 convolutions.

The last layer in the decoder section, a 1×1 convolution layer, and the "softmax" activation function generate three segmentation classes of the feature maps for each pixel of the given input image. Each pixel is assigned to the class according to the highest probability value among the classes.

2.5.4 The ResNet34-U-Net

The Residual Convolutional Neural Network (ResNet) [99] replaced the feature extraction part of the standard U-net architecture to improve multi-class segmentation further. Deeper neural networks are more effective for complex classification and segmentation tasks. On the other hand, the vanishing gradient problem appears in very deep CNNs during the training process. Also, employing a high number of CNN layers makes the training process slower, and the obtained value of the back-propagation derivative becomes insignificant in training. As a result, the model's accuracy is not improved, and the generalisation ability of the trained model is not satisfactory. To overcome this problem, skip connections are employed in the CNN to bypass one

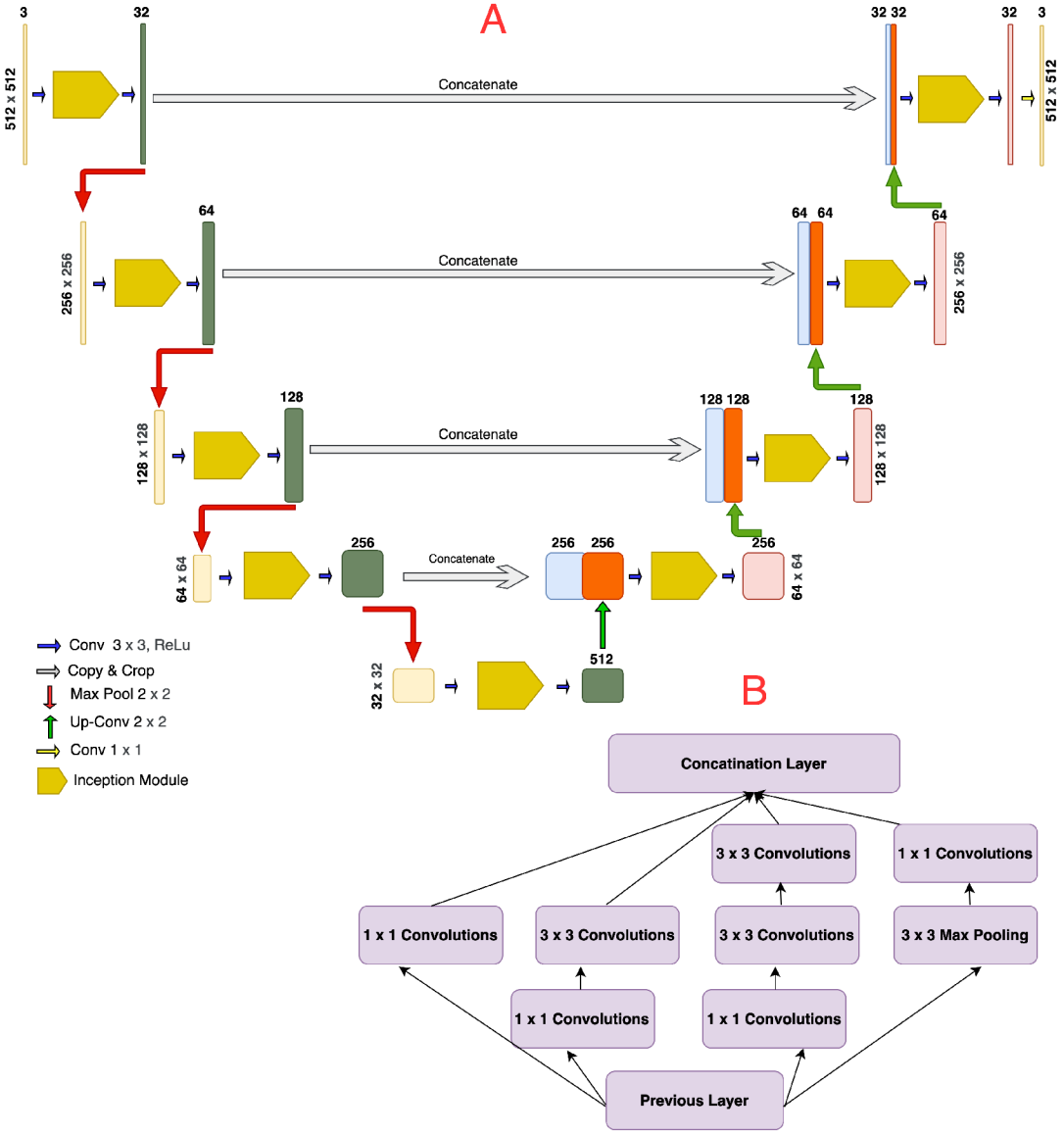

FIGURE 2.8: A) The Inception-U-Net architecture. B) The internal architecture of one inception module.

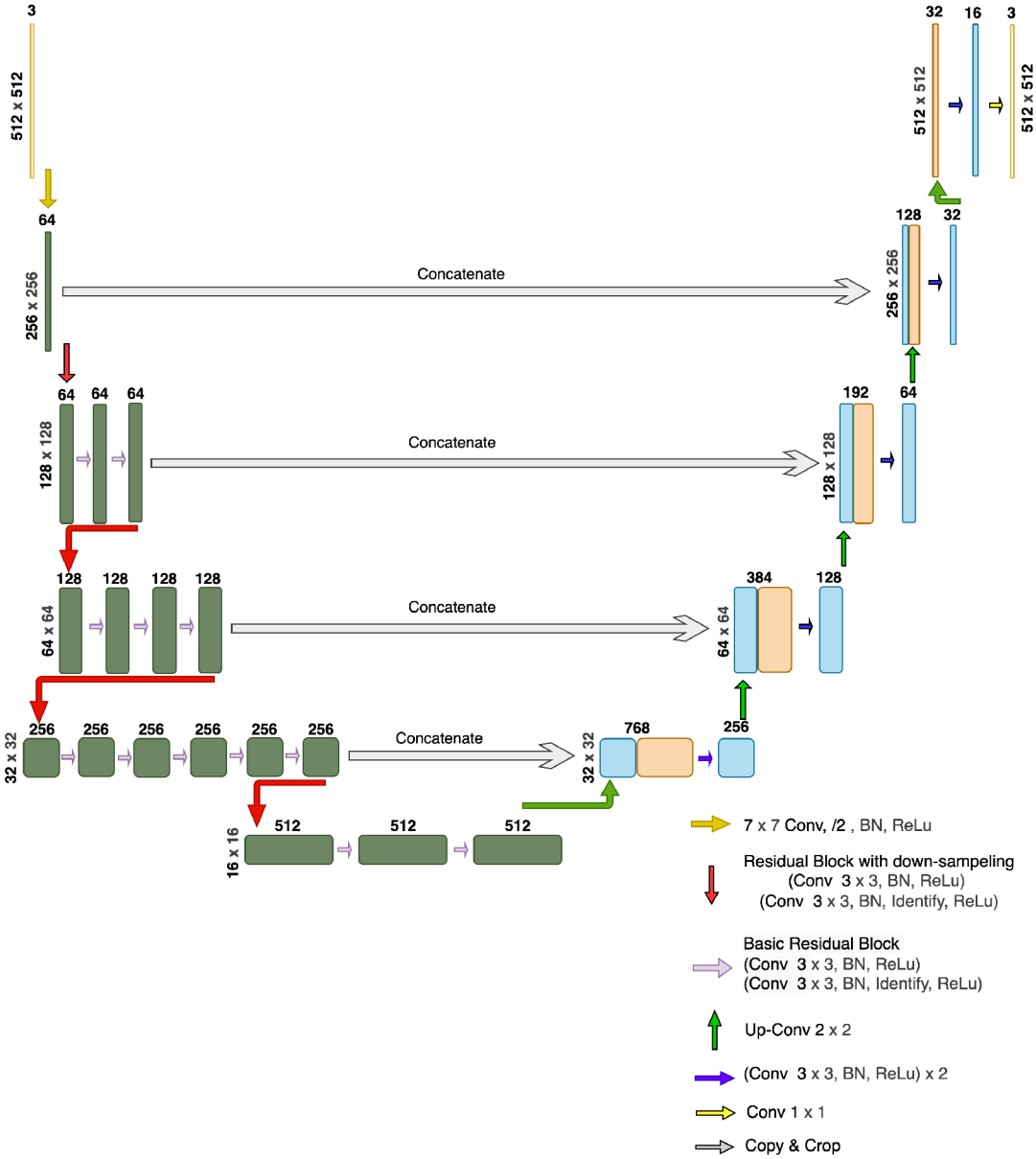

FIGURE 2.9: The hybrid ResNet-34-U-Net architecture.

or more layers and update the gradient values from one or more previous layers into the following layers.

The ResNet-34-U-Net architecture was implemented and applied in our research (Fig. 2.9). The proposed architecture has 34 layers and four residual convolution steps with a total of 16 residual blocks (red and purple arrows). The first convolution layer has 64 filters with a kernel size of 7×7, followed by a max-pooling layer. Each residual block consists of two 3×3 convolution layers followed by the ReLU activation function and batch normalisation with the identity shortcut connection.

The decoder section has the same structure as the simple U-Net architecture. The "softmax" activation function was applied to achieve the probability map across three different classes for each pixel of the input images.

2.6 Model training and evaluation

The implementation platform for cell segmentation was based on Python 3.9. The deep learning framework was Keras with the backend of Tensorflow [100]. The data sets were divided into training (80%) and testing (20%). A part (20%) of the training set was used for model validation in the training process to avoid over-fitting and achieve higher performance.

All data sets were resized to 512×512 pixels, the input image size for training models in the proposed CNNs. The optimised hyperparameter values for single- and multi-class segmentation (Tab. 2.1) were achieved and reported after training the most stable CNN models. The activation function in single- and multi-class segmentation was "LeakyReLU" and "ReLU", respectively. The early stopping hyperparameters were used to avoid over-fitting during the model's training. The patient value was 15 and 30 for training the single- and multi-class model, respectively. The batch size was set to the maximal value of 8 due to the complexity of the CNN structures and GPU-VRAM limitation. The Adam algorithm was chosen to optimise all neural networks. The learning rate was set to 10^{-3} for all CNN models.

TABLE 2.1: Hyperparameters setting for training the models.

Hyperparameter	**Single-class**	**Multi-class**
Activation function	LeakyReLU	ReLU
Learning rate	10^{-3}	10^{-3}
Number of classes	1	3
Batch size	8	8
Epochs number	100	200
Early stop	15	30
Optimizer	Adam	Adam
γ **for loss function**	2	2
Step per epoch	100	52

Image segmentation categorises pixels as either the background or cell classes. The Dice loss was used to compare the segmented cell image with the GT and minimise the difference between them as much as possible in the training process. The "binary focal loss" and "categorical focal loss" was used as the loss function for the single- and multi-class segmentation, respectively.

The segmentation models were evaluated by different metrics (Eqs. 2.1–2.5), where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative metrics, respectively [101]. The metrics were computed for all test sets and explained as mean values.

Overall pixel accuracy (Acc) represents a per cent of image pixels belonging to the correctly segmented cells:

$$\text{Acc} = \frac{\text{Correctly Predicted Pixels}}{\text{Total Number of Image Pixels}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{2.1}$$

Precision (Pre) is a proportion of the cell pixels in the segmentation results that match the GT:

$$\text{Pre} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Predicted Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.2}$$

The Recall (Recl) represents the proportion of cell pixels in the GT correctly identified through the segmentation process:

$$\text{Recl} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Actual Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.3}$$

The F1-score or Dice similarity coefficient states how the predicted segmented region matches the GT in location and level of details and considers each class's false alarm and missed value. This metric determines the accuracy of the segmentation boundaries [102] and has a higher priority than the Acc:

$$\text{Dice} = \frac{2 \times \text{Pre} \times \text{Recl}}{\text{Pre} + \text{Recl}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{2.4}$$

Another essential evaluation metric for semantic image segmentation is the Jaccard similarity index known as Intersection over Union (IoU). This metric is a correlation among the prediction and GT [6, 103], and represents the overlap and union area ratio for the predicted and GT segmentation:

$$\text{IoU} = \frac{|\, y_t \cap y_p \,|}{|\, y_t \,| + |\, y_p \,| - |\, y_t \cap y_p \,|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{2.5}$$

CHAPTER 3

Results and summary

3.1 Single-class segmentation results

The single-class segmentation models were well-trained and converged after 100 epochs, as evaluated by the training/validation loss and Jaccard plots per epoch. The best hyperparameter values reported in Table 2.1 were considered to train the model for the best training performance and stability. Then, the test data sets were used to evaluate the achieved models. All trained models were assessed (Tab. 3.2) using the metrics in Eqs. 2.1–2.5.

TABLE 3.1: Numbers of trainable parameters and the run time for single-class segmentation models.

Network	Run time	Training parameter
U-Net	3:42′:18″	31,402,501
Attention U-Net	4:04′:23″	34,334,665
Residual Att U-Net	4:11′:24″	39,090,377

Model training of the simple U-Net took the shortest run time with the fewest trainable parameters (Tab. 3.1). However, the difference in run time between the Attention U-Net and the Residual Attention U-Net is not huge in increasing trainable parameters. The computational costs also did not increase dramatically compared with the acceptable improvement in the model performance.

The simple U-Net segmentation results suffer from mis-segmentation of some unclear cell borders (Fig. 3.1A, black circle). The Attention U-Net (Fig. 3.1B) detected cells with unclear borders more efficiently than the simple U-Net. However, the Attention U-Net segmentation suffers from under-segmentation in some regions (visualised by the yellow circle). The outcome from the Residual Attention U-Net (Fig. 3.1C, red circle) achieved more accurate segmentation of the unclear cell borders. The watershed binary segmentation after the Residual Attention U-Net separated and identified the cells with the highest performance (Fig. 3.1).

According to the mean-IoU, mean-Dice, and accuracy metrics (Tab. 3.2), the Attention U-Net model showed better segmentation performance than the simple U-Net model in the same situation. The segmentation results were further slightly improved after applying the residual step into the Attention U-Net.

TABLE 3.2: Evaluation of the single-class segmentation models.

Network	Accuracy	Precision	Recall	m-IoU	m-Dice
U-Net	0.957418	0.988269	0.961264	0.950501	0.974481
Attention U-Net	0.959448	0.985663	0.965736	0.952471	0.975511
Residual Att U-Net	0.960010	0.986510	0.965574	0.953085	0.975840

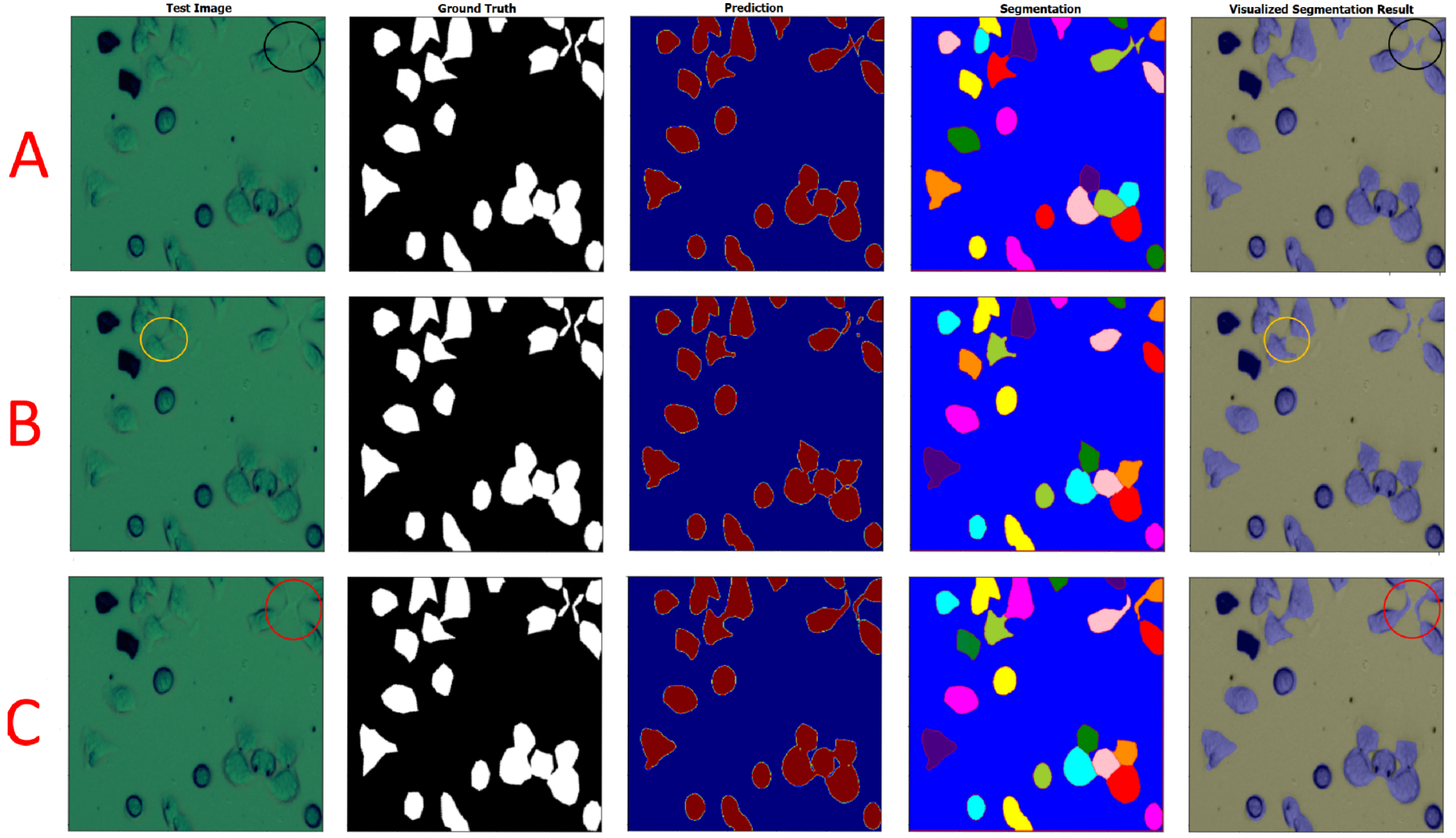

FIGURE 3.1: Segmentation results for A) the simple U-Net (the black circle highlights the non-segmented, unclear cell borders), B) Attention U-Net (the yellow circle highlights the under-segmentation problem), and C) the Residual Attention U-Net (red circle shows the successful segmentation of the cell borders. The image size is 512 × 512.

3.2 Multi-class segmentation results

Multi-class segmentation models were trained well and converged after 200 epochs by observing and evaluating training/validation loss and Jaccard plots. The hyperparameter values listed in Table 2.1 were used to achieve the best training performance and stability. Then, the performances of the trained models were assessed and evaluated using the test data sets and the metrics in Eqs. 2.1–2.5 (Tab. 3.4).

TABLE 3.3: Number of the trainable parameters and the run time for the multi-class models.

Network	Run time	Training parameter
U-Net	3:33′:29″	31,402,639
VGG19-U-Net	1:44′:38″	31,172,163
Inception-U-Net	1:05′:47″	18,083,535
ResNet34-U-Net	0:56′:22″	24,456,444

One of the critical factors in training high-performance models is optimising the computational costs. As presented in Table 3.3, the four methods had significantly different runtimes, the number of trainable parameters, and network structures. Training the simple U-Net took the longest runtime with the most training parameters. The VGG19-U-Net was trained well in a significantly shorter time due to the network structure; the number of training parameters was slightly lower than in the simple U-Net. The Inception-U-Net runtime was even faster than the previous two methods. This runtime reduction led to a further significant decrease in the number of trainable parameters and higher segmentation performance. The ResNet34-U-Net achieved the shortest computational costs with the best segmentation performance.

The results of the multi-class segmentation are shown in Figure 3.2. The simple U-Net obtained a lower categorical segmentation performance in the evaluation phase than the other models. The simple U-Net was inefficient in classifying the cell pixels into the right classes and suffers from wrongly segmented cells into the wrong classes (Fig. 3.2, yellow circle). The VGG19-U-Net showed better categorical segmentation regarding the evaluation metrics (Tab. 3.4). The cells wrongly segmented by the simple U-Net were caught slightly, but the wrong classifications still occurred (Fig. 3.2, purple circle). The Inception-U-Net applied to our data sets as the third hybrid CNN improved the multi-class segmentation results significantly in terms of evaluation metrics (Tab. 3.4). However, this method suffered from over-segmentation in all classes (Fig. 3.2, black circle). The hybrid ResNet34-U-Net obtained the best results in the segmentation and classification into all classes (Tab. 3.4).

TABLE 3.4: Evaluation of the U-Net models for multi-class segmentation.

Network	Accuracy	Precision	Recall	m-IoU	m-Dice
U-Net	0.9869	0.7897	0.8833	0.7062	0.8104
VGG19-Net	0.9865	0.8051	0.8614	0.7178	0.8218
Inception-Net	0.9904	0.8684	0.8905	0.7907	0.8762
ResNet 34-Net	0.9909	0.8795	0.8975	0.8067	0.8873

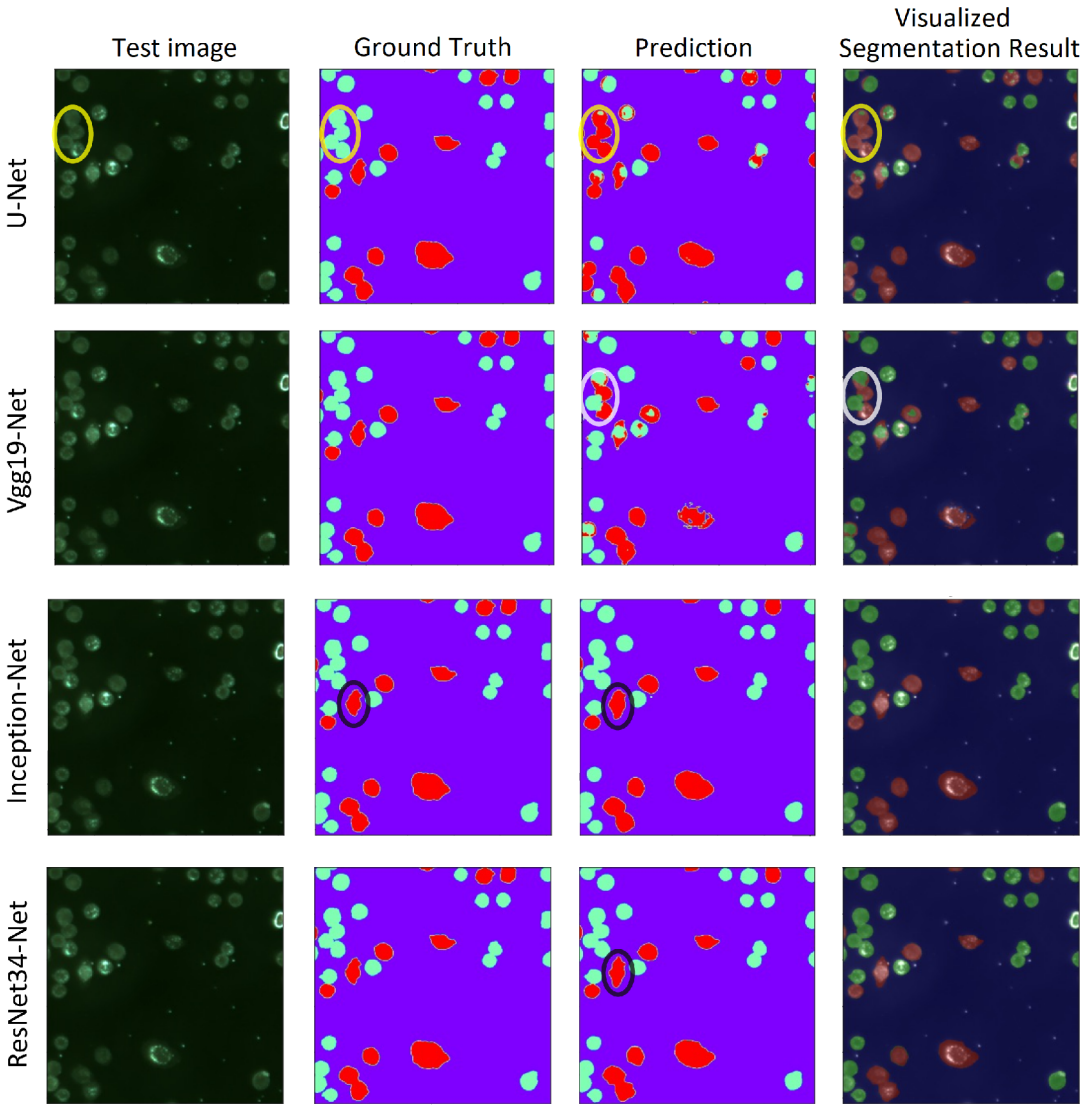

FIGURE 3.2: Test image, ground truth, prediction, and 8-bit visualisation of the segmentation results for the U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net. The yellow and white circles highlight the wrongly classified and segmented cells. The black circle highlights a different, smoother segmentation result achieved by the ResNet34-U-Net. The image size is 512 × 512.

3.3 Summary and conclusion

The main objective of single-class living HeLa cell segmentation research was to develop the most accurate and computationally reasonable method to classify image pixels into either cell or background region in light microscopy images. The image data sets were collected using a custom-made wide-field transmitted light microscope. Microscopy image analysis via deep learning methods was a convenient solution due to the complexity and variability of this data.

Different U-Net deep learning architectures were involved in this research: the simple U-Net, the Attention U-Net, and the Residual Attention U-Net. The simple U-Net showed the fastest training time. On the other hand, the Residual Attention U-Net achieved the best segmentation performance with a run time slightly higher than the other two U-Net models.

The second paper focuses on developing an efficient algorithm to detect and segment living HeLa cells and classify them according to their shapes and life-cycle stages. The time-lapse image series for this research were collected with the reflected setup of our unique wide-filed microscope. This research involved variants of hybrid U-Net-based CNN architecture: a simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-net.

The simple U-Net has the longest training time, the highest number of trainable parameters, and the lowest categorical segmentation performance. In contrast, the hybrid ResNet34-U-Net achieved the best categorical segmentation performance with a run time significantly lower than the other models. The Residual Convolutional Neural Network (ResNet) was applied as a hybrid with the U-Net to overcome the gradient vanishing and improve the generalisation ability during training. Using a series of residual blocks with skip connections in each level of the ResNet34-U-Net network resulted in better categorical segmentation.

In conclusion, DL-based methods to analyze microscopy images deliver accurate and promising outcomes for cell segmentation purposes. The proposed single- and multi-class cell segmentation methods successfully segmented living cells and classified them into categories with a high level of accuracy.

According to our best knowledge, not many similar researches on transmitted and reflected wide-field microscopy data have been done before. However, the achieved segmentation results were compared with other types of microscopy and medical research outcomes and show remarkable differences in segmentation results as reported in papers in Chapter 4. The proposed single and multi-class segmentation methods have general utilization for hyperparameters tuning and model training of different microscopy, medical or, even, remote sensing datasets.

Bibliography

[1] Telecentric objective (2023).
URL https://www.stemmer-imaging.com/s/category/products/optics/telecentric-lenses/0ZG6N00000000AhWAI?language=en_US&c__results_layout_state=%7B%7D

[2] D. Štys, T. Náhlík, P. Macháček, R. Rychtáriková, M. Saberioon, Least information loss (lil) conversion of digital images and lessons learned for scientific image inspection, in: F. Ortuño, I. Rojas (Eds.), Bioinformatics and Biomedical Engineering, Springer International Publishing, Cham, 2016, pp. 527–536.

[3] S. Abe, Support Vector Machines for Pattern Classification, Springer, 2010.

[4] Random forest (203).
URL https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/

[5] K-mean (2023).
URL www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

[6] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440. doi:10.1109/CVPR.2015.7298965.

[7] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9321, Springer, Cham, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

[8] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, E. Wu, X. Tang, Prostate segmentation using 2D Bridged U-Net, in: International Joint Conference on Neural Networks – IJCNN 2019, 2019, pp. 1–7. doi:10.1109/IJCNN.2019.8851908.

[9] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, V. K. Asari, Recurrent residual u-net (r2u-net) for medical image segmentation, Journal of Medical Imaging 6 (1) (2019) 014006. doi:10.1117/1.JMI.6.1.014006.

[10] F. D. Khameneh, S. Razavi, M. Kamasak, Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network, Computers in Biology and Medicine 110 (2019) 164–174. doi:10.1016/j.compbiomed.2019.05.020.

[11] W. F. Scherer, J. T. Syverton, G. O. Gey, Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix, Journal of Experimental Medicine 97 (5) (1953) 695–710. doi:10.1084/jem.97.5.695.

[12] R. Rahbari, T. Sheahan, V. Modes, P. Collier, C. Macfarlane, R. M. Badge, A novel L1 retrotransposon marker for HeLa cell line identification, Biotechniques 46 (4) (2009) 277–284. doi:10.2144/000113089.

[13] R. Rychtáriková, T. Náhlík, K. Shi, D. Malakhova, P. Macháček, R. Smaha, J. Urban, D. Štys, Super-resolved 3-d imaging of live cells' organelles from bright-field photon transmission micrographs, Ultramicroscopy 179 (2017) 1–14. doi:10.1016/j.ultramic.2017.03.018.

[14] K. Lonhus, R. Rychtáriková, G. Platonova, D. Štys, Quasi-spectral characterization of intracellular regions in bright-field light microscopy images, Scientific Reports 10 (1) (2020). doi:10.1038/s41598-020-75441-7.

[15] R. Coico, Gram staining, Current Protocols in Microbiology (2005). doi:10.1002/9780471729259.mca03cs00.

[16] R. C. Gonzalez, R. E. Woods, Digital Image Processing (3rd Edition), Prentice-Hall, Inc., Division of Simon and Schuster One Lake Street Upper Saddle River, NJ United States, 2006.

[17] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66. doi:10.1109/TSMC.1979.4310076.

[18] C. Callau, M. Lejeune, A. Korzynska, M. García-Rojo, G. Bueno, R. Bosch, J. Jaén, G. Orero, T. Salvadó, C. López, Evaluation of cytokeratin-19 in breast cancer tissue samples: a comparison of automatic and manual evaluations of scanned tissue microarray cylinders, Biomedical Engineering Online 14 (2015) S2. doi:10.1186/1475-925X-14-S2-S2.

[19] X. Zhou, F. Li, J. Yan, S. T. C. Wong, A novel cell segmentation method and cell phase identification using Markov model, IEEE Transactions on Information Technology in Biomedicine 13 (2) (2009) 152–157. doi:10.1109/TITB.2008.2007098.

[20] O. Schmitt, M. Hasse, Morphological multiscale decomposition of connected regions with emphasis on cell clusters, Computer Vision and Image Understanding 113 (2) (2009) 188–201. doi:10.1016/j.cviu.2008.08.011.

[21] Q. Wang, J. Niemi, C. M. Tan, L. You, M. West, Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy, Cytometry 77 (1) (2010) 101–110. doi:10.1002/cyto.a.20812.

[22] E. Meijering, Cell segmentation: 50 years down the road [life sciences], IEEE Signal Processing Magazine 29 (5) (2012) 140–145. doi:10.1109/MSP.2012.2204190.

[23] U. Adiga, R. Malladi, R. Fernandez-Gonzalez, C. O. de Solorzano, High-throughput analysis of multispectral images of breast cancer tissue, IEEE Image Processing 15 (8) (2006) 2259–2268. doi:10.1109/TIP.2006.875205.

[24] F. Li, X. Zhou, J. Ma, S. T. C. Wong, Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis, IEEE Transactions on Medical Imaging 29 (1) (2009) 96–115. doi:10.1109/TMI.2009.2027813.

[25] J. Cheng, J. C. Rajapakse, Segmentation of clustered nuclei with shape markers and marking function, IEEE Transactions on Biomedical Engineering 56 (3) (2009) 741–748. doi:10.1109/TBME.2008.2008635.

[26] X. Zhou, K. Y. Liu, P. Bradley, N. Perrimon, S. T. C. Wong, Towards automated cellular image segmentation for RNAi genome-wide screening, Medical Image Computing and Compute-Assisted Intervention 8 (Pt 1) (2005) 885–892. doi:10.1007/11566465_109.

[27] R. O. Duda, P. E. Hart, Use of the Hough transformation to detect lines and curves in pictures, Communications of the ACM 15 (1) (1972) 11–15. doi:10.1145/361237.361242.

[28] C. Zhang, F. Huber, M. Knop, F. A. Hamprecht, Yeast cell detection and segmentation in bright field microscopy, in: IEEE 11th International Symposium on Biomedical Imaging – ISBI 2014, 2014, pp. 1267–1270. doi:10.1109/ISBI.2014.6868107.

[29] P. Filipczuk, T. Fevens, A. Krzyzak, R. Monczak, Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies, IEEE Transactions on Medical Imaging 32 (12) (2013) 2169–2178. doi:10.1109/TMI.2013.2275151.

[30] K. R. Spring, J. C. Russ, M. J. Parry-Hill, T. J. Fellers, M. W. Davidson, Molecular expressions microscopy primer: Digital image processing (2022).
URL https://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/diffgaussians/index.html

[31] H. Peng, X. Zhou, F. Li, X. Xia, S. T. C. Wong, Integrating multiscale blob/curvilinear detector techniques and multi-level sets for automated segmentation of stem cell images, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009, pp. 1362–1365. doi:10.1109/ISBI.2009.5193318.

[32] F. Li, X. Zhou, H. Zhao, S. T. C. Wong, Cell segmentation using front vector flow guided active contours, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009. Lecture Notes in Computer Science, Vol. 5462, Springer Berlin Heidelberg, 2009, pp. 609–616. doi:10.1007/978-3-642-04271-3_74.

[33] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, Image and Vision Computing 22 (10) (2004) 761–767. doi:10.1016/j.imavis.2004.02.006.

[34] Z. Lu, G. Carneiro, A. P. Bradley, Automated nucleus and cytoplasm segmentation of overlapping cervical cells, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Lecture Notes in Computer Science, Vol. 8149, Springer Berlin Heidelberg, 2013, pp. 452–460. doi:10.1007/978-3-642-40811-3_57.

[35] C. Arteta, V. Lempitsky, J. A. Noble, A. Zisserman, Learning to detect cells using non-overlapping extremal regions, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. Lecture Notes in Computer Science, Vol. 7510, Springer Berlin Heidelberg, 2012, pp. 348–356. doi:10.1007/978-3-642-33415-3_43.

[36] F. Buggenthin, C. Marr, M. Schwarzfischer, P. S. Hoppe, O. Hilsenbeck, T. Schroeder, F. J. Theis, An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy, BMC Bioinformatics 14 (2013) 297. doi:10.1186/1471-2105-14-297.

[37] T. M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math, New York, 1997.

[38] C. M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.

[39] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297. doi:10.1007/BF00994018.

[40] T. Janssens, L. Antanas, S. Derde, I. Vanhorebeek, G. V. den Berghe, F. G. Grandas, CHARISMA: An integrated approach to automatic h&e-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting, Medical Image Analysis 17 (8) (2013) 1206–1219. doi:10.1016/j.media.2013.07.007.

[41] L. Cheng, N. Ye, W. Yu, A. Cheah, Discriminative segmentation of microscopic cellular images, in: G. Fichtinger, A. Martel, T. Peters (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, Springer Berlin Heidelberg, 2011, pp. 637–644. doi:10.1007/978-3-642-23623-5_80.

[42] T. Tikkanen, P. Ruusuvuori, L. Latonen, H. Huttunen, Training based cell detection from bright-field microscope images, in: 2015 9th International Symposium on Image and Signal Processing and Analysis – ISPA 2015, 2015, pp. 160–164. doi:10.1109/ISPA.2015.7306051.

[43] C. Sommer, L. Fiaschi, F. A. Hamprecht, D. W. Gerlich, Learning-based mitotic cell detection in histopathological images, in: Proceedings of the 21st International Conference on Pattern Recognition – ICPR2012, 2012, pp. 2306–2309.

[44] G. Lupica, N. M. Allinson, S. W. Botchway, Hybrid image processing technique for the robust identification of unstained cells in bright-field microscope images, in: Mohammadian, M (Ed.), 2008 International Conference on Computational Intelligence for Modelling Control & Automation, IEEE, New York, 2008, pp. 1053–1058. doi:10.1109/CIMCA.2008.144.

[45] D. von Winterveldt, W. Edwards, Decision Analysis and Behavioral Research, Cambridge University Press, 1986, Ch. Decision trees, p. 63–89.

[46] T. K. Ho, Random decision forests, in: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Vol. 1, 1995, pp. 278–282. doi:10.1109/ICDAR.1995.598994.

[47] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, S. Steidl, R. Buchholz, J. Hornegger, Unsupervised unstained cell detection by SIFT keypoint clustering and self-labeling algorithm, in: P. Golland, N. Hata, C. Barillot, J. Hornegger, R. Howe (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. Lecture Notes in Computer Science, Vol. 8675, Springer International Publishing, Cham, 2014, pp. 377–384. doi:10.1007/978-3-319-10443-0_48.

[48] S. A. Mah, R. Avci, P. Du, J.-M. Vanderwinden, L. K. Cheng, Supervised machine learning segmentation and quantification of gastric pacemaker cells, in: 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society – EMBC 2020, 2020, pp. 1408–1411. doi:10.1109/EMBC44109.2020.9176445.

[49] J. Gall, A. Yao, N. Razavi, L. Van Gool, V. Lempitsky, Hough forests for object detection, tracking, and action recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (11) (2011) 2188–2202. doi:10.1109/TPAMI.2011.70.

[50] A. Yao, J. Gall, C. Leistner, L. Van Gool, Interactive object detection, in: IEEE Conference on Computer Vision and Pattern Recognition – CVPR 2012, 2012, pp. 3242–3249. doi:10.1109/CVPR.2012.6248060.

[51] K. Liimatainen, P. Ruusuvuori, L. Latonen, H. Huttunen, Supervised method for cell counting from bright field focus stacks, in: 2016 IEEE 13th International Symposium on Biomedical Imaging – ISBI, 2016, pp. 391–394. doi:10.1109/ISBI.2016.7493290.

[52] Z. Yin, R. Bise, M. Chen, T. Kanade, Cell segmentation in microscopy imagery using a bag of local Bayesian classifiers, IEEE International Symposium on Biomedical Imaging (2010) 125–128doi:10.1109/ISBI.2010.5490399.

[53] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, A. Madabhushi, Expectation–Maximization-driven geodesic active contour with

overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology, IEEE Transactions on Biomedical Engineering 57 (7) (2010) 1676–1689. doi:10.1109/TBME.2010.2041232.

[54] G. Hinton, T. Sejnowski, Unsupervised Learning: Foundations of Neural Computation, MIT Press, MIT Press, 1999.

[55] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2nd ed., Springer, Berlin, 2011.

[56] J. B. MacQueen, Some methods for classification and analysis of multivariate observations, in: L. M. L. Cam, J. Neyman (Eds.), Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, 1967, p. 281–297.

[57] X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, Micron 107 (2018) 55–71. doi:10.1016/j.micron.2018.01.010.

[58] B. Antal, B. Remenyik, A. Hajdu, An unsupervised ensemble-based Markov Random Field approach to microscope cell image segmentation, in: 2013 International Conference on Signal Processing and Multimedia Applications – SIGMAP 2013, 2013, pp. 94–99.

[59] H. Schulz, S. Behnke, Deep learning, KI - Künstliche Intelligenz 26 (4) (2012) 357–363. doi:10.1007/s13218-012-0198-z.

[60] R. Yamashita, M. Nishio, R. K. G. Do, K. Togashi, Convolutional neural networks: an overview and application in radiology, Insights into Imaging 9 (2018) 611–629. doi:10.1007/s13244-018-0639-9.

[61] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, , Y. LeCun, OverFeat: Integrated recognition, localization and detection using convolutional networks (2013). doi:10.48550/arXiv.1312.6229.

[62] A. Sadafi, M. Radolko, I. Serafeimidis, S. Hadlak, Red blood cells segmentation: A fully convolutional network approach, in: 2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications – ISPA/IUCC/BDCloud/SocialCom/SustainCom, 2018, pp. 911–914. doi:10.1109/BDCloud.2018.00134.

[63] S. Lin, N. Norouzi, An effective deep learning framework for cell segmentation in microscopy images, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society – EMBC, 2021, pp. 3201–3204. doi:10.1109/EMBC46164.2021.9629863.

[64] D. C. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Lecture Notes in Computer Science, Vol. 8150, Springer Berlin Heidelberg, 2013, pp. 411–418. doi:10.1007/978-3-642-40763-5_51.

[65] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, T. Wang, Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning, IEEE Transactions on Biomedical Engineering 62 (10) (2015) 2421–2433. doi:10.1109/TBME.2015.2430895.

[66] F. Liu, L. Yang, A novel cell detection method using deep convolutional neural network and maximum-weight independent set, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9351, Springer International Publishing, 2015, pp. 349–357. doi:10.1007/978-3-319-24574-4_42.

[67] Y. Xie, X. Kong, F. Xing, F. Liu, H. Su, L. Yang, Deep voting: A robust approach toward nucleus localization in microscopy images, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9351, Springer International Publishing, 2015, pp. 374–382. doi:10.1007/978-3-319-24574-4_45.

[68] Y.-H. Chang, K. Abe, H. Yokota, K. Sudo, Y. Nakamura, M.-D. Tsai, Human induced pluripotent stem cell region detection in bright-field microscopy images using Convolutional Neural Networks, Biomedical Engineering: Applications Basis and Communications 31 (2) (2019) 1950009. doi:10.4015/S1016237219500091.

[69] P. Thi Le, T. Pham, Y.-C. Hsu, J.-C. Wang, Convolutional blur attention network for cell nuclei segmentation, Sensors 22 (4) (2022) 1586. doi:10.3390/s22041586.

[70] N. Kumar, R. Verma, D. Anand, A. Sethi, Multi-organ nuclei segmentation challenge, Retrieved on 05/05/2021 (2021).
URL https://monuseg.grandchallenge.org/

[71] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, A. E. Carpenter, Broad bioimage benchmark collection, Retrieved on 05/05/2021 (2021).
URL https://bbbc.broadinstitute.org/BBBC038

[72] J. Yi, P. Wu, M. Jiang, Q. Huang, D. J. Hoeppnerb, D. N. Metaxas, Attentive neural cell instance segmentation, Medical Image Analysis 55 (2019) 228–240. doi:10.1016/j.media.2019.05.004.

[73] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, A. C. Berg, DSSD: Deconvolutional single shot detector (2017). doi:10.48550/arXiv.1701.06659.

[74] T. Wan, S. Xu, C. Sang, Y. Jin, Z. Qin, Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks, Neurocomputing 365 (2019) 157–170. doi:10.1016/j.neucom.2019.06.086.

[75] F. Long, Microscopy cell nuclei segmentation with enhanced U-Net, BMC Bioinformatics 21 (2020) 8. doi:10.1186/s12859-019-3332-1.

[76] S. Bagyaraj, R. Tamilselvi, P. B. M. Gani, D. Sabarinathan, Brain tumour cell segmentation and detection using deep learning networks, IET Image Processing 15 (10) (2021) 2363–2371. doi:10.1049/ipr2.12219.

[77] E. Shibuya, K. Hotta, Cell image segmentation by using feedback and convolutional LSTM, The Visual Computer 38 (2022) 3791–3801. doi:10.1007/s00371-021-02221-3.

[78] S. Pereira, A. Pinto, V. Alves, C. A. Silva, Brain tumor segmentation using convolutional neural networks in mri images, IEEE Transactions on Medical Imaging 35 (5) (2016) 1240–1251. doi:10.1109/TMI.2016.2538465.

[79] J. Stawiaski, A pretrained DenseNet encoder for brain tumor segmentation, in: A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, T. van Walsum (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer International Publishing, Cham, 2019, pp. 105–115.

[80] S. P. Sunny, A. I. Khan, M. Rangarajan, A. Hariharan, P. B. N, H. J. Pandya, N. Shah, M. A. Kuriakose, A. Suresh, Oral epithelial cell segmentation from fluorescent multichannel cytology images using deep learning, Computer Methods and Programs in Biomedicine 227 (2022) 107205. doi:10.1016/j.cmpb.2022.107205.

[81] M. E. Bakir, v. H. Yalim Keles, Deep learning based cell segmentation using cascaded U-Net models, in: 2021 29th Signal Processing and Communications Applications Conference – SIU, 2021, pp. 1–4. doi:10.1109/SIU53274.2021.9477937.

[82] T. Piotrowski, O. Rippel, A. Elanzew, B. Nießing, S. Stucken, S. Jung, N. König, S. Haupt, L. Stappert, O. Brüstle, R. Schmitt, S. Jonas, Deep-learning-based multi-class segmentation for automated, non-invasive routine assessment of human pluripotent stem cell culture status, Computers in Biology and Medicine 129 (2021) 104172. doi:10.1016/j.compbiomed.2020.104172.

[83] H. Yu, F. Wang, G. Teodoro, J. Nickerson, J. Kong, MultiHeadGAN: A deep learning method for low contrast retinal pigment epithelium cell

segmentation with fluorescent flatmount microscopy images, Computers in Biology and Medicine 146 (2022) 105596. doi:10.1016/j.compbiomed.2022.105596.

[84] Y. Zhao, C. Fu, S. Xu, L. Cao, H. feng Ma, LFANet: Lightweight feature attention network for abnormal cell segmentation in cervical cytology images, Computers in Biology and Medicine 145 (2022) 105500. doi:10.1016/j.compbiomed.2022.105500.

[85] D. Eschweiler, T. V. Spina, R. C. Choudhury, E. Meyerowitz, A. C. J. Stegmaier, CNN-based preprocessing to optimize watershed-based cell segmentation in 3D confocal microscopy images, in: IEEE 16th International Symposium on Biomedical Imaging – ISBI 2019, 2019, pp. 223–227. doi:10.1109/ISBI.2019.8759242.

[86] R. Khan, J. Mir, White blood cells segmentation and classification using U-Net CNN and hand-crafted features, in: IEEE International Conference on IT and Industrial Technologies – ICIT, 2022, pp. 1–7. doi:10.1109/ICIT56493.2022.9988955.

[87] T. Tran, O.-H. Kwon, K.-R. Kwon, S.-H. Lee, K.-W. Kang, Blood cell images segmentation using deep learning semantic segmentation, in: IEEE International Conference on Electronics and Communication Engineering – ICECE, 2018, pp. 13–16. doi:10.1109/ICECOME.2018.8644754.

[88] W. Liu, A. Rabinovich, A. C. Berg, ParseNet: Looking wider to see better, in: 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, 2016, p. 11.

[89] Vision-telecentric-obj, https://www.vision-control.com/uploads/tx_vcproducts/article/TO4.5_43.4-48-F-WN_V1.0_en_20220406.pdf, downloaded 22/03/2023 (2023).

[90] Cl-41, https://www.optikamicroscopes.com/optikamicroscopes/product/cl-41/, downloaded 22/03/2023 (2023).

[91] Schott visiled s80-25 led brightfield ringlight 400225, https://spectraservices.com/product/schott-400-225.html, downloaded 22/03/2023 (2023).

[92] G. Platonova, D. Štys, P. Souček, K. Lonhus, J. Valenta, R. Rychtáriková, Spectroscopic approach to correction and visualization of bright-field light transmission microscopy biological data, Photonics 8 (8) (2021) 333. doi:10.3390/photonics8080333.

[93] D. Štys, T. Náhlík, P. Macháček, R. Rychtáriková, M. Saberioon, Least information loss (LIL) conversion of digital images and lessons learned for scientific image inspection, in: F. Ortuno, I. Rojas (Eds.), Bioinformatics and Biomedical Engineering. IWBBIO 2016. Lecture Notes

in Computer Science, Vol. 9656, Springer, Cham, 2016, pp. 527–536. doi:10.1007/978-3-319-31744-1_47.

[94] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, 2005, pp. 60–65. doi:10.1109/CVPR.2005.38.

[95] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning where to look for the pancreas, in: 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), 2018, pp. –.

[96] W. Zhang, D. Jiang, The marker-based watershed segmentation algorithm of ore image, in: 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011, pp. 472–474. doi:10.1109/ICCSN.2011.6014611.

[97] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: The 3rd International Conference on Learning Representations (ICLR2015), 2015, pp. 1–14. doi:10.48550/arXiv.1409.1556.

[98] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9. doi:10.1109/CVPR.2015.7298594.

[99] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[100] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, TensorFlow: large-scale machine learning on heterogeneous distributed systems, in: Proc. USENIX Symp. Oper. Syst. Des. Implement., 2016, pp. 265–283.

[101] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, Y. Fan, Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks, Neurocomputing 229 (2017) 88–99. doi:10.1016/j.neucom.2016.08.103.

[102] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure for semantic segmentation?, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2013, pp. 32.1–32.11. doi:10.5244/C.27.32.

[103] B. Vijay, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern

Anal. Mach. Intell. 39 (12) (2015) 228–233. doi:10.1109/TPAMI.2016.2644615.

CHAPTER 4

Original papers

Paper 1

Cell segmentation from telecentric bright-field transmitted light microscopy images using a Residual Attention U-Net: A case study on HeLa line

Authors: **Ghaznavi, A.**, Rychtáriková, R., Saberioon, M., and Štys, D.

Computers in Biology and Medicine 147 (2022) 105805

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Cell segmentation from telecentric bright-field transmitted light microscopy images using a Residual Attention U-Net: A case study on HeLa line

Ali Ghaznavi [a], Renata Rychtáriková [a,*], Mohammadmehdi Saberioon [b], Dalibor Štys [a]

[a] Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 373 33, Nové Hrady, Czech Republic

[b] Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Telegrafenberg, Potsdam 14473, Germany

ARTICLE INFO

Keywords:
Deep learning
Neural network
Cell detection
Microscopy image segmentation
Tissue segmentation
Semantic segmentation
Watershed segmentation

ABSTRACT

Living cell segmentation from bright-field light microscopy images is challenging due to the image complexity and temporal changes in the living cells. Recently developed deep learning (DL)-based methods became popular in medical and microscopy image segmentation tasks due to their success and promising outcomes. The main objective of this paper is to develop a deep learning, U-Net-based method to segment the living cells of the HeLa line in bright-field transmitted light microscopy. To find the most suitable architecture for our datasets, a residual attention U-Net was proposed and compared with an attention and a simple U-Net architecture.

The attention mechanism highlights the remarkable features and suppresses activations in the irrelevant image regions. The residual mechanism overcomes with vanishing gradient problem. The Mean-IoU score for our datasets reaches 0.9505, 0.9524, and 0.9530 for the simple, attention, and residual attention U-Net, respectively. The most accurate semantic segmentation results was achieved in the Mean-IoU and Dice metrics by applying the residual and attention mechanisms together. The watershed method applied to this best – Residual Attention – semantic segmentation result gave the segmentation with the specific information for each cell.

1. Introduction

Image object detection and segmentation can be defined as a procedure to localize a region of interest (ROI) in an image and separate an image foreground from its background using image processing and/or machine learning approaches. Cell detection and segmentation are the primary and critical steps in microscopy image analysis. These processes play an important role in estimating the number of the cells, initializing cell segmentation, tracking, and extracting features necessary for further analysis. In the text below, the segmentation methods were categorized as (1) traditional, feature- and machine learning (ML)-based methods and (2) deep learning (DL)-based methods.

1.1. Traditional cell segmentation methods

Traditional segmentation methods have achieved impressive results in cell boundary detection and segmentation, with an efficient processing time [1,2]. These methods include low-level pixel processing approaches. The region-based methods are more robust than the threshold-based segmentation methods [2]. However, in low-contrast images, cells placed close together or flat cell regions can be segmented as blobs. Rojas-Moraleda et al. [1] proposed a region-based method on the principles of persistent homology with an overall accuracy of 94.5%. The iterative morphological and Ultimate Erosion [3,4] suffer from poor segment performance when facing small and low-contrast objects. Guan et al. [5] detected rough circular cell boundaries using the Hough transform and the exact cell boundaries using fuzzy curve tracing. Compared with the watershed-based method [6], this method was more robust to the noise and the uneven brightness in the cells. Winter et al. [7] combined the image Euclidean distance transformation with the Gaussian mixture model to detect elliptical cells. This method requires solid objects for computing the distance transform. The target objects' large holes or extreme internal irregularities make the distance transform unreliable and reduce the method performance. Buggenthin et al. [8] identified nearly all cell bodies and segmented multiple cells instantly in bright-field time-lapse microscopy images by a fast, automatic method combining the Maximally Stable Extremal

* Corresponding author.
E-mail addresses: ghaznavi@frov.jcu.cz (A. Ghaznavi), rrychtarikova@frov.jcu.cz (R. Rychtáriková), saberioon@gfz-potsdam.de (M. Saberioon), stys@frov.jcu.cz (D. Štys).

https://doi.org/10.1016/j.compbiomed.2022.105805
Received 17 March 2022; Received in revised form 3 June 2022; Accepted 26 June 2022
Available online 28 June 2022

Regions (MSER) with the watershed method. The main challenges for this method remain the oversegmentation and poor performance for out-of-focus images.

The machine learning methods have expanded due to the microscopy images' complexity and the previous methods' low performance to detect and segment cells. The ML methods can be classified into two groups: supervised vs unsupervised. The supervised methods produce a mathematical function or model from the training data to map a new data sample [9]. Mualla et al. [10] utilized the Scale Invariant Feature Transform (SIFT) as a feature extractor and the Balanced Random Forest as a classifier to calculate the descriptive cell keypoints. The SIFT descriptors were invariant to illumination conditions, cell size, and orientation. Tikkanen et al. [11] developed a method based on the Histogram of Oriented Gradients (HOG) and the Support Vector Machine (SVM) to extract feature descriptors and classify them as a cell or a non-cell in bright-field microscopy data. The proposed method is susceptible to the number of iterations in the training process as a crucial step to eliminating false positive detections.

The unsupervised ML algorithms require no pre-assigned labels or scores for the training data [12]. The best known unsupervised methods are clustering methods. Mualla et al. [13] segmented unstained cells in bright-field micrographs using a combination of a SIFT to extract key points, a self-labelling, and two clustering methods. This method is fast and accurate but sensitive to the feature selection step to avoid overfitting.

1.2. Deep learning cell segmentation methods

In the last decade, Deep Learning has emerged as a new area of machine learning. The DL methods contain a class of ML techniques that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation for pattern analysis and classification. The Deep Convolutional Networks exhibited impressive performance in many visual recognition tasks [14]. Song et al. [15] used a multiscale convolutional network (MSCN) to extract scale-invariant features and graph-partitioning method for accurate segmentation of cervical cytoplasm and nuclei. This method significantly improved the Dice metric and standard deviation compared with similar methods. Shibuya et al. [16] proposed the Feedback U-Net using the convolutional Long Short-Term Memory (LSTM) network for cell image segmentation, working on four classes of *Drosophila* cell image dataset. However, the proposed method suffered from a low accuracy rate depending on the segmented class. Thi et al. [17] proposed a convolutional blur attention (CBA) network. The network consists of down- and upsampling procedures for nuclei segmentation in standard challenge datasets [18,19]. The authors achieved a good value of the aggregated Jaccard index. The reduced number of trainable parameters led to a reasonable decrease in the computational cost. Xing et al. [20] also proposed an automated nucleus segmentation method based on a deep convolutional neural network (DCNN) to generate a probability map. However, the proposed mitosis counting remains laborious and subjective to the observer.

One of the most popular models for semantic segmentation is Fully Convolutional Network (FCN) architectures. The FCN combines deep semantic information with a shallow appearance to achieve satisfactory segmentation results. The convolutional networks can take the arbitrary size of input images to train end-to-end, pixel-to-pixel, and produce an output of the corresponding size with efficient inference and learning to achieve semantic segmentation in complex images, including microscopy and medical images [21,22]. Ronneberger et al. [23] proposed a training strategy that relies on the strong use of data augmentation by applying U-Net Neural Network, contracting the path to capture context, and expanding the path symmetrically to achieve a precise localization. This method was optimized with a low amount of training labelled samples and efficiently performed electron microscopy image segmentation. Long et al. [24] proposed an enhanced U-Net-based architecture called light-weighted U-Net (U-Net+) with a modified encoded branch for potential low-resources computing of nuclei segmentation in bright-field, dark-field, and fluorescence microscopy images. However, the proposed method did not achieve higher accuracy in the Mean-IoU metric. Bagyaraj et al. [25] proposed two automatic deep learning networks called U-Net-based deep convolution network and U-Net with a dense convolutional network (DenseNet) for segmentation and detection of brain tumour cells. The authors achieved remarkable results by applying the DenseNet architecture.

As described above, traditional ML methods are not much efficient to segment cells in a microscopy image with a complex background, particularly bright-field microscopy tiny cells [8,11,13]. These methods cannot build sufficient models for big datasets. On the other hand, some Convolution Neural Networks (CNNs) require a vast number of manually labelled training datasets and higher computational costs compared with the ML methods [21,26].

Deep learning-based methods have delivered better outcomes in segmentation tasks than other methods. Therefore, the main objective of this research is to propose a highly accurate and reasonably computationally cost deep learning-based method to segment human HeLa cells in unique telecentric bright-field transmitted light microscopy images. The U-Net was chosen since it is one of the most promising methods used in semantic segmentation [23]. Different U-Net architectures such as Attention and Residual Attention U-Net were examined to find the most suitable architecture for our datasets.

Human Negroid cervical epithelioid carcinoma line HeLa [27] was chosen as a testing cell line for described microscopy image segmentation. The reason for choosing is that HeLa is the oldest, immortal, and most used model cell line ever. HeLa is cultivated in almost all tissue and cell laboratories worldwide and utilized in many fields of medical research, such as research on carcinoma or testing the material biocompatibility.

The processed microscopy data are specific to high-pixel resolution in rgb mode and requires preprocessing to suppress optical vignetting and camera noise. The data shows unlabelled living cells in their physiological state. The cells are shown in-focused and out-of-focus. Thus, the obtained segmentation method is applicable in a 3D visualization of the cell.

2. Materials and methods

2.1. Cell preparation and microscope specification

Human HeLa cell line (European Collection of Cell Cultures, Cat. No. 93021013) was cultivated to low optical density overnight at 37 °C, 5% CO_2, and 90% relative humidity. The nutrient solution consisted of Dulbecco's modified Eagle medium (87.7%) with high glucose (>1 g L^{-1}), fetal bovine serum (10%), antibiotics and antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all purchased from Biowest, Nuaille, France). The HeLa cells were maintained in a Petri dish with a cover glass bottom and lid at room temperature of 37 °C.

Time-lapse image series of living human HeLa cells on the glass Petri dish were captured using a high-resolved bright-field light microscope for observation of microscopic objects and cells. This microscope was designed by the Institute of Complex System (ICS, Nové Hrady, Czech Republic) and built by Optax (Prague, Czech Republic) and ImageCode (Brloh, Czech Republic) in 2021. The microscope has a simple construction of the optical path. The light from two light-emitting diods CL-41 (Optika Microscopes, Ponteranica, Italy) passes through a sample to reach a telecentric measurement objective TO4.5/43.4-48-F-WN (Vision & Control GmbH, Shul, Germany) and an Arducam AR1820HS 1/2.3-inch 10-bit RGB camera with a chip of 4912 × 3684 pixel resolution. The images were captured as a primary (raw) signal with theoretical pixel size (size of the object projected onto the camera pixel) of 113 nm. The software (developed by the ICS) controls the capture of the primary signal with the camera exposure of 2.75 ms. All these experiments were performed in time-lapse to observe cells' behaviour over time.

2.2. Data acquisition

Different time-lapse experiments on the HeLa cells were completed under the bright-field microscope (Section 2.1). The algorithm proposed in [28] was fully automated and implemented in the microscope control software to calibrate the microscope optical path and correct all image series to avoid image background inhomogeneities and noise.

After the image calibration, we converted the raw image representations to 8-bit colour (rgb) images of resolution (number of pixels) quarter of the original raw images. We employed quadruplets of Bayer mask pixels [29]: Red and blue camera filter pixels were adopted into the relevant image channel and each pair of green camera filter pixels' intensities were averaged to create the green image channel. Then, images were rescaled to 8-bits after creating the image series intensity histogram and omitting unoccupied intensity levels. This bit reduction ensured the maximal information preservation and mutual comparability of the images through the time-lapse series.

The means denoising method [30] minimized the background noise in the constructed RGB images at preserving the texture details. Afterwards, the image series were cropped to the 1024 × 1024 pixel size. The steps described above gave us 500 images from different time-lapse experiments. The image dataset is accessible at the Dryad [31].

The cells in the images were labelled manually by MATLAB (MathWorks Inc., Natick, Massachusetts, USA) as Ground-Truth (GT) single class masks with the dimension of 1024 × 1024 (Fig. 1). The labelled images (512 × 512 pixels) were used as training (80%), testing (20%), and evaluation (20% of the training set) sets in the proposed U-Net networks.

2.3. U-Net model architectures

The U-Net [23] is a semantic segmentation method proposed on the FCN architecture. The FCN consists of a typical encoder–decoder convolutional network. This architecture includes several feature channels to combine shallow and deep features. The deep features are used for positioning, whereas the shallow features are utilized for precise segmentation. The architecture of the simple U-Net was chosen (Fig. 2) for training the model with the specific size of input images.

The first layer of the encoder part consists of the input layer, which accepts RGB images with the size 512 × 512. Each level in the five-"level" U-Net structure includes two 3 × 3 convolutions. Batch normalization follows each convolution, and "LeakyReLu" activation functions follow a rectified linear unit. In the down-sampling (encoder) part (Fig. 2, left part), each "level" in the encoder consists of a 2 × 2 max pooling operation with the stride of two. The max-pooling process extracts the maximal value in the 2 × 2 area. By completing down-sampling in each level of the encoder part, convolutions will double the number of feature channels.

In the up-sampling (decoder) section (Fig. 2, right part), the height and width of the existing feature maps are doubled in each level from bottom to top. Then, the high-resolution deep semantic and shallow features were combined and concatenated with the feature maps from the encoder section. After concatenation, the output feature maps have channels twice the size of the input feature maps. The output decoder layer at the top with a 1 × 1 convolution size predicts the probabilities of pixels. Padding in the convolution process allowed to achieve the same input and output layers size. The computational result, combined with the Binary Focal Loss function, becomes the energy function of the U-Net.

Between each Encoder–Decoder layer in the simple U-Net (Fig. 2), there is a connection combining the down-sampling path with the up-sampling path to achieve the spatial information. Nevertheless, at the same time, this process brings also many irrelevant feature representations from the initial layers. The self-attention U-Net architecture (Fig. 3-*A*) with an impressive performance in medical imaging [32] was applied to prevent this problem and improve semantic segmentation result achieved by standard U-Net. As an extension to the standard U-Net model architecture, the attention gate at the skip connections between encoder and decoder layers highlights the remarkable features and suppresses activations in the irrelevant regions. The advanced function of an attention mechanism is to map a set of key–value pairs and a query to an output. The key, query, values, and outputs are vectors. The compatibility function of the query, together with the corresponding key, is computed to be assigned by weights. Then, weighted sums of the values are computed and generate the output. The weights represent the relative importance of the inputs (the keys) for a particular output (the query) [33]. In this way, the attention gate improves the model sensitivity and performance without requiring complicated heuristics.

The attention gate (Fig. 3-*B*) has two inputs: x^l and g. Input x^l comes from the skip connection from the encoder layers. Since coming from the early layers, input x^l contains better spatial information. Providing x^l is an output from layer l, a feature activation can be formulated as

$$x_i^l = \sigma_1\Big(\sum_{c' \in F_1} x_{c'}^{l-1} \circledast k_{c',c}\Big), \tag{1}$$

by applying a rectified linear unit $\sigma_1(x_{i,c}^l) = \max(0, x_{i,c}^l)$ repeatedly, where i and c correspond to spacial and channel dimensions, respectively, and F_1 denotes the number of feature maps in layer l and $\circledast$ indicates the convolution operation.

Input g – a gating signal – comes from a deeper network layer and contains a better feature representation and contextual information to determining the focus region. Attention coefficients $\alpha \in [0, 1]$ determine, extract, and preserve the valuable features corresponding to the important part of the image regions. The attention part weights different images' parts. This process will add the weights to the pixels based on their relevance in the training steps. The image's relevant parts will get higher weights than the less relevant parts. The output of the attention gate is the multiplication of the input feature maps $x_{i,c}^l$ and the achieved attention coefficient α:

$$\mathrm{p}_{att}^I = \psi^T(\sigma_1(W_x^T x_i^I + W_g^T g_i + b_g)) + b_\psi, \tag{2}$$

$$\alpha_i^I = \sigma_2(p_{att}^I(x_i^I, g_i; \Theta_{att})), \tag{3}$$

where parameter σ_2 represents the sigmoid activation function and Θ_{att} contains parameters including linear transformations W_x and W_g, function ψ and bias terms b_ψ and b_g [32]. The achieved weights are also trained in the training process and make the trained model more attentive to the relevant regions.

Another architecture used in this study and developed based on the U-Net models (originally for nuclei segmentation [34]) is the Residual U-Net. The simple U-Net architecture was built based on repetitive Convolutional blocks in each level (Fig. 4-*B*). Each of these Convolutional blocks consists of the input, two steps of the convolution operation followed by the activation function and the output. On the other hand, we face the vanishing gradient problem when dealing with very deep convolutional networks. The residual step was applied to update the weights in each convolutional block incrementally and continuously (Fig. 4-*C*) to enhance the U-Net architecture performance by overcoming the vanishing gradient problems.

In the traditional neural networks, each convolutional blocks feed the next blocks. The other problem in a DCNN-based network, such as stacking convolutional layers, is that a deeper structure of these kind of networks will affect generalization ability. To overtake this problem, the skip connections – the residual blocks – improve the network performance, with each layer feeding the next layer and layers about two or three steps apart (Fig. 4–*C*). The Residual and Attention U-Net architecture were connected to build more effective and high-performance models from our datasets and improve segmentation results.

The watershed algorithm based on morphological reconstruction [35] was applied after completion of the semantic segmentation by

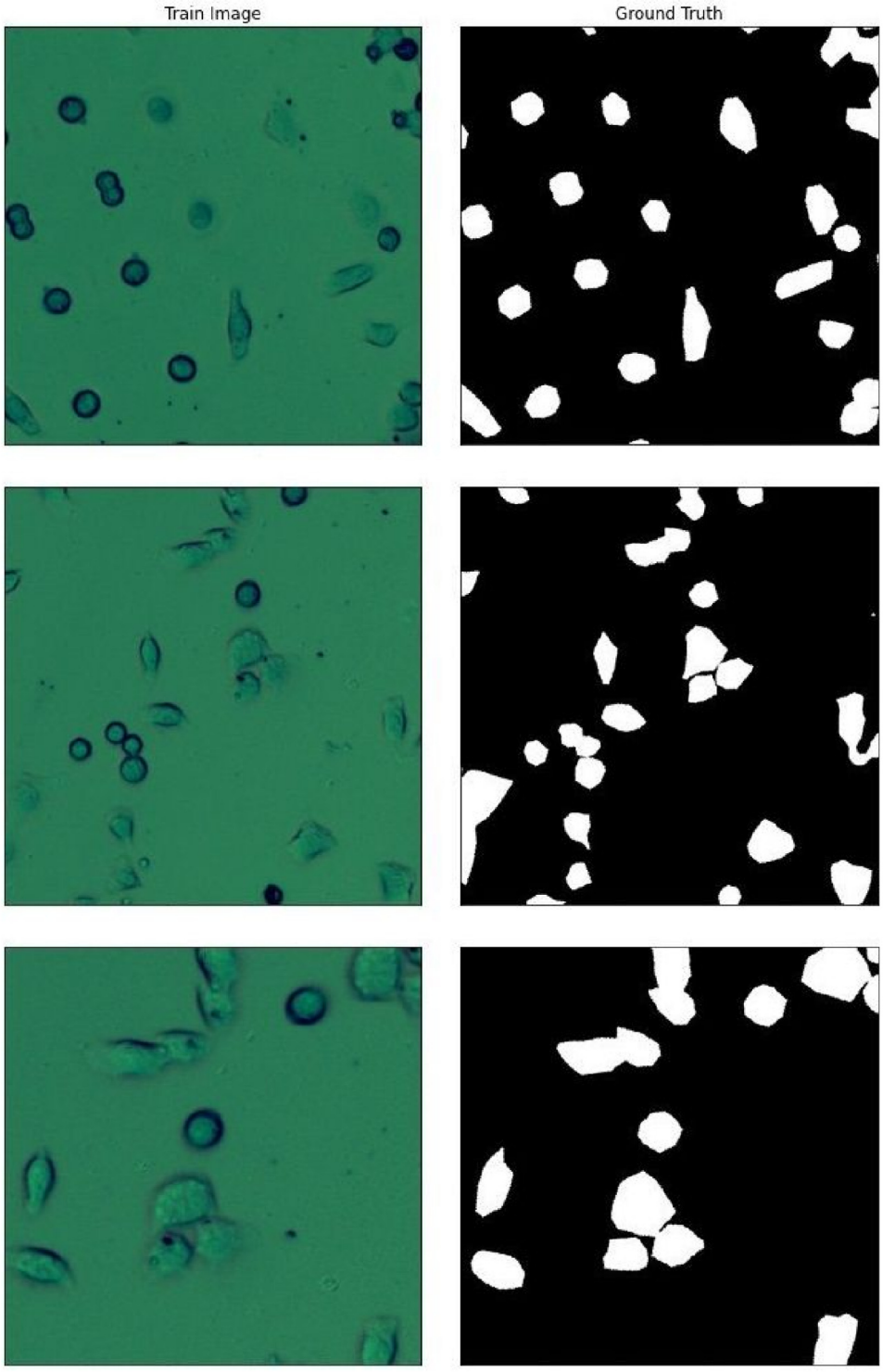

Fig. 1. Examples of the train sets and their ground truths. The image size is 512 × 512.

U-Net methods described above. The U-Net semantic segmentation results were first transformed into a binary image using the Otsu method [36]. After that, the background was determined using ten iterations of binary dilation. The simple Euclidean distance transform defined the foreground of eroded cell regions. The unknown region was achieved by subtraction of the particular foreground region from the background. The watershed method applied to the unknown regions separated the cell borders. The watershed segmentation further helped to solve the over- and under-segmented regions and specify each separated cell by, e.g., cell diameters, solidity, or mean intensity. The

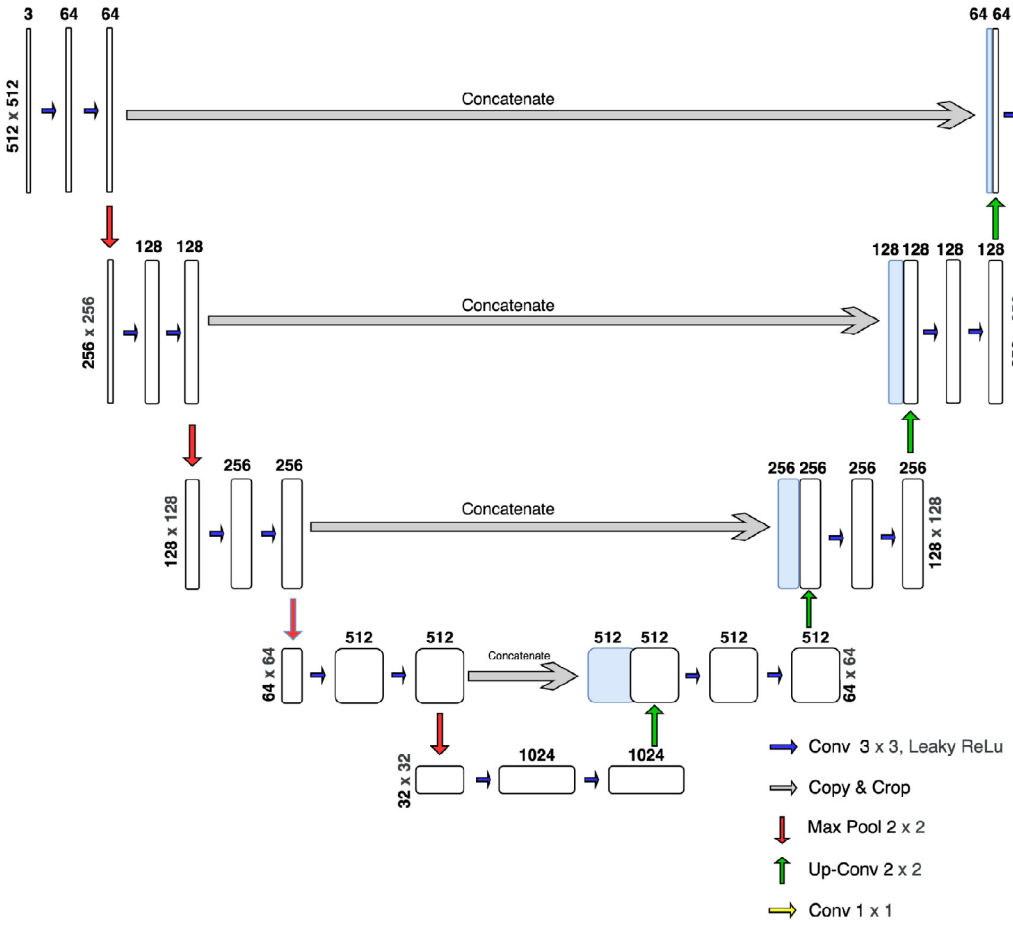

Fig. 2. Architecture of the proposed simple U-Net model.

Table 1
Number of the trainable parameters and the run time for each U-Net model.

Network	Run time	Training parameter
U-Net	3:42':18"	31,402,501
Attention U-Net	4:04':23"	34,334,665
Residual Att U-Net	4:11':24"	39,090,377

segmentation results were optimized using the marked images. Wrongly detected residual connections between different cell regions were cut off, which improved the method accuracy. Fig. 5 presents a general diagram of the proposed U-Net based methods. The U-Net models are hosted on the GitHub [37].

2.4. Training models

The computation was implemented in Python 3.7. The framework for deep learning was Keras, and the backend was Tensorflow [38]. The whole method, including the Deep Learning framework, was transferred and executed on the Google Colab Pro account with P100 and T4 GPU, 24 Gb of RAM, and 2 vCPU [39]. After data preprocessing (Section 2.2), The primary dataset was divided into training (80%) and test (20%). A part (20%) of the training set was used for model validation in the training process to avoid over-fitting and achieve higher performance. Among a 500-image dataset of the mixture of under-, over-, and focused images, 320 images were randomly selected to train the model, and 80 images were chosen randomly to validate the process. The rest of the 100 dataset images were considered for testing and evaluating the model after training.

Before the training, the images were normalized: the pixel values were rescaled in the range from 0 to 1. Since all designed network architectures work with a specific input image size, all datasets were resized to 512 × 512 pixel size. Data augmentation parameters were also applied in training all three U-Net architectures. The optimized values of the hyperparameters used in the training process are written in Table 2. The "rotation range" represents an angle of the random rotation, "width shift range" represents an amplitude of the random horizontal offset, "height shift range" corresponds to an amplitude of the random vertical offset, "shear range" is a degree of the random shear transformation, "zoom range" represents a magnitude of the random scaling of the image. Early stopping hyperparameters were applied to avoid over-fitting during the model training. The patient value was considered as 15. The activation function was set to the LeakyRelu, and the Batch size was set to 8. To optimize the network, we chose the Adam optimizer and set the learning rate to 10^{-3}.

Semantic image segmentation can be considered as a pixel classification as either the cell or background class. The Dice loss was used to compare the segmented cell image with the GT and minimize the difference between them as much as possible in the training process. One of the famous loss functions used for semantic segmentation is the Binary Focal Loss (Eq. (4)) [40]:

$$\text{Focal Loss} = -\alpha_t(1-p_t)^{\gamma}\log(p_t), \tag{4}$$

where $p_t \in [0,1]$ is the model's estimated probability for the GT class with label $y = 1$; a weighting factor $\alpha_t \in [0,1]$ for class 1 and $1-\alpha_t$ for

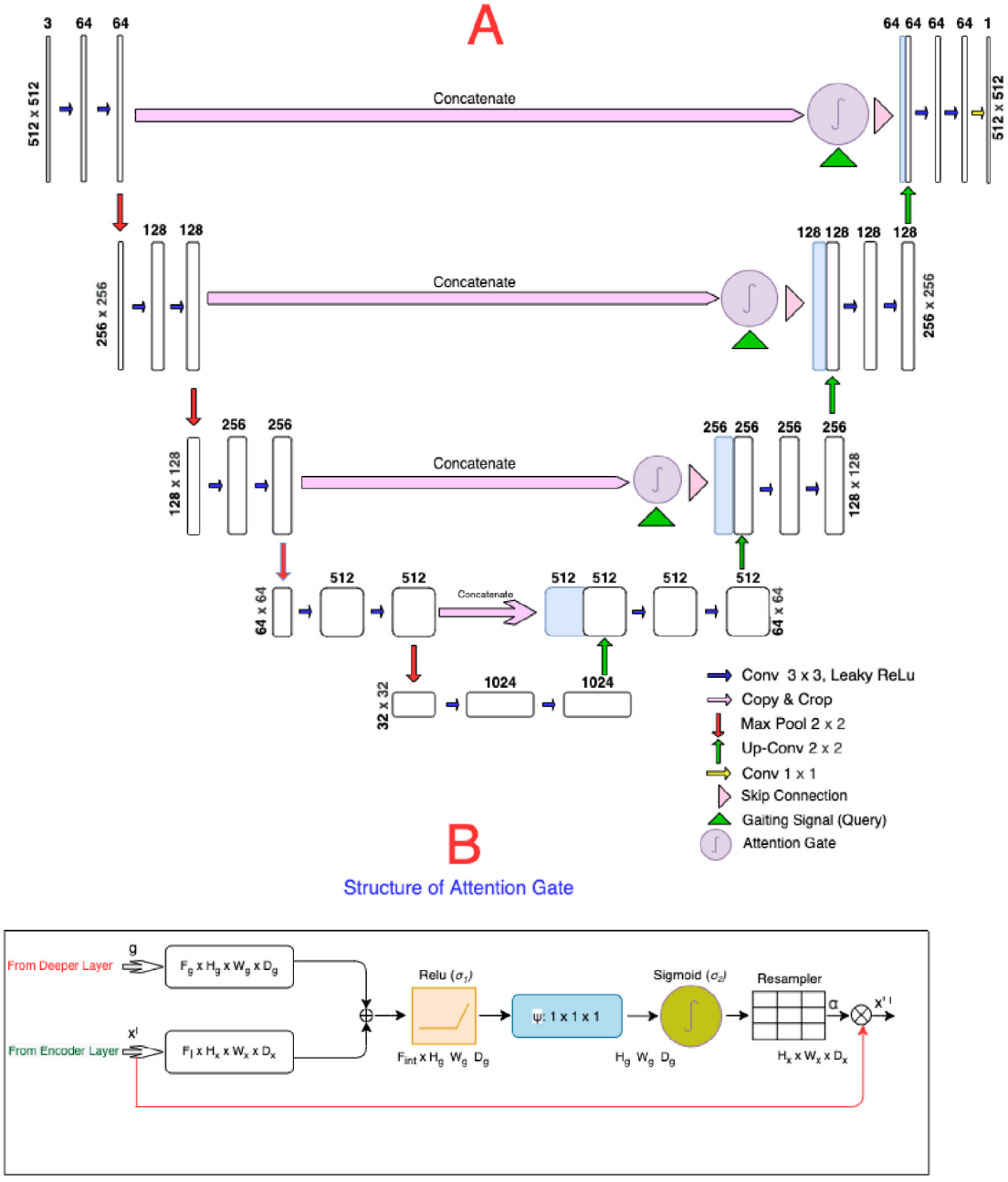

Fig. 3. (*A*) Architecture of the proposed Attention U-Net model, (*B*) the attentive module mechanism. The size of each feature map is shown in $H \times W \times D$, where H, W, and D indicate height, width, and number of channels, respectively.

class -1; $\gamma \geq 0$ is a tunable focusing parameter. The focal loss can be enhanced by the contribution of hardly segmented regions (e.g., cells with vanished borders) and distinguish parts between the background and the cells with unclear borders. The second benefit of the focal loss is that it controls and limits the contribution of the easily segmented pixel regions (e.g., sharp and apparent cells) in the image at the loss of the model. In the final step, updating the gradient direction is under the control of the model algorithm, dependent on the loss of the model.

2.5. Evaluation metrics

The proposed semantic segmentation models were evaluated by different metrics (Eqs. (5)–(9)), where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative metrics, respectively [41]. The metrics were computed for all test sets and explained as mean values (Table 3).

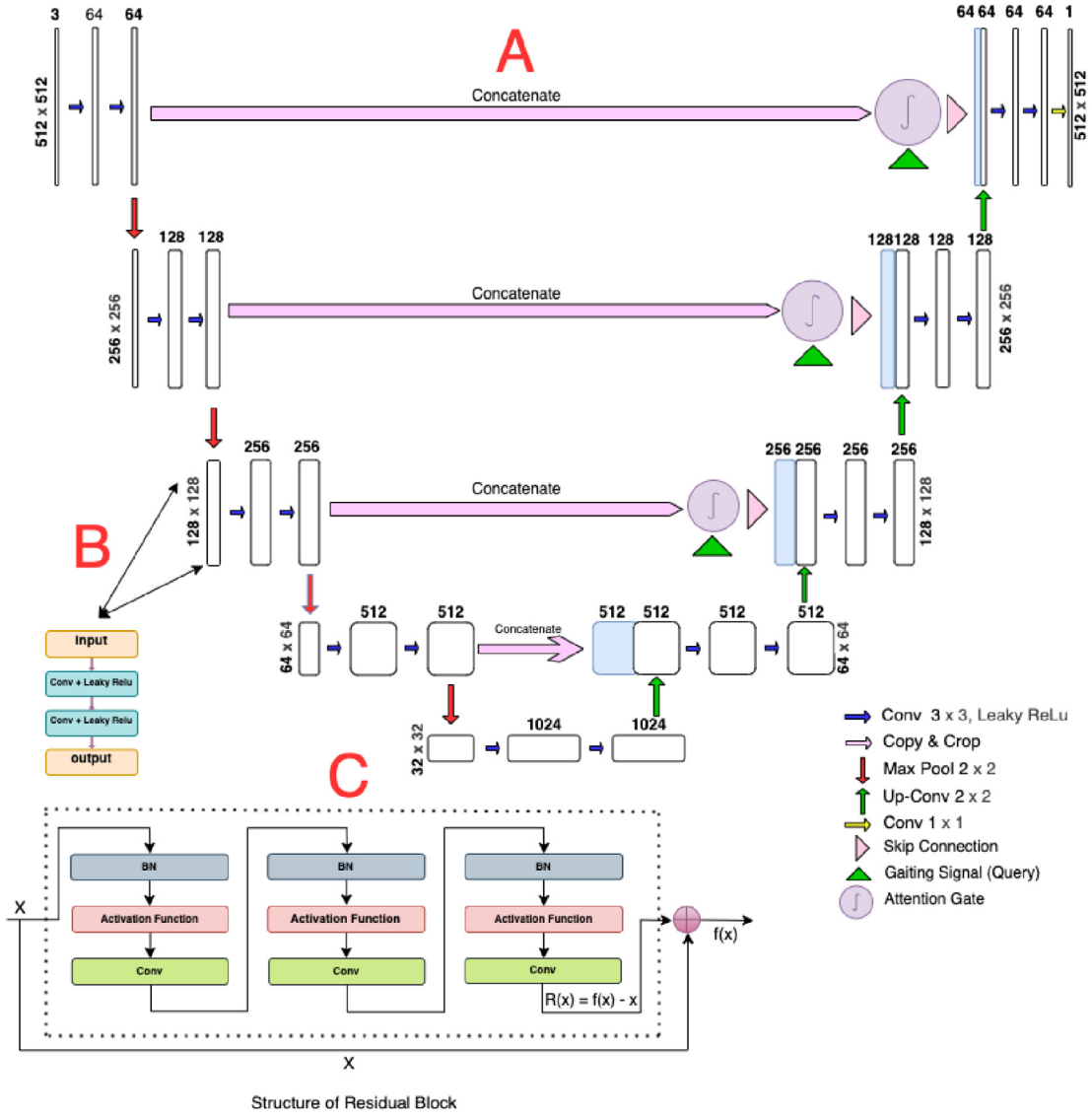

Fig. 4. (*A*) Architecture of the Residual Attention U-Net model. (*B*) Each U-Net layer structure. (*C*) The sample of residual block progress. *BN* refers to Batch Normalization.

Table 2
Hyperparameters setting for all three U-Net models.

Parameter name	Value
Activation function	LeakyRelu
Learning rate	10^{-3}
Batch size	8
Epochs number	100
Early stop	15
Step per epoch	100
Rotation range	90
Width shift range	0.3
Height shift range	0.3
Shear range	0.5
Zoom range	0.3

Overall pixel accuracy (Acc) represents a per cent of image pixels belonging to the correctly segmented cells. Precision (Pre) is a proportion of the cell pixels in the segmentation results that match the GT. The Recall (Recl) represents the proportion of cell pixels in the GT correctly identified through the segmentation process. The F1-score or Dice similarity coefficient states how the predicted segmented region matches the GT in location and level of details and considers each class's false alarm and missed value. This metric determines the accuracy of the segmentation boundaries [42] and have a higher priority than the Acc. Another essential evaluation metric for semantic image segmentation is the Jaccard similarity index known as Intersection over Union (IoU). This metric is a correlation among the prediction and GT [21,43], and represents the overlap and union area ratio for the predicted and GT segmentation.

$$\text{Acc} = \frac{\text{Correctly Predicted Pixels}}{\text{Total Number of Image Pixels}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{5}$$

$$\text{Pre} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Predicted Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recl} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Actual Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Dice} = \frac{2 \times \text{Pre} \times \text{Recl}}{\text{Pre} + \text{Recl}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{8}$$

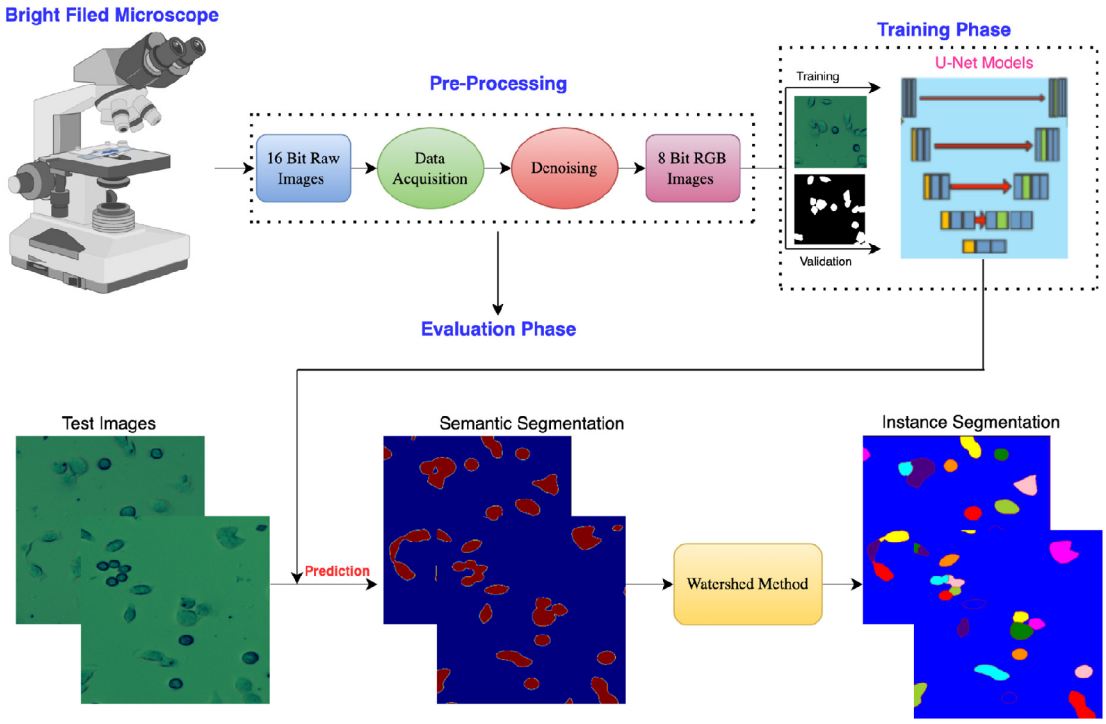

Fig. 5. Flowchart of methodology applied in this study.

$$\mathrm{IoU} = \frac{|\, y_t \cap y_p \,|}{|\, y_t \,| + |\, y_p \,| - |\, y_t \cap y_p \,|} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{9}$$

3. Results

All three models were well trained and converged after running 100 epochs based on training/validation loss and Jaccard plots per epochs (Fig. 6). The hyperparameter values listed in Table 2 were selected to tune for the best training performance and stability. Then, the test datasets were used to evaluating the achieved models. All trained models were assessed (Table 3) using the metrics in Eqs. (5) and (9).

Training the model with the simple U-Net method took the shortest run time with the lowest trainable number of parameters (Table 1). Compared with the Attention U-Net and Residual Attention U-Net, the run time difference is not huge in terms of increasing trainable parameters. The computational cost also did not increase dramatically compared with the acceptable improvement in the model performance. Fig. 7 presents the segmentation results achieved by three different U-Net models. The simple U-Net segmentation result did not distinguish some vanished cell borders (Fig. 7–*A*, black circle). The Attention U-Net (Fig. 7–*B*) detected cells with the vanish borders more efficiently than the simple U-Net. However, the Attention U-Net segmentation suffers from under-segmentation in some regions (visualized by the yellow circle). The outcome of the Residual Attention U-Net method (Fig. 7–*C*, red circle) achieved more accurate segmentation of the vanished cell borders. The watershed binary segmentation after the Residual Attention U-Net networks separated and identified the cells with the highest performance (Fig. 7).

As seen in Mean-IoU, Mean-Dice, and Accuracy metrics (Table 3), the Attention U-Net model showed better segmentation performance than the simple U-Net model in the same situation. The segmentation results were further slightly improved after applying the residual step into the Attention U-Net.

4. Discussion

The analysis of bright-field microscopy image sequences is challenging due to living cells' complexity and temporal behaviour. We have to face (1) irregular shapes of the cells, (2) very different sizes of the cells, (3) noise blobs and artefacts, and (4) vast sizes of the time-lapse datasets. Traditional machine learning methods, including random forests and support vector machines, cannot deal with some of these difficulties in terms of higher computational cost and longer run time for huge time-lapse datasets. The traditional methods suffer from low performance in vanishing and tight cell detection and segmentation and are sensitive to training steps [11,44]. The DL methods have been rapidly developed to overcome these problems. The U-Net is one of the most effective semantic segmentation methods for microscopy and biomedical images [23]. This method is based on the FCN architecture and consists of encoder and decoder parts with many convolution layers.

The image data used to train the Residual Attention model are specific in the way of acquisition. Firstly, the optical path was calibrated to obtain the number of photons that reaches each camera pixel with increasing illumination light intensity. This gave a calibration curve (image pixel intensity vs the number of photons reaching the relevant camera pixel) to correct the digital image pixel intensity. This step ensured homogeneity in digital image intensities to improve the quality of cell segmentation by the neural networks. We work with the low-compressed telecentric transmitted light bright-field high-pixel microscopy images. The bright-field light microscope allows us to observe living cells in their most natural state. Due to the object-sided telecentric objective, the final digital raw image of the observed cells is high-resolved and low-distorted, with no light interference halos around objects.

The procedure compressed the raw colour images to ensure the least information loss at the quarter-pixel-resolution decrease of the image. The final pixel resolution of the images inputting into the neural network is higher (512 × 512) than in the case of any other neural

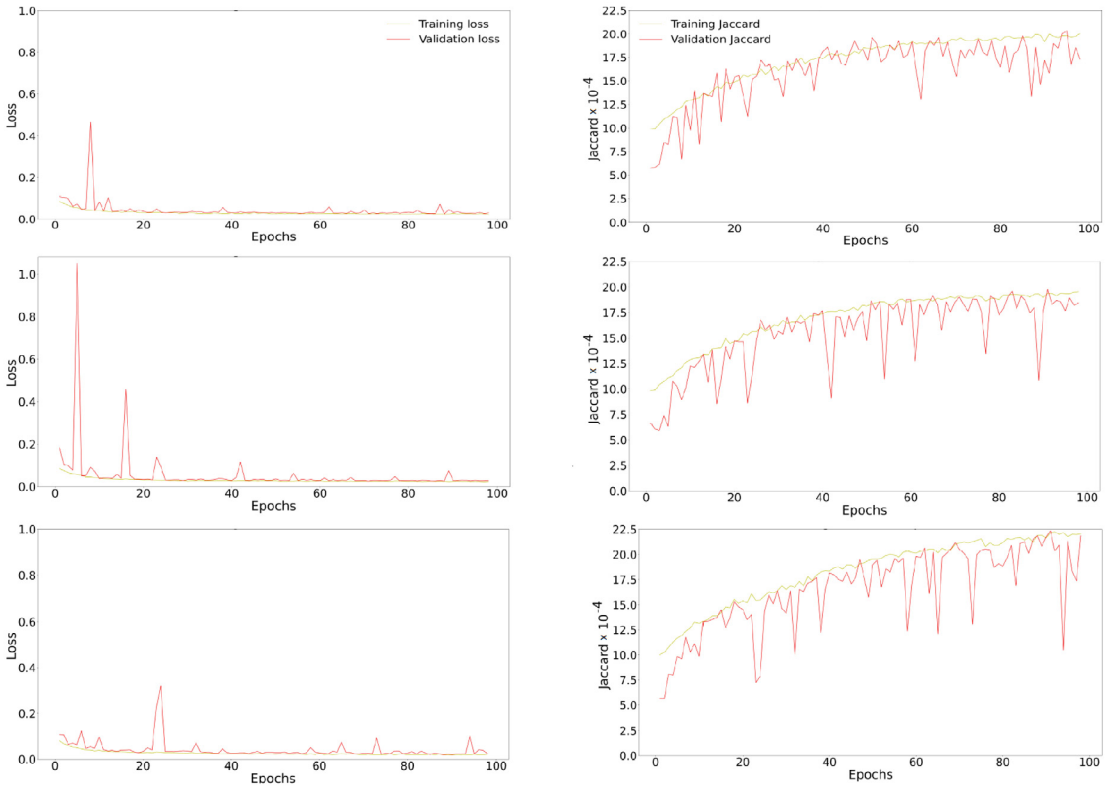

Fig. 6. Training/validation plots for Simple U-Net (left column), Attention U-Net (middle column), and Residual Attention U-Net (right column).

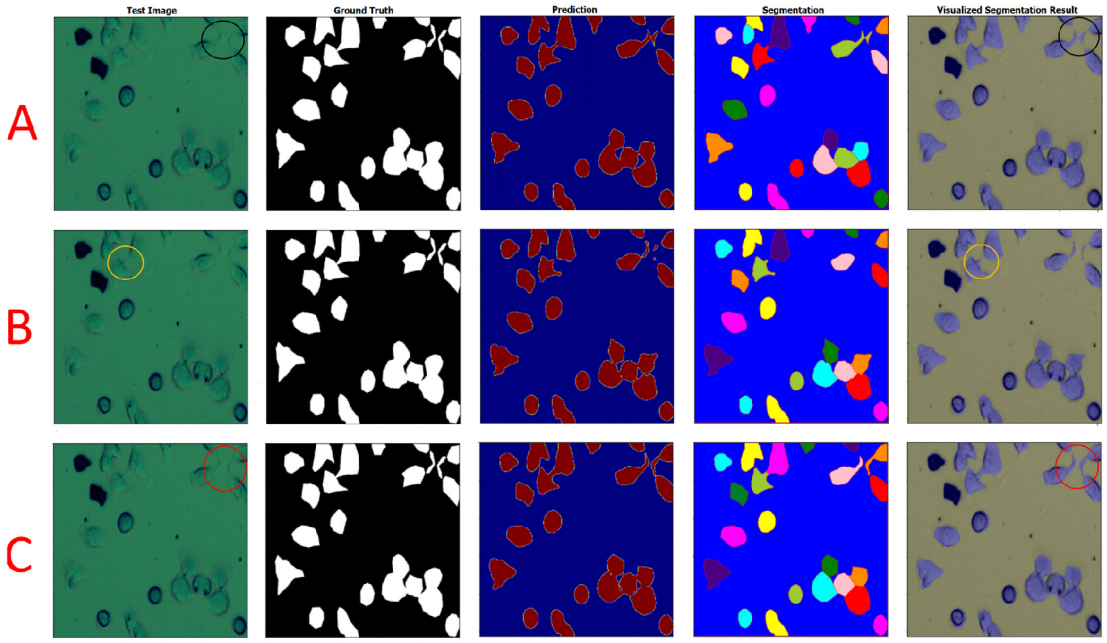

Fig. 7. Segmentation results for (*A*) the simple U-Net (the black circle highlights the non-segmented, vanished cell borders), (*B*) Attention U-Net (the yellow circle highlights the undersegmentation problem), and (*C*) the Residual Attention U-Net (red circle shows the successful segmentation of the cell borders). The image size is 512 × 512.

Table 3
Results for metrics evaluating the U-Net Models. Green values represent the highest segmentation accuracy for the related metric.

Network	Accuracy	Precision	Recall	m-IoU	m-Dice
U-Net	0.957418	0.988269	0.961264	0.950501	0.974481
Attention U-Net	0.959448	0.985663	0.965736	0.952471	0.975511
Residual Att U-Net	0.960010	0.986510	0.965574	0.953085	0.975840

Table 4
Performances of the proposed networks and other networks proposed for microscopy and medical applications. Green highlighted value represent the highest segmentation accuracy in term of mentioned metric.

Models	IoU	Dice	Acc
proposed U-Net	0.9505	0.9744	0.9574
proposed Att U-Net	0.9524	0.9755	0.9594
proposed ResAtt U-Net	0.9530	0.9758	0.9600
U-Net [23]	0.9203	0.9019	0.9554
U-Net [45]	0.7608	-	0.9235
U-Net+ [24]	0.567	-	-
DenseNet [25]	-	0.911	-
SegNet [45]	0.7540	-	0.9225
Attention U-Net [32]	-	0.840	0.9734
Residual Attention U-Net [46]	-	0.9081	0.9557
Residual U-Net [47]	-	0.8366	-
Residual Attention U-Net [48]	-	0.9655	0.9887

network datasets. By preserving high image resolution as much as possible, the demands on the neural network's computational memory and performance parameters were increased.

As our microscope and acquired microscopy data are unique, and were not used before in similar research, it is hard to compare the results with other works. Despite this, the performances of the proposed U-Net-based models were compared with similar microscopy and medical works (Table 4). Our first model was based on a simple U-Net structure and achieved the Mean-IoU score of 0.9505. We assume that better value of the Mean-IoU will be achieved after the hyper-parameter optimization (Table 2). Ronneberger et al. [23] achieved 0.920 and 0.775 Mean-IoU scores for U373 cell line in phase-contrast microscopy and HeLa cell line in Nomarski contrast, respectively. Pan et al. [45] segmented nuclei from medical, pathological MOD datasets with 0.7608 segmentation IoU accuracy score using the U-Net.

We further implemented an attention gate into the U-Net structure (so-called Attention U-Net) to further improve the U-Net model performance by weighing the relevant part of the image pixels containing the target object. In this way, the Mean-IoU metric was improved to 0.9524. The achieved IoU score represents a noticeable improvement in the trained model performance compared with the simple U-Net model. To the best of our knowledge, not many researchers have applied the Attention U-Net to microscopy datasets, but recent papers are prevalently about its application to medical datasets. Microscopy and medical datasets have their complexity and structure, complicating the comparison of the method performances. Applying the Attention U-Net, pancreas [32] and liver tumour [46] medical datasets showed 0.840 and 0.948 Dice metric segmentation accuracy, respectively.

The proposed model performance were improved by one step and obtained the Residual Attention U-Net to overcome the vanishing gradient problem and generalization ability. As a result, the segmentation accuracy was slightly improved by reaching the Mean-IoU of 0.953. The Residual Attention U-Net showed the Dice coefficient of 0.9655 in the testing phase of medical image segmentation [48]. The Recurrent Residual U-Net (R2U-Net) achieved the Dice coefficient of 0.9215 in the testing phase of nuclei segmentation [34]. Patel et al. [47] applied the Residual U-Net to bright-field absorbance image and achieved the Mean-Dice coefficient score of 0.8366. Long et al. [24] applied the enhanced U-Net (U-Net+) to bright-field, dark-field, and fluorescence microscopy images and achieved the Mean-IoU score of 0.567. The U-Net with a dense convolutional network (DenseNet) was applied to detect and segment brain tumour cells [25] with the Dice score of 0.911 and the Jaccard index of 0.839.

5. Conclusion

Microscopy image analysis via deep learning methods can be a convenient solution due to the complexity and variability of this kind of data. This research aimed to detect and segment living human HeLa cells in images acquired using an original custom-made bright-field transmitted light microscope. Three types of deep learning U-Net architectures were involved in this research: the simple U-Net, Attention U-Net, and Residual Attention U-Net. The simple U-Net (Table 1) has the fastest training time. On the other hand, the Residual Attention U-Net architecture achieved the best segmentation performance (Table 3) with a run time slightly higher than the other two U-Net models.

The Attention U-Net is a method to highlight only the relevant activations during the training process. This method can reduce the computational resource waste on irrelevant activations to generate more efficient models. The best segmentation performance was achieved due to the integration of the residual learning structure (to overcome the gradient vanishing) together with the attention gate mechanism (to integrate a low and high-level feature representation) into the U-Net architecture. After extracting semantic segmentation binary results (Table 3), the watershed segmentation method was applied to separate the cells from each other, avoid over-segmentation, label the cells individually, and extract vital information about the cells (e.g., the total number of the segmented cells, cell equivalent diameter, mean intensity and solidity). Nevertheless, future works are still essential to expand the knowledge on multi-class semantic segmentation with different and efficient CNN's architecture and combine the constructed CNN models in the prediction process to achieve the most accurate segmentation result.

Funding

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic – project CENAKVA (LM2018099), by the European Regional Development Fund in frame of the project ImageHeadstart (ATCZ215) in the Interreg V-A Austria–Czech Republic programme, and by the project GAJU, Czech Republic 017/2016/Z.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code availability

The U-Net models are hosted on the GitHub [37] and other data on the Dryad [31].

Acknowledgements

The authors would like to thanks our lab colleagues Šárka Beranová and Pavlína Tláskalová (both from the ICS USB) and Mohammad Mehdi Ziaei for their support of this study.

References

[1] R. Rojas-Moraleda, W. Xiong, N. Halama, K. Breitkopf-Heinlein, S. Dooley, L. Salinas, D.W. Heermann, N.A. Valous, Robust detection and segmentation of cell nuclei in biomedical images based on a computational topology framework, Med. Image Anal. 38 (2017) 90–103, http://dx.doi.org/10.1016/j.media.2017.02.009.

[2] J.R. Tang, N.A. Mat Isa, E.S. Ch'ng, A fuzzy-c-means-clustering approach: Quantifying chromatin pattern of non-neoplastic cervical squamous cells, PLoS One 10 (11) (2015) e0142830, http://dx.doi.org/10.1371/journal.pone.0142830.

[3] Z. Wang, A semi-automatic method for robust and efficient identification of neighboring muscle cells, Pattern Recognit. 53 (2016) 300–312, http://dx.doi.org/10.1016/j.patcog.2015.12.009.

[4] G. Fan, J.-W. Zhang, Y. Wu, D.-F. Gao, Adaptive marker-based watershed segmentation approach for T cell fluorescence images, in: 2013 International Conference on Machine Learning and Cybernetics, Vol. 02, 2013, pp. 877–883, http://dx.doi.org/10.1109/ICMLC.2013.6890407.

[5] P.P. Guan, H. Yan, Blood cell image segmentation based on the hough transform and fuzzy curve tracing, in: 2011 International Conference on Machine Learning and Cybernetics, Vol. 4, 2011, pp. 1696–1701, http://dx.doi.org/10.1109/ICMLC.2011.6016961.

[6] X. Zhou, F. Li, J. Yan, S.T. Wong, Novel cell segmentation method and cell phase identification using Markov model, IEEE Trans. Inf. Technol. Biomed. 13 (2) (2009) 152–157, http://dx.doi.org/10.1109/TITB.2008.2007098.

[7] M. Winter, W. Mankowski, E. Wait, E.C.D.L. Hoz, A. Aguinaldo, A.R. Cohen, Separating touching cells using pixel replicated elliptical shape models, IEEE Trans. Med. Imaging 38 (4) (2019) 883–893, http://dx.doi.org/10.1109/TMI.2018.2874104.

[8] F. Buggenthin, C. Marr, M. Schwarzfischer, P.S. Hoppe, O. Hilsenbeck, T. Schroeder, F.J. Theis, An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy, BMC Bioinform. 14 (2013) 297, http://dx.doi.org/10.1186/1471-2105-14-297.

[9] S.J. Russell, Artificial Intelligence: A Modern Approach, third ed., Prentice Hall, 2010.

[10] F. Mualla, S. Schöll, B. Sommerfeldt, A.K. Maier, J. Hornegger, Automatic cell detection in bright- field microscope images using SIFT, random forests, and hierarchical clustering, IEEE Trans. Med. Imaging 32 (12) (2013) 2274–2286, http://dx.doi.org/10.1109/TMI.2013.2280380.

[11] T. Tikkanen, P. Ruusuvuori, L. Latonen, H. Huttunen, Training based cell detection from bright-field microscope images, in: 2015 9th International Symposium on Image and Signal Processing and Analysis, ISPA, 2015, pp. 160–164, http://dx.doi.org/10.1109/ISPA.2015.7306051.

[12] G. Hinton, T. Sejnowski, Unsupervised Learning: Foundations of Neural Computation, in: MIT Press, MIT Press, 1999.

[13] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, S. Steidl, R. Buchholz, J. Hornegger, Unsupervised unstained cell detection by SIFT keypoint clustering and self-labeling algorithm, in: P. Golland, N. Hata, C. Barillot, J. Hornegger, R. Howe (Eds.), Medical Image Computing and Computer-Assisted Intervention, MICCAI 2014, in: Lecture Notes in Computer Science, vol. 8675, Springer International Publishing, Cham, 2014, pp. 377–384, http://dx.doi.org/10.1007/978-3-319-10443-0_48.

[14] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587, http://dx.doi.org/10.1109/CVPR.2014.81.

[15] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, T. Wang, Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning, IEEE Trans. Biomed. 62 (10) (2016) 2421–2433, http://dx.doi.org/10.1109/TBME.2015.2430895.

[16] E. Shibuya, K. Hotta, Cell image segmentation by using feedback and convolutional LSTM, Vis. Comput. (2021) 11, http://dx.doi.org/10.1007/s00371-021-02221-3.

[17] P. Thi Le, T. Pham, Y.-C. Hsu, J.-C. Wang, Convolutional blur attention network for cell nuclei segmentation, Sensors 22 (4) (2022) 1586, http://dx.doi.org/10.3390/s22041586.

[18] N. Kumar, R. Verma, D. Anand, A. Sethi, Multi-Organ Nuclei Segmentation Challenge, Retrieved on 05 May 2021 URL https://monuseg.grandchallenge.org/.

[19] J.C. Caicedo, A. Goodman, K.W. Karhohs, B.A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban, S. Singh, A.E. Carpenter, Broad Bioimage Benchmark Collection, Retrieved on 05 May 2021 URL https://bbbc.broadinstitute.org/BBBC038.

[20] F. Xing, Y. Xie, L. Yang, An automatic learning-based framework for robust nucleus segmentation, IEEE Trans. Med. Imaging 35 (2) (2016) 550–566, http://dx.doi.org/10.1109/TMI.2015.2481436.

[21] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440, http://dx.doi.org/10.1109/CVPR.2015.7298965.

[22] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection in deep learning and data labeling for medical applications, in: G. Carneiro, D. Mateus, L. c Peter (Eds.), Deep Learning and Data Labeling for Medical Applications DLMIA 2016, LABELS 2016, in: Lecture Notes in Computer Science, vol. 10008, Springer, Cham, 2016, pp. 77–85, http://dx.doi.org/10.1007/978-3-319-46976-8_9.

[23] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, in: Lecture Notes in Computer Science, vol. 9321, Springer, Cham, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[24] F. Long, Microscopy cell nuclei segmentation with enhanced U-Net, BMC Bioinform. 21 (8) (2020) http://dx.doi.org/10.1186/s12859-019-3332-1.

[25] S. Bagyaraj, R. Tamilselvi, P.B.M. Gani, D. Sabarinathan, Brain tumour cell segmentation and detection using deep learning networks, IET Image Process. 15 (10) (2021) 2363–2371, http://dx.doi.org/10.1049/ipr2.12219.

[26] W. Liu, A. Rabinovich, A.C. Berg, Parsenet: Looking wider to see better, in: 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, 2016, p. 11.

[27] I.N. Lyapun, B.G. Andryukov, M.P. Bynina, Hela cell culture: Immortal heritage of henrietta lacks, Mol. Genet. Microbiol. Virol. 34 (4) (2019) 195–200, http://dx.doi.org/10.3103/S0891416819040050.

[28] G. Platonova, D. Štys, P. Souček, K. Lonhus, J. Valenta, R. Rychtáriková, Spectroscopic approach to correction and visualization of bright-field light transmission microscopy biological data, Photonics 8 (8) (2021) 333, http://dx.doi.org/10.3390/photonics8080333.

[29] D. Štys, T. Náhlík, P. Macháček, R. Rychtáriková, M. Saberioon, Least information loss (LIL) conversion of digital images and lessons learned for scientific image inspection, in: F. Ortuno, I. Rojas (Eds.), Bioinformatics and Biomedical Engineering, IWBBIO 2016, in: Lecture Notes in Computer Science, vol. 9656, Springer, Cham, 2016, pp. 527–536, http://dx.doi.org/10.1007/978-3-319-31744-1_47.

[30] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, CVPR'05, 2005, pp. 60–65, http://dx.doi.org/10.1109/CVPR.2005.38.

[31] A. Ghaznavi, R. Rychtáriková, M. Saberioon, D. Štys, Telecentric bright-field transmitted light microscopic dataset, 2022, http://dx.doi.org/10.5061/dryad.80gb5mksp.

[32] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning where to look for the pancreas, in: 1st Conference on Medical Imaging with Deep Learning, MIDL 2018, 2018.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017, pp. 5998–6008, URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[34] M.Z. Alom, C. Yakopcic, T.M. Taha, V.K. Asari, Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net), in: IEEE National Aerospace and Electronics Conference, NAECON 2018, 2018, pp. 228–233, http://dx.doi.org/10.1109/NAECON.2018.8556686.

[35] W. Zhang, D. Jiang, The marker-based watershed segmentation algorithm of ore image, in: 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011, pp. 472–474, http://dx.doi.org/10.1109/ICCSN.2011.6014611.

[36] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66, http://dx.doi.org/10.1109/TSMC.1979.4310076.

[37] A. Ghaznavi, Github repository, 2022, URL https://github.com/AliGhaznavi1986/U-Net-Networks-for-Segmentation.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, in: OSDI'16: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, 2016, pp. 265–283.

[39] Google, System Spec, Retrieved on 12 December 2021 URL https://research.google.com/colaboratory/faq.html.

[40] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 318–327, http://dx.doi.org/10.1109/TPAMI.2018.2858826.

[41] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, Y. Fan, Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks, Neurocomputing 229 (2017) 88–99, http://dx.doi.org/10.1016/j.neucom.2016.08.103.

[42] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure for semantic segmentation? in: Proceedings of the British Machine Vision Conference, BMVA Press, 2013, pp. 32.1–32.11, http://dx.doi.org/10.5244/C.27.32.

[43] B. Vijay, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2015) 228–233, http://dx.doi.org/10.1109/TPAMI.2016.2644615.

[44] C. Sommer, C. Straehle, U. Köthe, F.A. Hamprecht, Ilastik: Interactive learning and segmentation toolkit, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011, pp. 230–233, http://dx.doi.org/10.1109/ISBI.2011.5872394.

[45] X. Pan, L. Li, D. Yang, Y. He, Z. Liu, H. Yang, An accurate nuclei segmentation algorithm in pathological image based on deep semantic network, IEEE Access 7 (2019) 110674–110686, http://dx.doi.org/10.1109/ACCESS.2019.2934486.

[46] Z. Wang, Y. Zou, P.X. Liu, Hybrid dilation and attention residual U-Net for medical image segmentation, Comput. Biol. Med. 134 (2021) 104449, http://dx.doi.org/10.1016/j.compbiomed.2021.104449.

[47] G. Patel, H. Tekchandani, S. Verma, Cellular segmentation of bright-field absorbance images using residual U-Net, in: 2019 International Conference on Advances in Computing, Communication and Control, ICAC3, 2019, pp. 1–5, http://dx.doi.org/10.1109/ICAC347590.2019.9036737.

[48] J. Qiangguo, M. Zhaopeng, S. Changming, C. Hui, S. Ran, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, Front. Bioeng. Biotechnol. 8 (2020) 1471, http://dx.doi.org/10.3389/fbioe.2020.605132.

Paper 2

Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images

Version June 2, 2023 submitted to Biomedical Signal Processing and Control, Elsevier

Authors: **Ghaznavi, A.**, Rychtáriková, R., Císař, P., Ziaei, M., and Štys, D.

Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images

Ali Ghaznavi[a], Renata Rychtáriková[a], Petr Císař[a], Mohammadmehdi Ziaei[b], Dalibor Štys[a]

[a]*Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 373 33 Nové Hrady, Czech Republic*
[b]*Faculty of Science, University of South Bohemia, Branišovská 1760, 37005 České Budějovice, Czech Republic*

Abstract

Multi-class segmentation of unlabelled living cells in time-lapse light microscopy images is challenging due to the temporal behaviour and changes in cell life cycles and the complexity of images of this kind. The deep learning-based methods achieved promising outcomes and remarkable success in single- and multi-class medical and microscopy image segmentation. The main objective of this study is to develop a hybrid deep learning-based categorical segmentation and classification method for living HeLa cells in reflected light microscopy images. Different hybrid convolution neural networks – a simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net architectures – were proposed and mutually compared to find the most suitable architecture for multi-class segmentation of our datasets.

The inception module in the Inception-U-Net contained kernels with different sizes within the same layer to extract all feature descriptors. The series of residual blocks with the skip connections in each ResNet34-U-Net's level alleviated the gradient vanishing problem and improved the generalisation ability. The m-IoU scores of multi-class segmentation for our datasets reached 0.7062, 0.7178, 0.7907, and 0.8067 for the simple U-Net, VGG19-U-Net, Inception-U-

*Corresponding author:Ali Ghaznavi
Email address: ghaznavi@frov.jcu.cz (Ali Ghaznavi)

Net, and ResNet34-U-Net, respectively. For each class and the mean value across all classes, the most accurate multi-class semantic segmentation was achieved using the ResNet34-U-Net architecture (evaluated as the m-IoU and Dice metrics).

Keywords: `Categorical segmentation, Neural network, Cell detection, Microscopy image segmentation, U-Net, Tissue segmentation, Semantic segmentation, Bright-Field Microscopy cell segmentation, Cell analysis`

1. Introduction

Cell detection and segmentation is a fundamental process in microscopy cell image analysis. This is also a challenging task due to the complexity of these images. On the other hand, the information from the segmented living cells can play an essential role in further analysis, such as observing and estimating cell behaviour, their number and dimensions. Recently developed artificial intelligence (AI) methods have achieved promising outcomes in this field. The segmentation methods for analysing cell cultures can be categorised as machine learning (ML) or deep learning (DL).

1.1. Cell culture segmentation with machine learning methods

The number of cell detection-segmentation ML methods has grown rapidly as a result of the low performance of simple techniques such as threshold-based [1], region-based [2], or morphological approaches [3, 4] when processing such complex images. The ML methods can be further classified as supervised or unsupervised.

The supervised methods generate a mathematical function or a model from the training data to map a new data sample [5]. Trained and optimised parameters using the graph-based Supervised Normalized Cut Segmentation (SNCS) with loosely annotated images separate overlapping and curved cells better than the traditional image processing methods [6]. The Fast Random Forest (FRF)

classification using Trainable WEKA Segmentation outperformed the Decision Table and Naïve Bayes classification methods in sensitivity, accuracy, and F-measure when extracting the Interstitial cells of Cajal networks from 3D confocal microscopy images. However, the method showed higher computational costs due to the FRF's structure [7]. A method combining the Histogram of Oriented Gradients and the Support Vector Machine (SVM) extracted and classified the feature descriptors as cells or non-cells in bright-field microscopy data. The method was susceptible to the number of iterations in the training process, which is a crucial step to eliminate false positive detections [8]. A Logistic Regression classification with intensity values of 25 focal planes as features, followed by the binary erosion with a large circular structuring element, counted the cells in bright-field microscopy images. However, the method showed miss-segmentation and a low recall rate [9].

The unsupervised ML algorithms require no pre-assigned labels or scores for the training data [10]. Unsupervised segmentation using the Markov Random Field considered an image as a series of planes based on Bit Plane Slicing. The planes were used as initial labelling for an ensemble of segmentations. The robust cell segmentation was achieved with pixel-wise voting. However, this method was too sensitive to the confidence threshold [11]. A combination of a Scale-Invariant Feature Transform, a self-labelling, and two clustering methods segmented unstained cells in bright-field micrographs. The method was fast and accurate but sensitive to the feature selection to avoid overfitting [12]. A self-supervised (i.e., a kind of unsupervised) learning approach combined unsupervised initial coarse segmentation (K-means clustering) followed by supervised segmentation refinement (SVM pixel classifier) to separate white blood cells. However, the unsupervised part of the method generates a rough segmentation result. In the case of complex datasets, the supervised part of the method cannot work efficiently due to fuzzy boundaries [13].

1.2. Cell culture segmentation with deep learning methods

In recent years, a subset of new machine learning techniques – deep learning (DL) methods – has been developed to solve cell segmentation problems with higher accuracy and performance. The deep neural networks have integrated low/medium/high-level features and classifiers into a comprehensive multi-layer structure. The depth of the network, or the number of layers stacked, determines the "levels" of features [14].

Mask RCNN with a Shape-Aware Loss generated the HeLa cell's segmentation masks with a good performance [15]. A Convolutional Blur Attention (CBA) network consisted of down- and up-sampling procedures for nuclei segmentation in standard challenge datasets [16, 17], with a good value of the aggregated Jaccard index. The reduced number of trainable parameters led to a reasonable decrease in the computational cost [18]. The size of input images of a convolutional network can be of different custom sizes so that it can be trained end-to-end, pixel-to-pixel, and produce an output of the appropriate size. Effective inference and learning can achieve successful semantic segmentation in complex microscopic and medical images [19, 20].

A U-Net architecture containing a contracting path to capture context and a symmetric expanding path for precise localisation showed strong data augmentation in the training process. It was optimised when applied to small datasets and performed efficiently in semantic segmentation of photon microscopy (phase contrast and DIC) images [21]. A Feedback U-Net with the convolutional Long Short-Term Memory network, working on *Drosophila* cell image dataset and mouse cell image dataset, generally showed a low level of accuracy, depending on the segmented class (cytoplasm, cell membrane, mitochondria, synapses) [22]. A Residual Attention U-Net-based method segmented living HeLa cells in bright-field light microscopy data with a high IoU metric. The method combined the self-attention mechanism to highlight the remarkable features and suppress activations in the irrelevant image regions, and the residual mechanism to overcome with vanishing gradient problem [23]. Multi-class cell segmentation in fluorescence images combining U-Net (a deeper network) with ResNet-34 (a

residual mechanism) achieved a good value of IoU score [24]. A two-step U-Net method segmented HeLa cells in microscopy images. The first U-Net localised the position of each cell. The second U-Net was trained with the first U-Net to determine the cell boundaries [25]. A fully automated U-Net-based algorithm recognised different classes (colonies, single, differentiated, and dead) of human pluripotent stem cells from each other with a satisfying m-IoU value in phase contrast images [26].

1.3. Our motivation for a new image segmentation method

In segmentation, especially of tiny cells, the traditional ML methods struggle with microscopy images with complex backgrounds. [8, 7]. The ML methods were also not very efficient in training the multi-class segmentation models in large time-lapse image series. Compared with the ML methods, some Convolution Neural Networks (CNNs) architectures require many manually labelled training datasets and higher computational costs [19]. Deep learning methods have shown better results in segmentation tasks than other methods.

The main goal of our research is to develop and compare variants of a fully convolutional network as the encoder part of the original U-Net architecture and find the most accurate categorical segmentation algorithm. The U-Net was chosen since it is one of the most promising methods for semantic segmentation [21]. Later, the encoder part of the U-Net architecture was modified and replaced with a VGG-19, Inception, and ResNet34 encoder architecture and was examined to find the most suitable architecture for multi-class segmentation. We used unique telecentric bright-field reflected light microscopy multi-class labelled images of the cells to be automatically classified according to their morphological shapes to predict their cell cycle phases.

We captured image series of HeLa cells to test the algorithms. The HeLa is a cell line of human Negroid cervical epithelioid carcinoma that is used in tissue culture laboratories as the gold standard. Each image contains HeLa cells in different cell cycle states. The raw microscopy data is specific for its high pixel resolution in rgb mode and requires pre-processing steps to suppress optical

vignetting and camera noise. The data shows unlabelled in-focused and out-of-focus living cells in their physiological state. Thus, the obtained segmentation method is applicable to observing and predicting cell behaviour in time-lapse experiments during their life cycles and 3D visualisation of the cell.

2. Materials and methods

2.1. Cell preparation and microscope specification

The cells were prepared as written in [23], Section 2.1. Human HeLa cell line (European Collection of Cell Cultures, Cat. No. 93021013) was prepared and cultivated to low optical density overnight at 37°C, 5% CO_2, and 90% relative humidity. The nutrient solution consisted of Dulbecco's modified Eagle medium (87.7%) with high glucose (>1 g L^{-1}), fetal bovine serum (10%), antibiotics and antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all purchased from Biowest, Nuaille, France). The HeLa cells were maintained in a Petri dish with a cover glass bottom and lid at room temperature of 37°C.

The data was collected by running several time-lapse image series experiments of living human HeLa cells on a glass Petri dish using a high-resolved reflected light microscope to observe the microscopic objects and cells. This microscope was designed by the Institute of Complex System (ICS, Nové Hrady, Czech Republic) and built by Optax (Prague, Czech Republic) and ImageCode (Brloh, Czech Republic) in 2021. The microscope has a simple construction of the optical path. The light from a Schott VisiLED S80-25 LED Brightfield Ringlight was reflected from a sample to reach a telecentric measurement objective TO4.5/43.4-48-F-WN (Vision & Control GmbH, Shul, Germany) and an Arducam AR1820HS 1/2.3-inch 10-bit RGB camera with a chip of 4912×3684 pixel resolution. The images were captured as a primary (raw) signal with a theoretical pixel size (size of the object projected onto the camera pixel) of 113 nm. The software (developed by the ICS) controls the capture of the primary signal with a camera exposure of 998 ms. All these experiments were performed in time-lapse to observe cells' behaviour over time.

2.2. Data preparation and pre-processing

Several time-lapse experiments were completed with HeLa cells using a reflected bright-field microscope (Sect. 2.1). The microscope control software calibrated the microscope optical path and corrected all image series using the algorithm proposed in [27] to avoid image background inhomogeneities and noise.

After the calibration step, the raw image representations were converted to 8-bit colour (rgb) images of resolution (number of pixels) quarter of the original raw images. The Bayer mask pixels quadruplets [28] were merged as follows: each pair of green camera filter pixels' intensities were averaged as the green image channel. The red and blue camera filter pixels were adopted into the relevant image channel. Then, images were rescaled to 8 bits after creating the image series intensity histogram and omitting unoccupied intensity levels. This bit reduction ensured the maximal information preservation and mutual comparability of the images through the time-lapse series.

After generating 8-bit images, the denoising method [29] was applied to minimise the background noise in the constructed rgb images at preserving the texture details. Afterwards, the image series from different time-lapse experiments were cropped into the 1024×1024 pixel size to achieve 650 images as the main dataset. The image dataset is accessible at the Dryad data publishing platform [30].

For multi-class segmentation, one of three cell states was assigned to each cell manually using Apeer platform [31]: (1) a background class containing no cells, (2) a cell class containing larger dilated adhered or migrating cells with unclear borders by which we anticipate they are growing, and (3) a cell class including roundish cells with sharper borders when the cells are assumed in their early stage of the life cycle, having no division state yet, or at the beginning of the division. The detection of the ratio of cells in mitosis plays an important role in many biomedical activities, such as biological research and medical diagnosis [32]. Figure 1 depicts a sample of the resized dataset and relevant generated mask classes as ground truth of the size of 512×512 pixels. The labelled images were used as training (80%), testing (20%), and evaluation

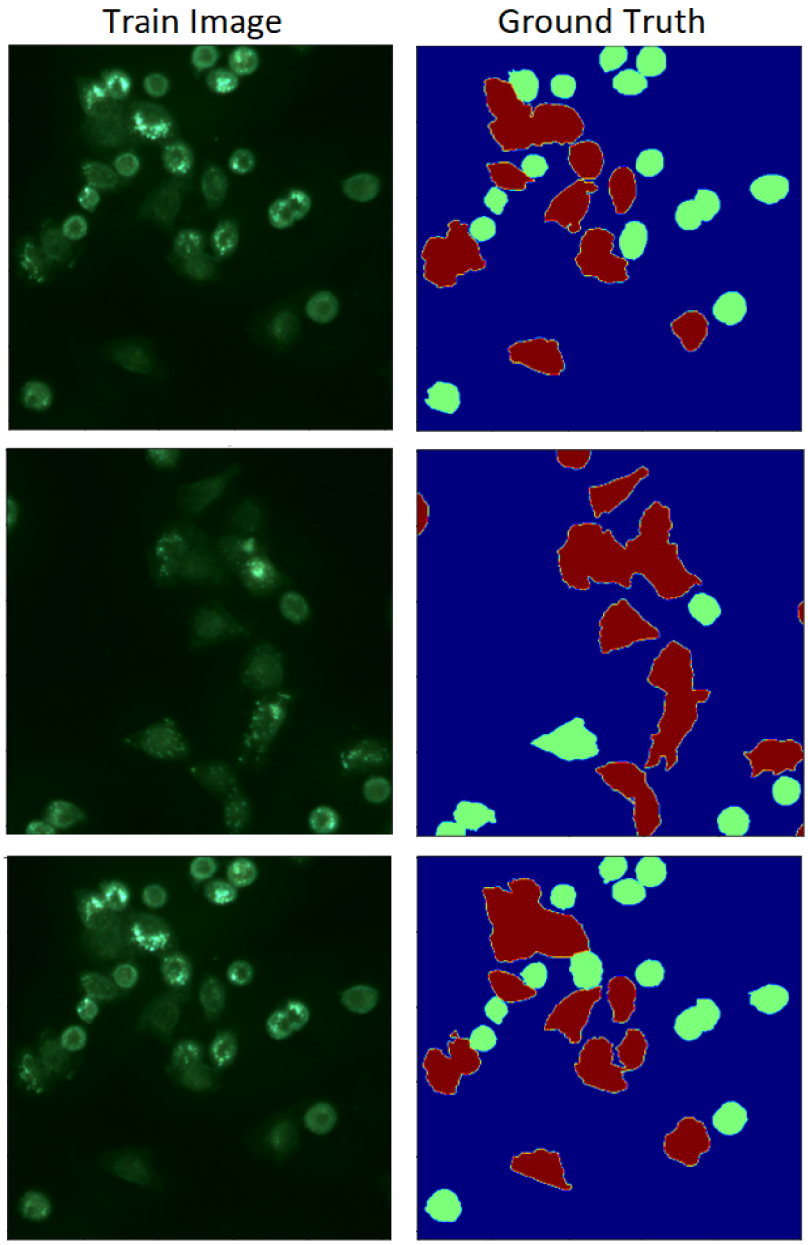

Figure 1: Examples of the train sets and their ground truths. The image size is 512×512. The green and red class represents the roundish sharp cells and the migrating unclear cells, respectively.

(20% of the training set) sets in the proposed neural network architectures.

2.3. The Neural Network Model Architectures

2.3.1. U-Net

The U-Net [21] is well-known as a deep neural network for semantic image segmentation. The U-Net architecture is based on encoder-decoder layers. The U-Net combines many shallow and deep feature channels. In this research, a five-"level" simple U-Net was implemented as the first method for multi-class segmentation purposes. The extracted deep features served for object localisation, whereas the shallow features were used for precise segmentation.

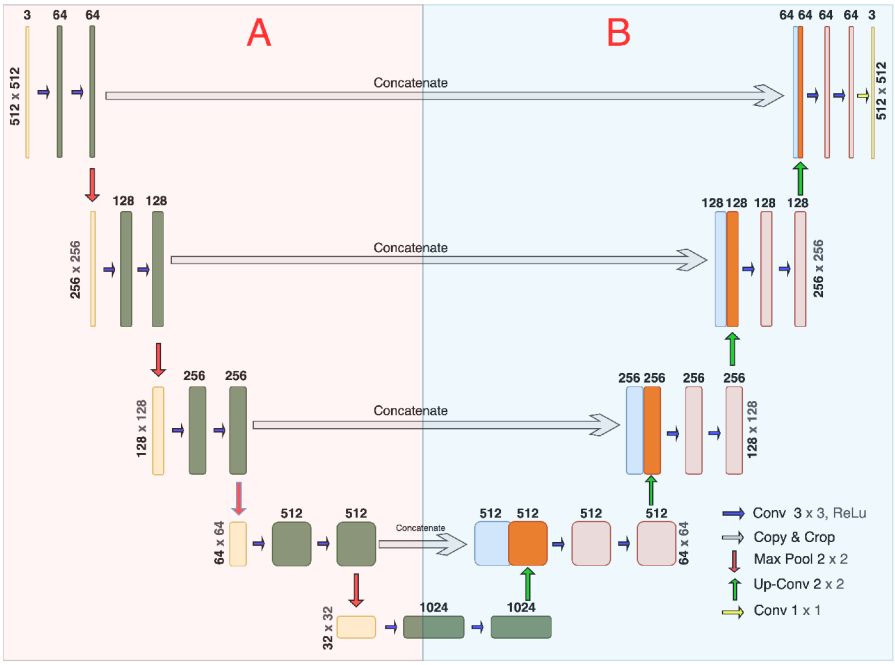

Figure 2: The simple U-Net model architecture. A) The encoder section. B) The decoder section.

The first input layer accepts rgb 512×512-sized training set images. Each level of the proposed U-Net includes two 3×3 convolutions. Batch normalisation follows each convolution, and "ReLU" is used as an activation function. In

the down-sampling (encoder) part (Fig. 2A), each encoder "level" consists of a 2×2 max-pooling operation with a stride of two. The max-pooling process extracts the maximal value in the 2×2 area. By completing the down-sampling in each level of the encoder part, convolutions will double the number of feature channels.

In each level (from bottom to top) of the up-sampling (decoder) section (Fig. 2B), the height and width of the existing feature maps are doubled. In the concatenation step, the high-resolution deep semantic and shallow features were combined with the feature maps from the encoder section. After concatenation, the output feature maps have channels twice the size of the input feature maps. The "softmax" activation function in the top, 1×1 convolution-sized, output decoder layer predicts the occurrence of each pixel in each of the three classes. Padding in the convolution process allowed us to achieve the same input and output layers size. Each of those classes, achieved by the softmax activation, represents the probability of belonging each pixel into each class. In the final step, the "argmax" operation assigned each pixel to the class, where the highest probability value was achieved. This computational result, combined with the Categorical Focal Loss function, becomes the energy function of the U-Net.

2.3.2. The VGG19-U-Net

Many modified artificial neural networks, such as AlexNet [33], ZFNet [14], and VGG [34], have been developed as hybrids with the U-Net to simplify U-Net. In this study, a VGG-Net architecture replaced the U-Net encoder path. In this way, we combined two powerful architectures to improve the categorical segmentation of our unique microscopy dataset. The VGG-Net was proposed by Simonyan and Zisserman [34] from Oxford's Visual Geometry Group (VGG). A VGG-16 proved to be one of the most efficient classification networks. However, a VGG-19 performed even more effectively than VGG-16 [35]. The VGG-19 comprises a network with a deeper topology and smaller convolution kernels to simulate a perceptual field of view. This architecture is designed to reduce the number of trainable parameters and decrease computational costs compared

with the simple U-Net. Figure 3 represents the VGG19-U-Net proposed in this study. The left side of the network (Fig. 3A) shows the architecture of the VGG-19 encoder section with 16 convolution layers, three fully connected layers, and 5 MaxPool layers in 5 blocks. The convolution blocks at each level are followed by a 2×2 max-pooling operation with the stride of two to extract the maximal value in the 2×2 area. The first layer of the VGG network has 64 channels, and each subsequent layer is doubled up to 512 channels. The right side of the network (Fig. 3B) is a schema of the decoder part with five blocks. A concatenation step between each VGG-19 encoder layer and each U-Net decoder layer (Fig. 3) combines the feature maps from the encoder part with the high-resolution deep semantic and shallow features from the decoder part. The last decoder layer has a convolution size of 1×1 and predicts the probability values for each pixel and each of the three classes using the "softmax" activation function.

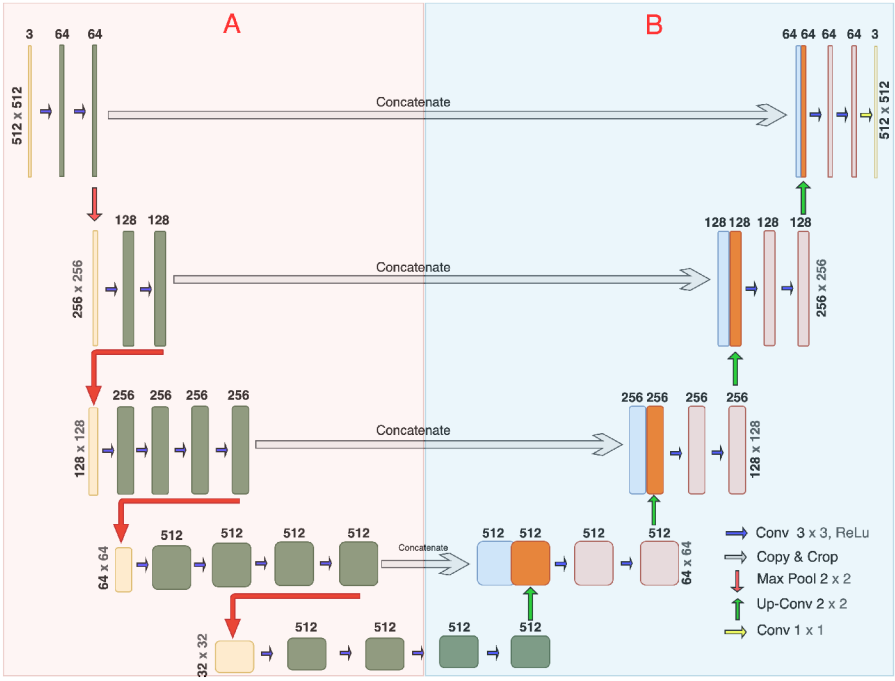

Figure 3: The hybrid VGG19-U-Net architecture. A) The VGG-19 encoder part. B) The U-Net decoder part.

2.3.3. The Inception-U-Net

The complexity of the U-Net network about the number of trainable parameters leads to higher runtime and computational costs (Tab. 4). On the other hand, in image analysis, applying fixed kernel size in all convolution layers can make it difficult to extract all feature descriptors of different sizes. For example, in microscopy image analysis, some (tiny) features are at the local level, and some (larger) are at the global level. The network cannot extract the representative features for big objects when the small kernel is selected in convolution operations. If the kernel size is big, the network will miss extracting the features representative at the pixel level. In other words, the larger kernel can extract a global feature representation over a large image area, and the smaller kernel has been considered for detecting area-specific features. Google's inception deep learning method [36], known as the Inception architecture, was selected to build a hybrid Inception-U-Net architecture (Fig. 4) to improve segmentation results in our datasets further.

The inception module is well known for its computational efficiency by integrating different sizes of convolutions. The inception module applies kernels of different sizes within the same architecture layer and becomes wider (instead of deeper) with the layers (Fig. 4B). The convolution layers were replaced with an inception module (Fig. 4A) in all five levels of the encoder and decoder sections of the original U-Net structure. The inception module consists of multiple sets of 3×3 convolutions, 1×1 convolutions, 3×3 max-pooling, and cascaded 3×3 convolutions. The number of filters at each convolution layer was doubled on the encoder side. The size of the output feature map (height and width) was halved on the last encoder layer.

The up-sampling (decoder) architecture section (Fig. 4A, left side) was also equipped with an inception module at each level. The skip connection connected the encoder and decoder parts to produce a finer prediction. The spatial feature maps from the encoder are concatenated with the decoder feature maps. The rectified linear unit (ReLU) was selected as an activation function for each

layer and performed batch normalisation in each inception module. At the last layer, a 1×1 convolution layer together with the "softmax" activation function generated three segmentation classes of the feature maps for the given input image. Each pixel was assigned to one class according to the highest probability value achieved among the classes. The Categorical Focal Loss function has been considered an energy function for this Inception-U-Net.

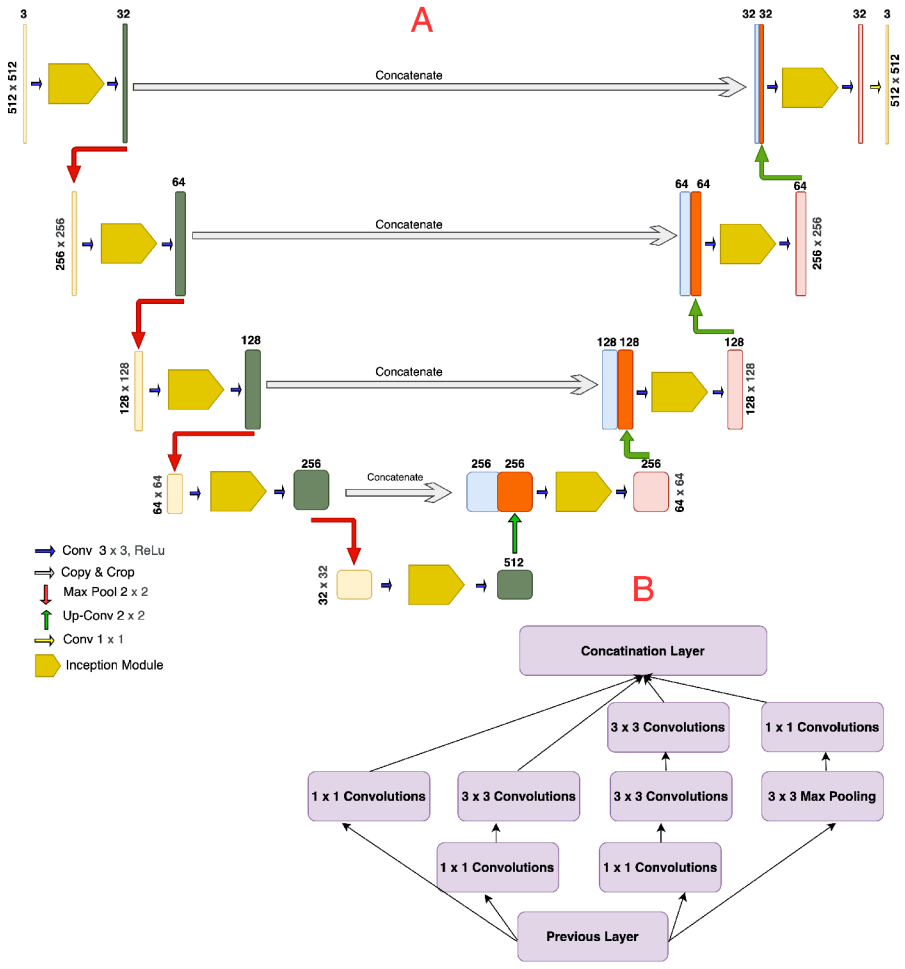

Figure 4: A) The Inception-U-Net architecture. B) The internal architecture of one inception module.

2.3.4. The ResNet34-U-Net

To further improve the categorical segmentation of our datasets, the Residual Convolutional Neural Network (ResNet) [37] was joint to the U-net. Neural networks with deeper architecture are more effective for complex classification and segmentation tasks. However, during the training process, the vanishing gradient problem appears in the very deep CNN. Moreover, a high number of CNN layers makes the training process slower, and the calculated value of the backpropagation derivative becomes increasingly insignificant. Thus, the model's accuracy gets saturated and rapidly declines instead of improving. The series of residual blocks with the skip connections were implemented into the CNN to alleviate the gradient vanishing and improve the network's generalisation ability during the training process. The skip connections were added to the deep neural networks to bypass one or more layers and update the gradient values from one or more previous layers into the following layers.

The ResNet-34-U-Net architecture used in our study (Fig. 5) has 34 layers and four residual convolution steps with a total of 16 residual blocks (red and purple arrows). The first convolution layer has 64 filters with a kernel size of 7×7, followed by a max-pooling layer. Each residual block consists of two 3×3 convolution layers followed by the ReLU activation function and batch normalisation with the identity shortcut connection.

After the first 7×7 convolution layer, the feature map size halved to 256×256. At the first residual level, three residual convolution blocks were applied to the achieved feature maps, and the output size of the feature maps was halved to 128×128. Four residual convolution blocks in the second residual step decreased the size of the output feature maps to 64×64. Six residual convolution blocks in the third residual step gave a feature map size of 32×32. The last residual step consists of three residual convolution blocks to achieve a feature map with a size of 16×16.

The up-sampling section of the network (Fig. 5B) gets the input with the feature map size of 16×16 with 512 channels and a 2×2 up-convolution step with

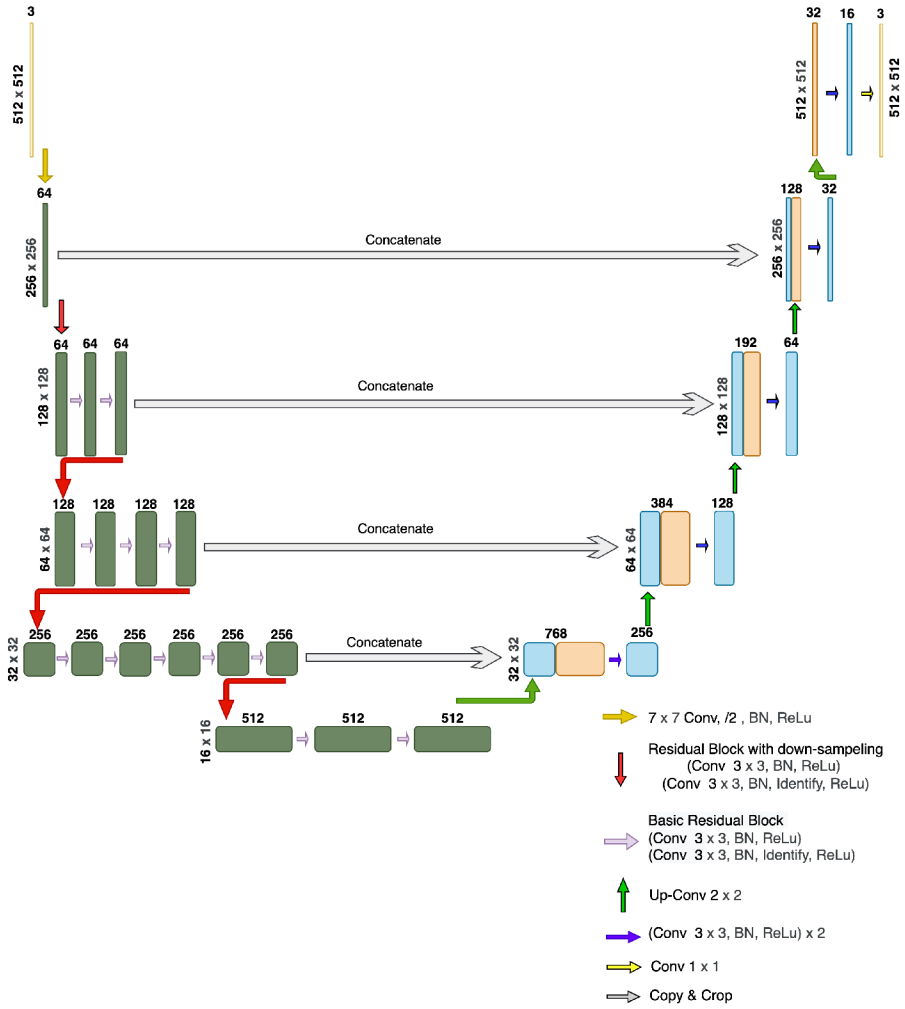

Figure 5: The hybrid ResNet-34-U-Net architecture.

a stride of two. The decoder section has the same structure as the simple U-Net architecture. After passing the U-Net decoder part, the "softmax" activation function was employed to achieve the probability map across three different classes for each pixel of the input images. Afterwards, each pixel was assigned to a certain class according to the highest probability value selected by the "argmax" function.

With the usage of the ResNet-34, the number of trainable parameters de-

creased significantly compared with the VGG-Net and the simple U-Net. Thus, the runtime for training the model was shortened.

2.4. Training Models

The implementation platform for this research was based on Python 3.9. The deep learning framework was Keras with the backend of Tensorflow [38]. All CNN architectures were first developed and completed on a personal computer and then transferred to the Google Colab Pro+ premium cluster account to train the most stable models. The Google Colab Pro+ cluster is equipped with an NVIDIA Tesla T4 or the NVIDIA Tesla P100 GPU with 16 GB of GPU VRAM, 52 GB of RAM, and two vCPUs [39].

The basic dataset included 650 images from different time-lapse experiments and consisted of under-, over-, and focused images. As a trainset, 416 images (64%) were randomly selected to train the model, and 104 images (16%) were chosen randomly to validate the process to avoid over-fitting. The rest of the 130 dataset images (20%) were considered for testing and evaluating the model after training.

Table 1: Number of the trainable parameters and the run time for the U-Net models.

Network	**Run time**	**Training parameter**
U-Net	3:33':29"	31,402,639
VGG19-U-Net	1:44':38"	31,172,163
Inception-U-Net	1:05':47"	18,083,535
ResNet34-U-Net	0:56':22"	24,456,444

All images were normalised (see the pre-processing step in Sect. 2.2) and resized to 512×512 pixels suitable for inputting the designed neural networks. The optimised hyperparameter values (Tab. 2) correspond to training the most stable CNN models. The ReLU was selected as the activation function for all architecture. The early stopping hyperparameter was employed to avoid over-fitting during the model training. The patient value was considered 30. The batch size was set to the maximal value of eight due to the complexity of the CNN structures and GPU-VRAM limitation. The Adam algorithm was

chosen to optimise the neural networks. The learning rate was set to 10^{-3} for all proposed CNN models. The suitable number of object classes was set as 3 (Sect. 2.2). The best number-of-steps-per-epoch value equals 52 (achieved after dividing the length of the trainset of value 416 by the batch size of value 8). The number of epochs when all CNN models converged and were well-trained was 200.

Table 2: Hyperparameters setting for training all proposed models.

Hyperparameters name	**Value**
Activation function	ReLU
Learning rate	10^{-3}
Number of classes	3
Batch size	8
Epochs number	200
Early stop	30
Step per epoch	52
γ for loss function	2

Categorical image segmentation is a pixel classification into either one of the cell classes or the background class. During training progress, all segmented cell images were compared to the GT to minimise the difference between these two as much as possible by using the Dice loss. One of the well-known loss functions used for categorical segmentation, which is an extension of the cross entropy loss, is the Categorical Focal Loss [40].

The Categorical Focal Loss is more efficient for the multi-class classification of imbalanced datasets, when some classes are classified easily and others are not. During training progress, the loss function down-weights easy classes and focuses training on hard-to-classify classes. Thus, the focal loss reduces the loss value for "well-classified" examples (e.g., roundish sharp cells) and increases the loss for hard-to-classify objects (e.g., migrated vanish cells) by tuning the right value of the focusing parameter γ in the categorical focal loss function. In summary, the categorical focal loss turns the model's attention towards the difficult-to-classify pixels to achieve more precise classification results.

343 *2.5. Evaluation metrics*

344 All categorical semantic segmentation models were evaluated using the com-
345 mon metrics (Eqs. 1–5). The TP, FP, FN, and TN correspond to the true
346 positive, false positive, false negative, and true negative metric, respectively
347 [41]. The metrics were computed for all test sets in each class and explained as
348 mean values for all classes (Tab. 4).

349 Overall pixel accuracy (Acc) represents a per cent of image pixels belonging
350 to the correctly segmented cells.

$$\text{Acc} = \frac{\text{Pixels Predicted Correctly}}{\text{Total Number of Image Pixels}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{1}$$

351 Precision (Pre) is a proportion of the cell pixels in the segmentation results
352 that match the GT. The Pre, known as a positive predictive value, is a valuable
353 segmentation performance metric due to its sensitivity to over-segmentation.

$$\text{Pre} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Predicted Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

354 The Recall (Recl) represents the proportion of cell pixels in the GT correctly
355 identified through the segmentation process. This metric says what proportion
356 of the objects annotated in the GT was captured as a positive prediction.

$$\text{Recl} = \frac{\text{Correctly Predicted Cell Pixels}}{\text{Total Number of Actual Cell Pixels}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

357 The Pre and Recl together give another important metric–F1 score–to eval-
358 uate the segmentation result. The F1-score or Dice similarity coefficient states
359 how the predicted segmented region matches the GT in location and level of
360 details and considers each class's false alarm and missed value. This metric
361 determines the accuracy of the segmentation boundaries [42] and has a higher
362 priority than the Acc.

$$\text{Dice} = \frac{2 \times \text{Pre} \times \text{Recl}}{\text{Pre} + \text{Recl}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{4}$$

Another essential evaluation metric for semantic image segmentation is the Jaccard similarity index, known as Intersection over Union (IoU). This metric is a correlation among the prediction and GT [19, 43], and represents the overlap and union area ratio for the predicted and GT segmentation.

$$\mathrm{IoU} = \frac{\mid y_t \cap y_p \mid}{\mid y_t \mid + \mid y_p \mid - \mid y_t \cap y_p \mid} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

3. Results

The models were trained well and converged after running 200 epochs (evaluated as training/validation loss and Jaccard criterion vs epochs, Fig. 6). The hyperparameter values listed in Table 2 were used to achieve the best training performance and stability. Then, the performances of the trained models were assessed and evaluated using the test datasets and the metrics in Eqs. 1–5 (Tab. 4).

The computational cost is one of the critical factors in training high-performance models based on the lowest computational resources. The four described methods differ significantly in runtime, the number of trainable parameters, and network structures (Tab. 1). Training the simple U-Net took the longest runtime with the highest number of training parameters. The VGG19-U-Net was trained well in a significantly shorter time due to the network structure; the number of training parameters was slightly lower than in the simple U-Net. The Inception-U-Net runtime was even faster than the previous two methods. This runtime reduction was followed by a further significant decrease in the number of trainable parameters and higher segmentation performance. The last – ResNet34-U-Net method – achieved the shortest computational cost with the best segmentation performance.

Figure 7 presents the segmentation results for the U-Net-based models proposed in this paper. At the same conditions, the simple U-Net achieved a lower categorical segmentation performance than the other models (when the evaluation metrics are compared). The simple U-Net was inefficient in classifying

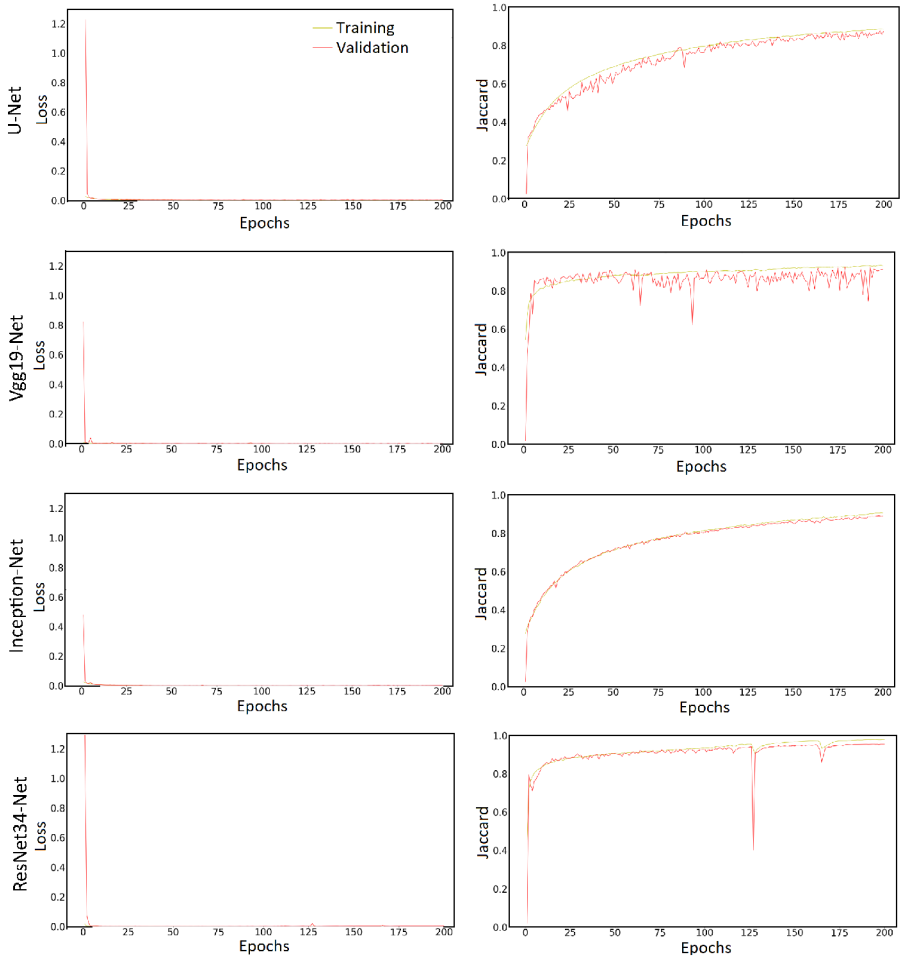

Figure 6: Training/validation plots for the loss criterion (left) and the Jaccard criterion (right) for the simple U-Net (1st row), Vgg19-U-Net (2nd row), Inception-U-Net (3rd row), and ResNet34-U-Net (4th row).

Table 3: m-IoU values for the classes. C1 – background, C2 – divided and unclear cells, C3 – roundish and sharp cells, green – the highest m-IoU value for the relevant class.

Network	**m-IoU C1**	**m-IoU C2**	**m-IoU C3**	**m-IoU**
U-Net	0.9894	0.4839	0.6452	0.7062
VGG19-Net	0.9885	0.5489	0.6160	0.7178
Inception-Net	0.9915	0.6614	0.7194	0.7907
ResNet 34-Net	0.9911	0.6911	0.7378	0.8067

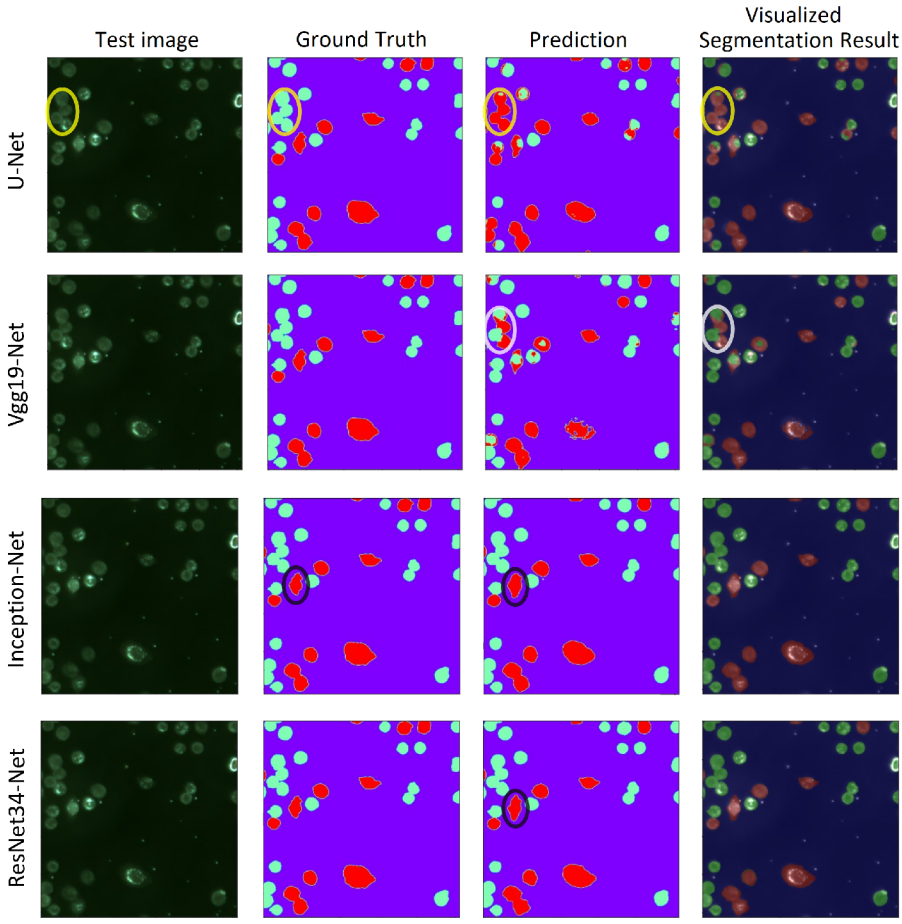

Figure 7: Test image, ground truth, prediction, and 8-bit visualisation of the segmentation results for the U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-Net. The yellow and white circles highlight the wrongly classified and segmented cells. The black circle highlights a different, smoother segmentation result achieved by the ResNet34-U-Net. The image size is 512×512.

the cell pixels into the suitable classes and suffered from wrongly segmented cells into the wrong classes (Fig. 7, yellow circle). Applying the VGG19-U-Net improved the categorical segmentation performance in terms of the evaluation metrics (Tab. 3–4). The cells segmented wrongly by the simple U-Net were improved slightly, but wrong classifications still occurred (Fig. 7, purple circle). The Inception-U-Net was applied to our datasets as the third hybrid CNN method. It leads to significant improvement of the multi-class segmentation results in terms of evaluation metrics (Tab. 3–4). However, this method suffers from over-segmentation in all classes (Fig. 7, black circle). The hybrid ResNet34-U-Net was employed to improve further the object segmentation and classification (Tab. 3–4).

Table 3 shows the mean value of the IoU metric for all combinations of class and method. Achieving a higher IoU value for the class of divided unclear cells (C2) was challenging for all methods. The ResNet34-U-Net achieved the highest m-IoU value in all classes.

Table 4: Results for metrics evaluating the U-Net models. Green values represent the highest segmentation accuracy for the related metric.

Network	**Accuracy**	**Precision**	**Recall**	**m-IoU**	**m-Dice**
U-Net	0.9869	0.7897	0.8833	0.7062	0.8104
VGG19-Net	0.9865	0.8051	0.8614	0.7178	0.8218
Inception-Net	0.9904	0.8684	0.8905	0.7907	0.8762
ResNet 34-Net	0.9909	0.8795	0.8975	0.8067	0.8873

4. Discussion

The light microscope enables observing living cells in their most natural possible states. However, analysing live cell behaviour in an ordinary light transmission (bright-field) microscope over time is difficult for these technical and biological reasons: (1) The cell morphology and position change significantly depending on the life cycle. (2) Illumination conditions are unstable over image and time. (3) The field of view is small to ensure sufficient statistics on cell

behaviour. (4) The images of observed cells are insufficiently spatially resolved and distorted by microscope optics. (5) The traditional image processing methods, including machine learning approaches, were sensitive to the number of iterations in the training process, showed mis-segmentation, low computational and runtime performance and recall rate.

Therefore we enhanced the method described in [23] and developed a microscopic technique with a connecting deep-learning multi-class image segmentation to obviate these complications: (1) Locating the object-sided telecentric objective on the side of the light source (reflection mode) enables us to capture "simple", high-resolved and low-distorted images on a black background (similar to fluorescence images). (2) Calibrating the microscope optical path balanced the intensities in the whole images for following processing by the CNNs. (3) The larger field of view provides a satisfactory number of cells per snapshot for the evaluation of cell behaviour. (4) The images of individual cells were segmented and categorised according to their current physiological state.

One of the most well-known efficient semantic segmentation methods for microscopy and biomedical images is U-Net [21]. The U-Net consists of encoder and decoder parts with many convolution layers. The encoder part of the network was replaced with other different and more effective architecture as the hybrid architecture of the U-Net for more challenging segmentation purposes like categorical segmentation over microscopy images.

The microscope and relevant image data used in this study are unique. No similar research on categorical segmentation of light reflection microscopy data has ever been performed before. Thus, comparing the results achieved in this study with the literature is hard. Despite this, the performances of the proposed hybrid U-Net-based models were compared with similar microscopy and medical works (Tab. 5). The first proposed model was based on a simple U-Net structure and achieved the m-IoU score of 0.7062 as the mean value of all classes for categorical segmentation purposes. We assume that a better value of the m-IoU will be achieved after the hyperparameter optimization (Tab. 2).

Sugimoto et al. [44] achieved a m-Dice score of 0.799 for multi-class segmen-

tation of cancer and non-cancer cells over the medical PD-L1 dataset. Nishimura et al. [45] applied a U-Net-based weakly supervised method on various microscopy datasets and reached a m-Dice segmentation score of 0.618 as an average over all datasets. Piotrowski et al. [26] applied a U-Net-based multi-class segmentation method over human induced pluripotent stem cell images and achieved segmentation IoU and Dice accuracy scores of 0.777 and 0.753, respectively. Long [46] applied the enhanced U-Net (U-Net+) to bright-field, dark-field, and fluorescence microscopy images and achieved the m-IoU score of 0.567 for single class semantic segmentation.

Table 5: Values of the evaluation metrics of the CNNs designed for microscopy and medical applications. Comparison with the literature. Green highlights the highest segmentation accuracy value for each metric.

Models	**IoU**	**Dice**	**Acc**
prop. U-Net	0.7062	0.8104	0.9869
prop. VGG19-U-Net	0.7178	0.8218	0.9865
prop. Inception-U-Net	0.7907	0.8762	0.9904
prop. ResNet34-U-Net	0.8067	0.8873	0.9909
Self-Attention U-Net [44]	-	0.799	-
U-Net [26]	0.777	0.753	-
U-Net [45]	-	0.618	-
U-Net+ [46]	0.567	-	-
VGG16-U-Net [47]	-	-	0.961
VGG19-U-Net [48]	-	0.8715	0.8764
Inception-U-Net [49]	-	0.887	-
Inception-U-Net [24]	-	0.95	-
ResNet34-U-Net [50]	0.6915	-	-
SMANet [51]	0.665	0.769	-
DMMN-M3 [52]	0.706 - 0.870	-	-

The U-Net encoder part was replaced with the VGG-19 architecture to improve the multi-class segmentation result. The final VGG19-U-Net was optimized for our dataset to reduce the number of trainable parameters in the convolution layers and improve the computational costs and segmentation performance using a dipper network topology and a smaller convolution kernel. In this way, the categorical segmentation accuracy increased to 0.7178 for the m-IoU score in the testing phase. Pravitasari et al. [47] applied a VGG16-U-Net

with transfer learning to single-class semantic segmentation of brain tumours in magnetic resonance images and achieved an accuracy of 0.961. Nillmani et al. [48] applied a VGG19-U-Net to X-ray images for single-class segmentation of Covid-19 infections and achieved accuracy and Dice scores of 0.8764 and 0.8715, respectively.

In the next step, we replaced Google's inception architecture for the U-Net encoder and made a hybrid Inception-U-Net network. The inception module contained kernels of various sizes in the same layer to make the network topology wider instead of deeper and extract more representative features. The m-IoU metric for categorical segmentation increased significantly to 0.7907. The number of trainable parameters was reduced. The computational costs were improved efficiently. Haichun et al. [49] proposed an Inception-U-Net for single-class segmentation of brain tumours and achieved the m-Dice score of 0.887 in the testing phase. Sunny et al. [24] applied an Inception-U-Net to categorical segmentation of fluorescence microscopy datasets and achieved the average Dice metric over all segmentation classes of 0.95.

The model performance was further improved using a hybrid ResNet34-U-Net architecture. The series of residual blocks with the skip connection was implemented into the CNN architecture during the training process to overcome the vanishing gradient and generalisation ability in very deep neural networks. It increased the m-IoU to 0.8067 after the multi-class segmentation. Sunny et al. [24] built up a ResNet34-U-Net which showed the m-IoU of 0.6915 in the cross-validation phase of fluorescence microscopy multi-class image segmentation. Gao et al. [51] applied a selected Multi-Scale Attention Network (SMANet) for multi-class segmentation in pancreatic pathological images and achieved m-Dice and m-IoU scores of 0.769 and 0.665. Ho et al. [52] proposed Multi-Encoder Multi-Decoder Multi-Concatenation (DMMN-M3) deep CNN for multi-class segmentation in two different image sets of breast cancer and reached m-IoU of 0.870 and 0.706.

5. Conclusion

The main objective of this research was to develop an efficient algorithm to detect and segment living human HeLa cells and classify them according to their shapes and life cycles stages. Deep learning approaches to reflected light microscopy data analysis delivered efficient and promising outcomes. This research involved variants of hybrid U-Net-based CNN architecture: a simple U-Net, VGG19-U-Net, Inception-U-Net, and ResNet34-U-net.

The simple U-Net (Tab. 1) has the longest training time, the biggest number of trainable parameters, and the lowest categorical segmentation performance. On the other hand, the hybrid ResNet34-U-Net achieved the best categorical segmentation performance (Tab. 4) with a run time significantly lower than the other proposed models. The computational cost and the number of trainable parameters of the inception network are lower than in the U-Net. Thus, the inception networks are better utilisable for bigger datasets. However, running the inception network requires a higher computational GPU memory.

The Residual Convolutional Neural Network (ResNet) was applied as a hybrid with the U-Net to overcome the gradient vanishing and improve the generalisation ability during training. Using a series of residual blocks with skip connection in each level of the ResNet34-U-Net network resulted in better categorical segmentation. The skip connections in each level of the deep neural networks bypass one or more layers and continuously update the gradient values from one or more previous layers into the layers ahead.

The categorical segmentation gradually improves from simple U-Net to ResNet34-U-Net (as evaluated using performance metrics, Tab. 4). The ResNet34 encoder network achieved the best categorical segmentation by integrating the residual learning structure to overcome the gradient vanishing with the U-Net as a hybrid ResNet34-U-Net method. Nevertheless, future works are still essential to expand the knowledge on multi-class semantic segmentation using the weakly supervised method to generate the ground truth for huge datasets independently and apply ensemble learning steps to combine different and efficient CNN ar-

chitectures in prediction to achieve the most accurate segmentation result.

FUNDING

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (project CENAKVA, LM2018099), from the European Regional Development Fund in the frame of the project ImageHeadstart (ATCZ215) in the Interreg V-A Austria–Czech Republic programme, and the project GAJU 114/2022/Z.

DECLARATION OF COMPETITING INTEREST

The authors declare no conflict of interest, or known competing financial interests, or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENT

The authors would like to thank our lab colleagues Šárka Beranová and Pavlína Tláskalová (both from the ICS USB), Jan Procházka (from the USB), and Guillaume Dillenseger (from the FS USB) for their support.

DATA AND CODE AVAILABILITY

The implemented methods and trained models are hosted on the GitHub [53] and other data on the Dryad [30].

References

[1] J. R. Tang, N. A. Mat Isa, E. S. Ch'ng, A fuzzy-c-means-clustering approach: Quantifying chromatin pattern of non-neoplastic cervical squamous cells, PLoS ONE 10 (11) (2015) e0142830. doi:10.1371/journal.pone.0142830.

[2] R. Rojas-Moraleda, W. Xiong, N. Halama, K. Breitkopf-Heinlein, S. Dooley, L. Salinas, D. W. Heermann, N. A. Valous, Robust detection and segmentation of cell nuclei in biomedical images based on a computational topology framework, Med. Image Anal. 38 (2017) 90–103. doi:10.1016/j.media.2017.02.009.

[3] Z. Wang, A semi-automatic method for robust and efficient identification of neighboring muscle cells, Pattern Recogn. 53 (2016) 300–312. doi:10.1016/j.patcog.2015.12.009.

[4] F. Buggenthin, C. Marr, M. Schwarzfischer, P. S. Hoppe, O. Hilsenbeck, T. Schroeder, F. J. Theis, An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy, BMC Bioinform. 14 (2013) 297. doi:10.1186/1471-2105-14-297.

[5] S. J. Russell, Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall, 2010.

[6] X. Huang, C. Li, M. Shen, K. Shirahama, J. Nyffeler, M. Leist, M. Grzegorzek, O. Deussen, Stem cell microscopic image segmentation using supervised normalized cuts, in: IEEE Int. Conf. Image Processing (ICIP), 2016, pp. 4140–4144. doi:10.1109/ICIP.2016.7533139.

[7] S. A. Mah, R. Avci, P. Du, J.-M. Vanderwinden, L. K. Cheng, Supervised machine learning segmentation and quantification of gastric pacemaker cells, in: 42nd Ann. Int. Conf Proc. IEEE Eng. Med. Biol. Soc., 2020, pp. 1408–1411. doi:10.1109/EMBC44109.2020.9176445.

[8] T. Tikkanen, P. Ruusuvuori, L. Latonen, H. Huttunen, Training based cell detection from bright-field microscope images, in: 9th International Symposium on Image and Signal Processing and Analysis (ISPA), 2015, pp. 160–164. doi:10.1109/ISPA.2015.7306051.

[9] K. Liimatainen, P. Ruusuvuori, L. Latonen, H. Huttunen, Supervised method for cell counting from bright field focus stacks, in: Proc. IEEE

13th Int. Symp. Biom. Imaging, 2016, pp. 391–394. doi:10.1109/ISBI.2016.7493290.

[10] G. Hinton, T. Sejnowski, Unsupervised Learning: Foundations of Neural Computation, MIT Press, MIT Press, 1999.

[11] B. Antal, B. Remenyik, A. Hajdu, An unsupervised ensemble-based markov random field approach to microscope cell image segmentation, in: 2013 International Conference on Signal Processing and Multimedia Applications (SIGMAP), 2013, pp. 94–99. doi:10.5220/0004612900940099.

[12] F. Mualla, S. Schöll, B. Sommerfeldt, A. Maier, S. Steidl, R. Buchholz, J. Hornegger, Unsupervised unstained cell detection by SIFT keypoint clustering and self-labeling algorithm, in: P. Golland, N. Hata, C. Barillot, J. Hornegger, R. Howe (Eds.), Med. Image Comput. Comput. Assist. Interv. 2014. Lect. Notes Comput. Sci., Vol. 8675, Springer International Publishing, Cham, 2014, pp. 377–384. doi:10.1007/978-3-319-10443-0_48.

[13] X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, Micron 107 (2018) 55–71. doi:10.1016/j.micron.2018.01.010.

[14] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional neural networks, ECCV: Computer Vision – ECCV 2014, Lect. Notes Comp. Sci. 8689 (2014) 818–833. doi:10.1007/978-3-319-10590-1_53.

[15] S. Lin, N. Norouzi, An effective deep learning framework for cell segmentation in microscopy images, in: Ann. Int. Conf. Proc. IEEE Eng. Med. Biol. Soc., 2021, pp. 3201–3204. doi:10.1109/EMBC46164.2021.9629863.

[16] N. Kumar, R. Verma, D. Anand, A. Sethi, Multi-organ nuclei segmentation challenge, Retrieved on 05/05/2021 (2018).
URL https://monuseg.grandchallenge.org/

[17] J. C. Caicedo, A. Goodman, K. W. Karhohs, B. A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, M. Rohban,

S. Singh, A. E. Carpenter, Broad bioimage benchmark collection, Retrieved on 05/05/2021 (2018).
URL https://bbbc.broadinstitute.org/BBBC038

[18] P. Thi Le, T. Pham, Y.-C. Hsu, J.-C. Wang, Convolutional blur attention network for cell nuclei segmentation, Sensors 22 (4) (2022) 1586. doi:10.3390/s22041586.

[19] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440. doi:10.1109/CVPR.2015.7298965.

[20] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection in deep learning and data labeling for medical applications, in: G. Carneiro, D. Mateus, L. Peter (Eds.), Deep Learning and Data Labeling for Medical Applications DLMIA 2016, LABELS 2016. Lect. Notes Comp. Sci., Vol. 10008, Springer, Cham, 2016, pp. 77–85. doi:10.1007/978-3-319-46976-8_9.

[21] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Med. Image Comput. Comput. Assist. Interv. Lect. Notes Comp. Sci., Vol. 9321, Springer, Cham, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

[22] E. Shibuya, K. Hotta, Cell image segmentation by using feedback and convolutional LSTM, Vis. Comput. 38 (2021) 3791–3801. doi:10.1007/s00371-021-02221-3.

[23] A. Ghaznavi, R. Rychtáriková, M. Saberioon, D. Štys, Cell segmentation from telecentric bright-field transmitted light microscopy images using a residual attention U-net: A case study on HeLa line, Comp. Biol. Med. 147 (147) (2022) 105805. doi:10.1016/j.compbiomed.2022.105805.

[24] S. P. Sunny, A. I. Khan, M. Rangarajan, A. Hariharan, P. Birur N, H. J. Pandya, N. Shah, M. A. Kuriakose, A. Suresh, Oral epithelial cell segmentation from fluorescent multichannel cytology images using deep learning, Comput. Methods Programs Biomed. 227 (2022) 107205. doi:10.1016/j.cmpb.2022.107205.

[25] M. E. Bakir, v. H. Yalim Keles, Deep learning based cell segmentation using cascaded U-Net models, in: 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021, pp. 1–4. doi:10.1109/SIU53274.2021.9477937.

[26] T. Piotrowski, O. Rippel, A. Elanzew, B. Nießing, S. Stucken, S. Jung, N. König, S. Haupt, L. Stappert, O. Brüstle, R. Schmitt, S. Jonas, Deep-learning-based multi-class segmentation for automated, non-invasive routine assessment of human pluripotent stem cell culture status, Comp. Biol. Med. 129 (2021) 104172. doi:10.1016/j.compbiomed.2020.104172.

[27] G. Platonova, D. Štys, P. Souček, K. Lonhus, J. Valenta, R. Rychtáriková, Spectroscopic approach to correction and visualization of bright-field light transmission microscopy biological data, Photonics 8 (8) (2021) 333. doi:10.3390/photonics8080333.

[28] D. Štys, T. Náhlík, P. Macháček, R. Rychtáriková, M. Saberioon, Least information loss (LIL) conversion of digital images and lessons learned for scientific image inspection, in: F. Ortuno, I. Rojas (Eds.), Bioinformatics and Biomedical Engineering. IWBBIO 2016. Lect. Notes Comp. Sci., Vol. 9656, Springer, Cham, 2016, pp. 527–536. doi:10.1007/978-3-319-31744-1_47.

[29] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Vol. 2, 2005, pp. 60–65. doi:10.1109/CVPR.2005.38.

[30] A. Ghaznavi, R. Rychtáriková, P. Císař, M. Ziaei, D. Štys, Telecentric bright-field reflected light microscopic dataset (2023).

URL https://datadryad.org/stash/share/D19v1HCHpsvAos7DPFZaJOAvyMP80ZjiskRruodzwKs

[31] Zeiss, APPEAR – automated image analysis, Retrieved on 12/12/2021. URL https://www.apeer.com/

[32] Y. Lu, A.-A. Liu, Y.-T. Su, Chapter 6 - mitosis detection in biomedical images, in: M. Chen (Ed.), Computer Vision for Microscopy Image Analysis, Computer Vision and Pattern Recognition, Academic Press, 2021, pp. 131–157. doi:10.1016/B978-0-12-814972-0.00006-0.

[33] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90. doi:10.1145/3065386.

[34] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: The 3rd International Conference on Learning Representations (ICLR2015), 2015, pp. 1–14. doi:10.48550/arXiv.1409.1556.

[35] W. A. Hamwi, M. M. Almustafa, Development and integration of VGG and dense transfer-learning systems supported with diverse lung images for discovery of the Coronavirus identity, Inform. Med. Unlocked 32 (2022) 101004. doi:10.1016/j.imu.2022.101004.

[36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9. doi:10.1109/CVPR.2015.7298594.

[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, TensorFlow: large-scale machine learning on heterogeneous distributed systems, in: Proc. USENIX Symp. Oper. Syst. Des. Implement., 2016, pp. 265–283.

[39] Colab, Google Research, Colaboratory, Retrieved on 12/12/2021. URL https://colab.research.google.com/?utm_source=scs-index

[40] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 318–327. doi:10.1109/TPAMI.2018.2858826.

[41] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, Y. Fan, Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks, Neurocomputing 229 (2017) 88–99. doi:10.1016/j.neucom.2016.08.103.

[42] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure for semantic segmentation?, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2013, pp. 32.1–32.11. doi:10.5244/C.27.32.

[43] B. Vijay, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2015) 228–233. doi:10.1109/TPAMI.2016.2644615.

[44] T. Sugimoto, H. Ito, Y. Teramoto, A. Yoshizawa, R. Bise, Multi-class cell detection using modified self-attention, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022, pp. 1854–1862. doi:10.1109/CVPRW56347.2022.00202.

[45] K. Nishimura, C. Wang, K. Watanabe, D. F. E. Ker, R. Bise, Weakly supervised cell instance segmentation under various conditions, Med. Image Anal. 73 (2021) 102182. doi:10.1016/j.media.2021.102182.

[46] F. Long, Microscopy cell nuclei segmentation with enhanced U-Net, BMC Bioinform. 21 (2020) 8. doi:10.1186/s12859-019-3332-1.

[47] A. A. Pravitasari, N. Iriawan, M. Almuhayar, T. Azmi, I. Irhamah, K. Fithriasari, S. W. Purnami, W. Ferriastuti, UNet-VGG16 with transfer learning for mri-based brain tumor segmentation, TELKOMNIKA 18 (3) (2020) 1310–1318. doi:10.12928/telkomnika.v18i3.14753.

[48] Nillmani, N. Sharma, L. Saba, N. N. Khanna, M. K. Kalra, M. M. Fouda, J. S. Suri, Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans, Diagnostics 12 (9) (2022) 2132. doi:10.3390/diagnostics12092132.

[49] H. Li, A. Li, M. Wang, A novel end-to-end brain tumor segmentation method using improved fully convolutional networks, Comp. Biol. Med. 108 (2019) 150–160. doi:10.1016/j.compbiomed.2019.03.014.

[50] G. Patel, H. Tekchandani, S. Verma, Cellular segmentation of bright-field absorbance images using residual U-Net, in: 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 2019, pp. 1–5. doi:10.1109/ICAC347590.2019.9036737.

[51] E. Gao, H. Jiang, Z. Zhou, C. Yang, M. Chen, W. Zhu, F. Shi, X. Chen, J. Zheng, Y. Bian, D. Xiang, Automatic multi-tissue segmentation in pancreatic pathological images with selected multi-scale attention network, Comp. Biol. Medicine 151 (Pt A) (2022) 106228. doi:10.1016/j.compbiomed.2022.106228.

[52] D. J. Ho, D. V. Yarlagadda, T. M. D'Alfonso, M. G. Hanna, A. Grabenstetter, P. Ntiamoah, E. Brogi, L. K. Tan, T. J. Fuchs, Deep Multi-Magnification Networks for multi-class breast cancer image segmentation, Comput. Med. Imaging Graph. 88 (2021) 101866. doi:10.1016/j.compmedimag.2021.101866.

[53] A. Ghaznavi, Github repository (2023).
URL https://github.com/AliGhaznavi1986/Hybrid-CNNs-for-multi-class-segmentation

Paper 3

Comparative Performance Analysis of simple U-Net, Residual Attention U-Net, and VGG16-U-Net for Inventory Inland Water Bodies

Version May 25, 2023 submitted to Remote Sensing, MDPI

Authors: **Ghaznavi, A.**, Saberioon, M., Brom, J., and Itzerott, S.

Comparative Performance Analysis of simple U-Net, Residual Attention U-Net, and VGG16-U-Net for Inventory Inland Water Bodies

Ali Ghaznavi[a,b], Mohammadmehdi Saberioon[b], Jakub Brom[c], Sibylle Itzerott[b]

[a]*Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 373 33 Nové Hrady, Czech Republic*
[b]*Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Telegrafenberg, Potsdam 14473, Germany*
[c]*Department of Applied Ecology, Faculty of Agriculture and Technology, University of South Bohemia in České Budějovice, Studentská 1668, České Budějovice 37005, Czech Republic*

Abstract

Inland water bodies play a vital role at all scales in the terrestrial water balance and Earth's climate variability. Thus, an inventory of inland waters is crucially important for hydrologic and ecological studies and management. Therefore, the main aim of this study was to develop a new method for inventorying and mapping inland water bodies using high-resolution satellite imagery automatically and accurately. Three different deep learning, U-Net-based algorithms were used to segment inland waters, including simple U-Net, Residual Attention U-Net, and VGG16–U–Net. All three algorithms were trained using a combination of Sentinel-2 visible bands (Red [B04; 665nm], Green[B03; 560nm], and Blue[B02; 490 nm]) in 10-meter spatial resolution. VGG16-U-Net provided the best segmentation results with 0.9850 in terms of mean-IoU score, which improved slightly compared to other proposed U-Net base architecture. Although the accuracy of the model based on VGG16-U-Net doesn't make a difference from Residual Attention U-Net, the computation costs for training VGG16-U-Net were dramatically lower than Residual Attention U-Net.

Keywords: `Automated mapping, Deep learning, Land cover, Satellite`

*Corresponding author:Mohammadmehdi Saberioon
Email address: `saberioon@gfz-potsdam.de` (Mohammadmehdi Saberioon)

imagery, Segmentation, Water bodies

1. Introduction

Inland waters (i.e., rivers, streams, lakes, reservoirs, wetlands, and flood plains) significantly impact hydrological and biogeochemical cycles. They play a vital role at all scales in the terrestrial water balance and Earth's climate variability[1, 2]. Furthermore, inland waters provide vital resources for humans and are the sole habitat for an extraordinarily rich, endemic, and sensitive biota. However, like many other ecosystems over the past century, humans' high demands on freshwater, continuous demographic pressure, and climate change have threatened the existence of inland water resources and biodiversity around the world[3]. Consequently, tracking and quantifying human and climate change influence on global inland water is essential, particularly for small water bodies, and delineating them is a prerequisite for further monitoring, modeling, and management.

Since the 1970s, remote sensing techniques have become increasingly popular for detecting and mapping inland waters regionally and globally[4, 5]. Since the launch of Sentinel-2, this trend has increased as Sentinel-2 is continuously acquiring high-resolution images from the land surface. Therefore, the scientific community and public and private sectors have used Sentinel-2 data extensively for land cover/use monitoring, including water bodies detection[6, 7]. Many former studies using methods like spectral indices [8, 9], single band density slicing [10], or supervised classification [11, 12] for detecting and mapping water bodies as water bodies appear dark in optical remote sensing due to high absorbance of irradiance in the near-infrared (NIR) spectrum. However, these methods have limitations, and some times challenging to inventory the inland waters with satisfactory accuracy. For instance, because of variations in the physical environment over space and time, it is often not straightforward to establish a constant threshold value [13]. In water body classification, shadows produced by mountains, trees, buildings, and river banks can contaminate

satellite imagery classification of water bodies [14]. Therefore, a new method is still desirable for detecting and mapping inland waters where high-resolution orbital remote sensing data automatically and accurately.

Deep learning algorithms, particularly deep learning-based semantic segmentation algorithms, are widely used in the classification of remote sensing images [15, 16]. Although recently, several studies have shown that U-Net-based algorithms have better results; for instance, however, Zhang et al. [17] used and compared six different deep learning-based algorithms, including the network using architecture shape like 'U' well known as (U-Net), fully convolutional DenseNet (FC-DenseNet), full-resolution residual network (FRRN), bilateral segmentation network (BiSeNet), DeepLab version 3 plus (DeepLabV3+), and pyramid scene parsing network (PSPNet) for classification of land covers for medium resolution remote sensing data. They have found that the architecture based on encoder–decoder mechanism, including U-Net, is the most competitive network with the appropriate outcome to detect and map land covers of medium-resolution images. An et al. [18] proposed new architecture based on U-net where the convolution layer in U-Net was replaced with a bottleneck structure for water bodies extraction. They found that their proposed architecture can accurately (98.13%) segment water bodies and greatly reduce the size of the model and prediction time.

It is still necessary to continue studying U-Net-based models with different architectures for the segmentation of different scenarios or types of features. Therefore, the main objective of this research was to develop and implement an accurate deep learning segmentation method with reasonable computational cost to detect and segment inland water bodies from high spatial resolution remote sensing images. We choose the U-Net for our research cause it is one of the methods with strong outcomes in semantic segmentation tasks. In addition, two other U-Net architectures, Residual Attention U-Net, and VGG16-U-Net were also investigated to achieve the best architecture for automated inland water detection based on the accuracy and computation cost.

2. Materials and Pre-Processing

2.1. Data preparation and pre-processing

This study acquired the raw images using the sentinel-2 Harmonized dataset archived on the Google Earth Engine javascript platform (GEE). The southern part of the Czech Republic, including the South Bohemian region, was selected as the region of interest (Fig. 1). This part of czech republic were considered to train the model because of the more water bodies in and artificial lakes existing in this region of the country. Including images with more related RoI regions were helpful to train more efficient models to predict the water bodies. Sentinel-2 images acquired during summer 2022 with less than 10% of cloud covering were considered as datasets for training and testing algorithms.

In this study, the combination of visible bands of sentinel-2 (Red [B04; 665nm], Green[B03; 560nm], and Blue[B02; 490 nm]) were considered and used to obtain true color images for segmentation purpose. The reason of considering RGB bands is because the more bands used, the more complex and computationally expensive the segmentation model. In other words, increasing model development and deploy the model requires more time and computation power. Additionally, not all bands may provide useful information for segmenting of water bodies, so it's often more efficient to select a relevant subset of bands. Therefore, using only the RGB bands, which produce true color images, was a reasonable choice, given their sufficiency in achieving good accuracy in segmenting water bodies. Using fewer bands can also help reduce overfitting, which occurs when a model becomes too complex and fits the training data too closely, resulting in poor generalization to new data. By using a simpler model with fewer input features, the risk of overfitting can be reduced and the generalization performance of the segmentation model can be improved.

To achieve RGB images and render the image as a true-color composite, The Earth Engine visualization parameters and specific bands are configured as 'B4'(665nm), 'B3' (560 nm), and 'B2' (490nm) for red, green, and blue color channels with 10-meter spatial resolution, respectively. The "min" and "max"

values in visualization parameters are suitable for displaying reflectance from typical Earth surface targets. The min value was set to zero, the max value was considered equal to 4000, and the Gamma correction factor was set to 1.4. After collecting the raw images from the Google Earth Engine (GEE) javascript platform, Raw images were downloaded and transferred into the QGIS software for further processing.

Figure 1: The map of the study area. The red region represented the area selected for the data collection phase.

After transferring the raw image data into the QGIS, the specific parts of the south bohemian region (Fig 1, The red region) was selected as the main dataset. On the other hand, the labeled data from Czech Republic inland waters provided by ZABAGED [19] were imported into the QGIS to generate the shape file of the inland water for all parts of the Czech Republic. Then, the same specific coordination from the GEE image and the labeled data were exported as "Tiff" file with a big size of $46K \times 46K$ pixel resolution.

In the next step, the image and mask in big size were patchified into smaller parts (Fig 2). That process generated the main dataset for further analysis. The patchifing step splits images into small patches by given patch cell size [20] (ie. like cropping image in big size into the small parts). Images were patchified and

masked into the 2048×2048 pixel resolution to achieve suitable region of interest (ROI) area and avoid pixelating and blurring problems in the smaller size of the images. The patchifying step helped us to convert the image in big size into the images in smaller size to use in training step. After patchifying the image and mask into smaller parts, we achieved 504 images as the main dataset. The main dataset was split into three parts: (1) train set by randomly considering 322 images (80% of the main dataset), (2) test set by randomly considering 101 images (20% of the main dataset), (3) for model validation progress, 20% of the train set randomly selected (81 images) to prevent over-fitting problem during training progress and reach more stable performance for generated models.

2.2. Neural network architecture

2.2.1. Simple U-Net

Deep neural network methods delivered promising outcomes in classification and segmentation tasks in terms of accuracy when dealing with a large dataset. One of the promising neural network architectures for semantic segmentation is U-Net. The U-Net based methods deliver promissing outcome in different sensetive research fields including medical and microscopy regions [21, 22]. The U-Net was proposed and created for semantic segmentation based on the convolutional neural network (CNN) architecture and comprised of an encoder-decoder convolutional network topology. The encoder and decoder blocked in each level were connected to each other via a bridge to combine features from the encoder part with extracted features from the decode section. The feature representation extracted by the decoder part is useful for positioning, whereas encoder part features are efficient in achieving accurate segmentation. The proposed architecture for the simple U-Net method applied in this research is displayed in Fig. 3.

The first layer of the encoder part (fig. 3, Part A) accepts images with the size 512×512 with three color channel (RGB) mode as input. The proposed U-Net structure has five levels. Each level consists of two 3×3 convolutions followed by Batch normalization for each convolution layer and applying a rec-

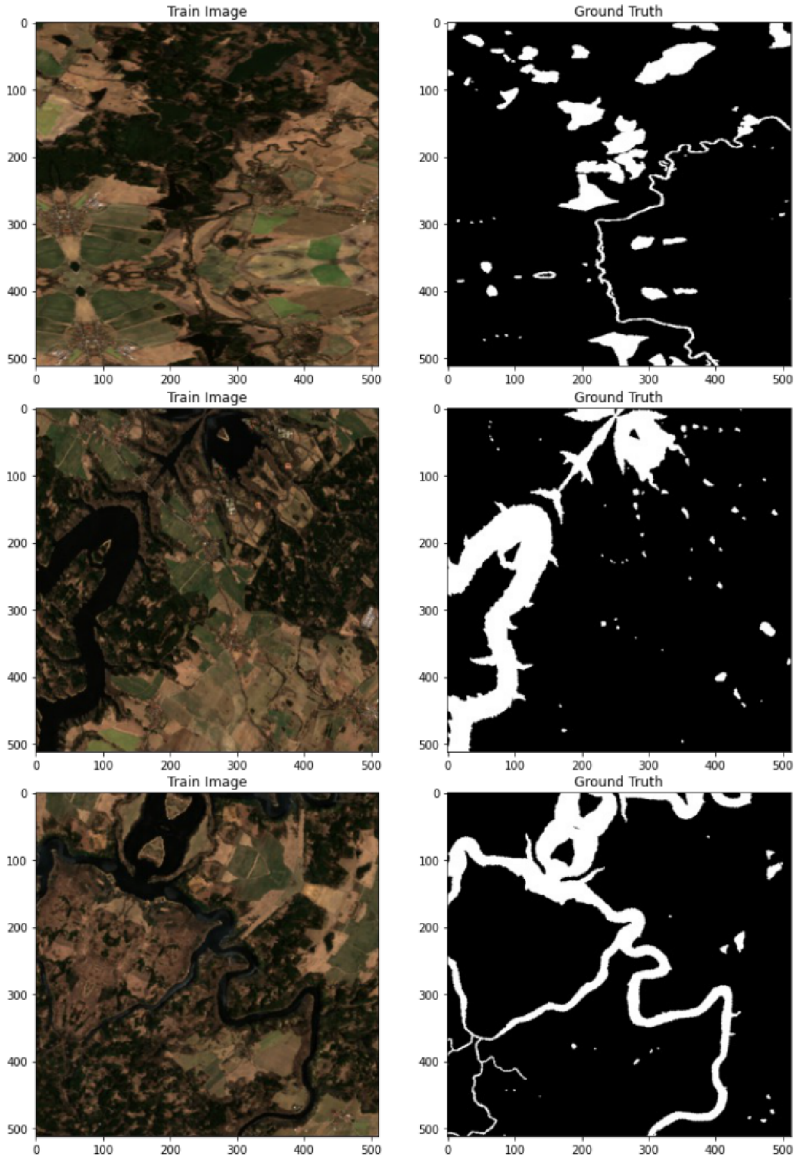

Figure 2: Train set images and corresponded ground truth images. The size of image is 512×512.

tified linear unit "ReLu" as activation functions. In each level of the encoder part (down-sampling), The image size was halved by applying 2×2 max pooling

operation, and the number of feature channels was doubled using convolutions. The maximum value was selected in the 2×2 area with the stride of two by max pooling operation. The encoder part of the network extracts the features and learns an abstract representation of the input image through a sequence of the encoder blocks.

In the decoder or up-sampling section (Fig. 3, Part B), the dimension of the feature maps in each level was doubled from the layer at the bottom to the top layer till achieved the exact same size as the input images. The bridge connection combined the extracted features from the encoder part into the decoder section. As a result of the concatenation step, the channels of the output feature maps will be twice as big as the size of the input features. The Concatenation step of feature maps in U-Net gives us better localization information. The output of the last decoder layer at the top includes 1×1 convolution with Sigmoid activation to predict the probabilities value of pixels for classification purposes. The size of the feature map at the output layer was achieved the exactly as same size as the input layer by applying Padding in the convolution process. The decoder part of the network used extracted abstract representation from the encoder part and generated a semantic segmentation mask. The Binary Focal Loss was used as loss function of the U-Net.

2.2.2. Residual Attention U-Net

The architecture of U-Net consists of encoder and decoder blocks that are connected via a bridge at each level (Fig. 3). The bridge connections are responsible for merging the down-sampling and up-sampling paths together to reach spatial information. On the other hand, the concatenation step may transfer many unimportant and useless feature representations from the encoder part during the combination process. The attention mechanism implemented based on U-Net architecture (Fig. 4, part D) was proposed by Oktay et al. [23] with a promising outcome in medical imaging. The soft attention mechanism was implemented to keep and highlight the most representative features and enhance achieved segmentation results by simple U-Net. The soft attention mechanism

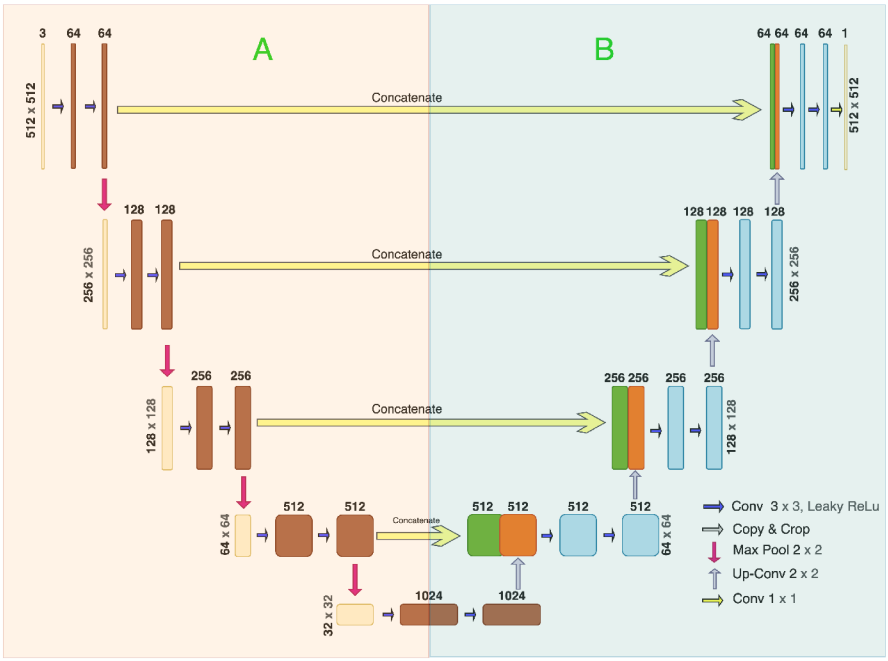

Figure 3: The simple U-Net Architecture. Part A represent the encoder section and part B represent decoder section

remark the important features and represses activations in the unrelated regions. As a result, model sensitivity and performance were slightly improved by employing the attention gate without requiring complicated and heavy computational costs [22].

The employed soft attention gate (Fig. 4, part D) getting two inputs, x and g. The input x was achieved by the concatenation bridges from the early layers of the encoder part and includes better spatial information. Input g comes from the deeper layers of the network known as the gating signal, which includes more efficient feature representation and contextual information to identify the focus region and gives weight to the different parts of the images. The attention coefficients $\alpha \in [0, 1]$ identify, extract, and assign weights to the features belong to the important part of the image regions in our case the water bodies. The attention mechanism progress, getting the weights to the pixels according to

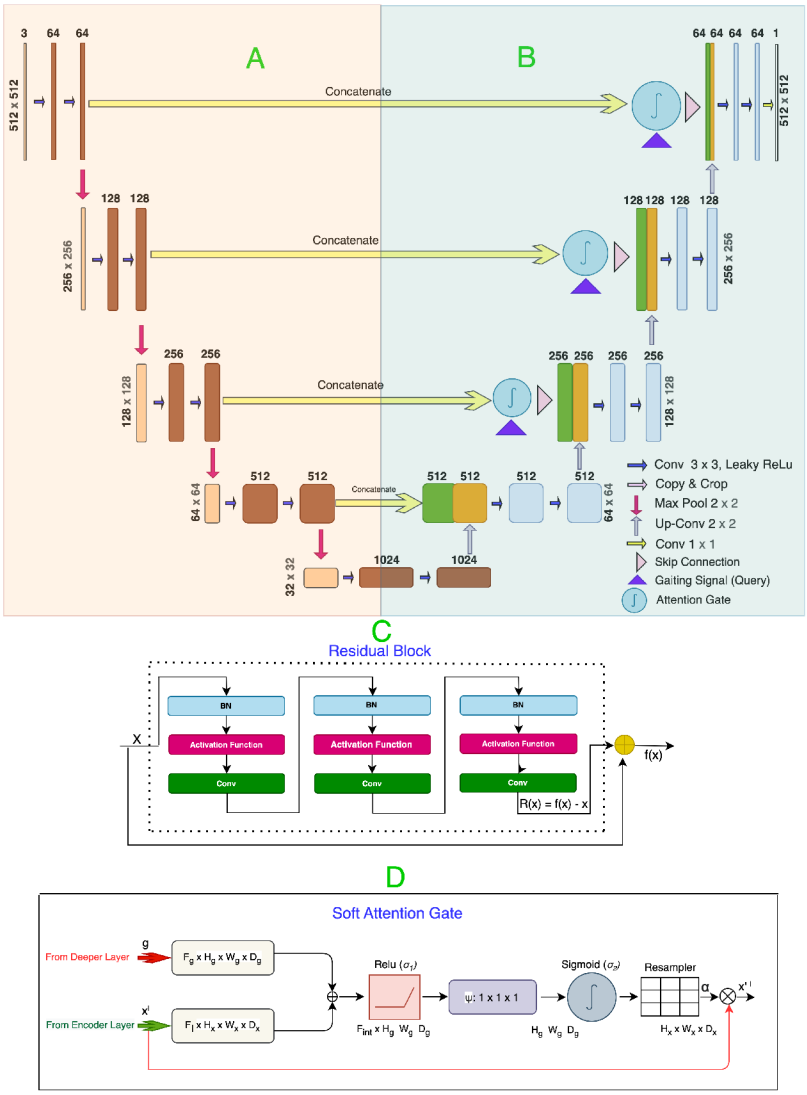

Figure 4: The proposed architecture for Residual attention U-Net. Part A represents the encoder section, and part B represents the decoder section. Part C represents the residual mechanism. Part D represent the soft Attention mechanism. Each feature map has size as $H \times W \times D$, which H, W, and D represent height, width, and number of channels.

their relevance in training steps [23]. The more relevant part of the image will get weights bigger than the less relevant parts. So, by applying the achieved

weights in the training process, we trained model that is more attentive to the relevant image parts. The multiplication of the input feature maps x^l and the achieved attention coefficient α generate the output of the attention gate:

$$\mathrm{q}_{att}^{I} = \psi^{T}(\sigma_1(W_x^T x_i^I + W_g^T g_i + b_g)) + b_\psi, \tag{1}$$

$$\alpha_i^I = \sigma_2(p_{att}^I(x_i^I, g_i; \Theta_{att})), \tag{2}$$

whereas the σ_1 and σ_2 parameters correspond to the relu and sigmoid activation functions and Θ_{att} indicate different parameters including linear transformations W_x and W_g, function ψ and bias terms b_ψ and b_g [23].

Deeper neural networks deliver more effective performance in complex classification and segmentation tasks [24]. Each level of the proposed U-Net-based architectures consists of many convolutional blocks (Fig. 4). The input value enters into the Convolutional blocks, the convolution operation, and the activation function applied in the input value and generates the output. In neural networks, the output of each convolutional block is the input of the next convolutional block. So, by making the neural network architecture deeper, the calculated gradient value from one block to another will be smaller because of the gradient vanishing effect, and the accuracy of the trained model will degrade rapidly instead of improving. The gradient vanishing problem appeared during the training procedure and affected the model's generalization ability. To mitigate this problem, the residual mechanism was implemented and applied to the proposed method to continuously update the calculated gradient values in each convolutional block and improve the performance of trained models [25]. The proposed residual blocks, known as skip connections, will bypass one or more layers and update the gradient values from one or more previous layers into the layer step ahead. By combining the soft attention mechanism with the residual mechanism, we will get the weights into the important part of the image and overcome the gradient vanishing problem during training progress.

2.2.3. VGG16-U-Net

Different CNN architectures have been proposed to be combined with the U-Net architecture for improving the trained model accuracy and computational cost of the U-Net and reducing the number of trainable parameters in comparison to the original U-Net. The VGG is the basis of CNN architecture proposed by Simoniyan et al. [26] and developed by the Visual Geometry Group from Oxford university. The VGG was developed and proposed to reduce the number of trainable parameters in the Convolutional layers and improve the training time because of the structure of the developed architecture proposed by [26]. The VGG architecture has many different variants depending on the number of layers from VGG11 to VGG19. The VGG16 efficiently performed many object detection and image classification tasks [27, 28]. Due to this, in this research, the hybrid VGG16-U-Net architecture was chosen and implemented to compare with two other methods and improve the semantic segmentation results in term of performance and computational costs. To implement the proposed hybrid network, the encoder part of the U-Net, which is responsible for extracting the feature representation, was completely replaced with the VGG16 structure (Fig. 5, part B). The VGG16 architecture at the encoder part (Fig. 5, part A) consists of sixteen layers, including thirteen convolutional layers and three dense layers. The 3 fully connected layers of Vgg16 (Fig, 5, part A, green rectangles) were replaced with architecture that resembled the decoding part of U-Net, which formed the expanding path with convolution layers and upsampling layers (Fig. 5, part B). Hence, the VGG16 without the final 3 fully connected layers was retained as the contracting path [29].

The first layer of the encoder section takes the input image with the size of 512×512 in RGB color mode and has 64 channels. Each convolutional blocks in each level have max pooling progress with the size of 2×2 and a stride of two to extract the maximal value. In each level of the encoder section, the size of the image was half, and the size of feature channels was doubled from 64 to a maximum of 512. The right side of the network (Fig 6, Part B) represents the

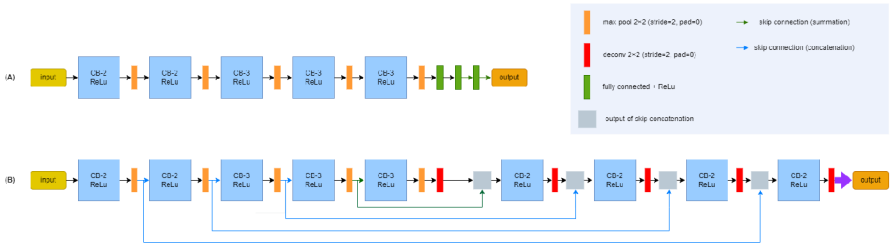

Figure 5: Architecture of the VGG16 and its variants. *A*) represent the VGG16 network architectur. *B*) represent VGG16–U–Net architecture.

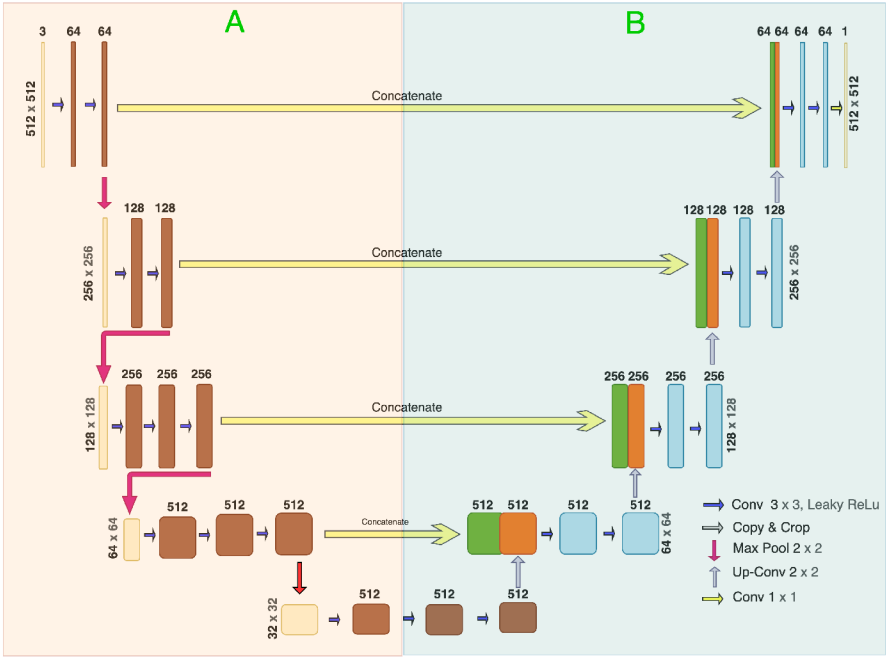

Figure 6: Architecture of the proposed Hybrid VGG16-U-Net model. *A*) represent the encoder part of VGG16 architecture, *B*) represent the decoder part of U-Net respectively.

decoder part with five levels. The structure of the decoder section remained the same as we applied in the simple U-Net method. Each level of the encoder and decoder parts was connected via a concatenation bridge. The concatenation step combines features extracted from the encoder section with the decoder section, and this concatenation step is important for achieving localization information. The last encoder layer has 1×1 convolutional size to predict the probability value of each pixel and generate the semantic segmentation by applying the "Sigmoid" activation function.

2.3. Training Models

The computational platform used for implementing all methods is Python 3.9. All deep learning frameworks were implemented using Keras with the back-end of Tensorflow [30] to train the best stable models. After developing methods and completing of implementation phase for all CNN architectures, the complete method was transferred and compiled on the Google Collab Pro + cluster account. The google clusters are equipped with two vCPU as processors, 24 Gb of RAM as memory, and P100 and T4 graphical processor unit (GPU) [31]. By the completion of the data pre-processing step (Sect. 2), 80% of the main dataset was chosen randomly as a train set (322 images), and the rest of 20% was considered randomly as a test set (101 images) for testing and evaluating the generated models' performance. Meanwhile, 20% of the training set was chosen randomly as the validation set (81 images) to validate the model and prevent over-fitting problems during the training process.

The input image size used in proposed CNN architectures was 512×512 px. All dataset images were resized from 2048×2048 px into 512×512 px as proper and specific input image size for proposed CNN's. We employed data augmentation variables during model training for all three CNN methods. The best-achieved values for each hyperparameter were reported in Tab. 1. The early stopping parameters are useful to prevent the over-fitting problem in the training phase. The threshold for patient value is set equal to 20. The "Relu" was selected as an activation function, and the Batch size value was considered

8. As a description of data Augmentation parameters, the "rotation range" means randomly rotating images between [-90,90] degrees. The "width shift range" shift the image to the left or right (horizontal shifts), and the "height shift range" parameter shifts the image vertically (up or down). The "shear range" parameter shows a distorted image along an axis to create or rectify the perception angle. The random zoom for the training images was obtained by the "zoom range" parameter. For optimizing the network, we choose the 'Adam' optimizer. The learning rate value was considered to 10^{-3}.

Table 1: The value of Hyperparameters used for all CNN models.

Hyperparameter	**Value**
Activation function	Relu
Learning rate	10^{-3}
Size of the Bach	8
Number of the Epochs	70
Early stopping	20
Number of steps in each epochs	100
Rotation range	90
Width shift	0.3
Height shift	0.3
Shear range	0.5
Zoom range	0.3

Semantic segmentation progress could be defined as a classification task at the pixel level to classify those pixels into water bodies or other classes. The segmented water bodies' images with the ground truth (GT) were compared to minimize the difference between them during the training using the Dice loss. The Binary Focal Loss was used as a loss function for semantic segmentation (Eq. 3) [32]:

$$\text{Focal Loss} = -\alpha_t(1-p_t)^{\gamma}\log(p_t), \tag{3}$$

Which $p_t \in [0,1]$ represents the predicted probability value achieved by the model for the ground truth class with label y = 1; $\alpha_t \in [0,1]$ corresponding to the weighting factor for class 1 and $1-\alpha_t$ for class 0; and $\gamma \geq 0$ represent-

ing tunable focusing parameter. Applying focal loss efficiently achieved better segmentation performance in regions of images that are challenging to segment (e.g., narrow inland water bodies or inland bodies with a similar texture to forest) and separate sensitive inland water bodies from the background. On the other hand, the focal loss as loss function manages and reduces the participation of the pixels belonging to the specific region that can be segmented easier (e.g., big and visible inland waters) over the image region in the model training progress. The model has the responsibility of updating the gradient direction. This progress depends on the loss of the model.

2.4. Evaluation metrics

To evaluate segmentation models generated by CNN's, different evaluation metrics were used (Eqs. 4–8). The TP represents a true positive, FP indicates a false positive, FN corresponds to a false negative, and TN represents true negative values, respectively [33]. The generated models were evaluated with the test sets using described metrics, and mean values of each metric were reported in table 3.

The accuracy (Acc) metric indicates the percentage of the pixels which segmented correctly from water bodies. The Precision (Pre) metric represents a ratio of the pixels segmented as water bodies that exactly match the masks (GT). The Recall metric indicates the ratio of pixels belonging to the water bodies in the mask (GT), which is detected properly over the segmentation process. The Dice coefficient, known as F1-score, indicates if the segmented area is equal to the mask of the image (GT) in terms of location and level of detail. The F1-score represents ascertaining how accurate is the segmentation result in boundary regions[34] and is more important than the ACC metric for evaluating model performance. The most important metric for segmentation model evaluation is Intersection over Union (IoU), also known as the Jaccard similarity index. The mentioned metric represents the correlation between the prediction of the model and mask (GT) [35, 36], and indicates the overlap and union area proportion for the model predicted and mask (GT).

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{4}$$

$$\mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

$$\mathrm{Recl} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{Dice} = \frac{2 \times \mathrm{Pre} \times \mathrm{Recl}}{\mathrm{Pre} + \mathrm{Recl}} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{IoU} = \frac{\mid y_t \cap y_p \mid}{\mid y_t \mid + \mid y_p \mid - \mid y_t \cap y_p \mid} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

3. Results and discussion

The proposed neural network models were well trained by processing 70 epochs according to the training/validation loss and accuracy plots (Fig. 7). To achieve the best training performance and stability, we assume all models were trained well according to the best-optimized hyperparameter values listed in Table 1. The best hyperparameter values were achieved by training several models based on different values of hyperparameters to achieve the best model performance and training stability. The trained models were evaluated using a test dataset to assess the performance of the proposed models based on the metrics written in Eqs. 4–8.

The simple U-Net model had an average computational cost in comparison with the Residual attention and VGG16-U-Net architecture. However, the number of the trainable parameters in the Residual attention U-net increased dramatically because of soft attention and residual mechanism, which cause the highest computational cost by this architecture. On the other hand, VGG16-U-Net had the lowest number of trainable parameters and, as a result, the shortest run time because of the structure of this architecture and achieved the best performance compared with the other two proposed methods (Tab. 2).

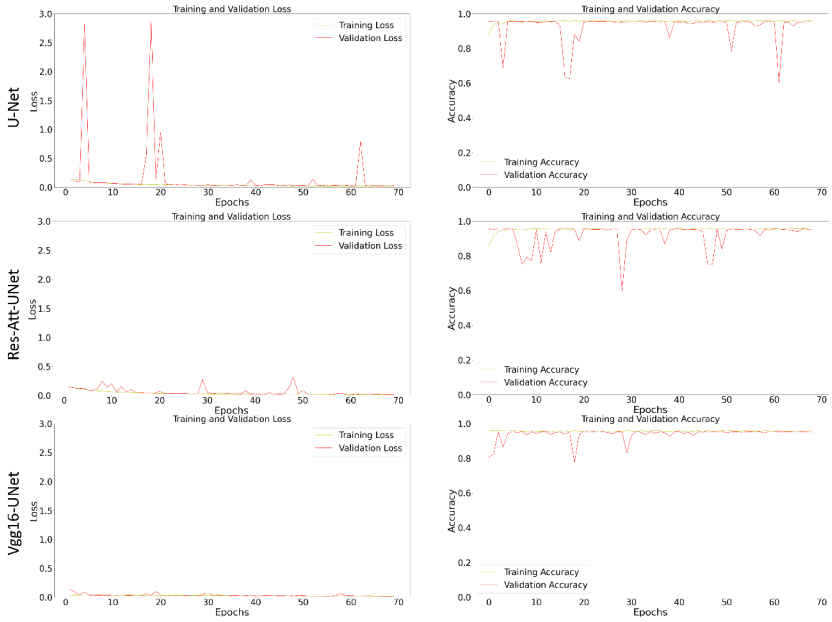

Figure 7: The training loss and accuracy plots for U-Net (first raw), Residual Attention U-Net (second raw), and VGG16-U-Net (third raw).

Figure 8 shows the segmentation results achieved by different proposed CNN architectures. The result of segmentation accomplished by U-Net did not manage to segment all the water bodies over the test set image and suffered from a miss segmentation problem (Fig. 8, red circle). The Residual Attention U-Net segmented the borders of water bodies in complete shape, and the segmentation result was improved in comparison with the simple U-Net. Nevertheless, the result achieved by Residual Attention U-Net faced the under-segmentation problems in some water bodies regions to detect and segment some edges as visualized in Fig. 8, green circle. The best performance of the segmentation was achieved by the VGG16-U-Net method. The result represents a more precise and accurate segmentation of the water bodies' borders, especially in the edge region and sensitive areas (Fig. 8, light blue circle).

Table 3 displays the evaluation of different U-Net-based proposed models with different evaluation metrics using (Eqs. 4–8) as the mean value for all

Table 2: CNN's architecture trainable parameters and runtimes.

Network name	Training time	Trainable parameters
U-Net	3:01':47"	31,402,501
Residual Attention U-Net	4:17':23"	39,090,377
VGG16-U-Net	2:53':19"	25,862,337

the metrics. The simple U-Net achieved the lowest segmentation performance according to the value of Mean-IoU and other evaluation metrics. The Residual Attention U-Net model represents a more improved segmentation result in comparison with the U-Net model in terms of the same test set image and evaluation metric values. In one more step, the segmentation result was further improved after applying the VGG16 encoder architecture with U-Net as a hybrid VGG16-U-Net method.

Table 3: The performance of the CNN Models evaluated by the different metrics. Green highlighted values indicate the best performance of segmentation according to the reported metrics.

Network	Accuracy	Precision	Recall	m-IoU	m-Dice
U-Net	0.9710	0.9997	0.9709	0.9707	0.9849
Residual Attention U-Net	0.9852	0.9986	0.9861	0.9848	0.9923
VGG16-U-Net	0.9855	0.9981	0.9869	0.9850	0.9924

The original U-Net architecture is one of the promising semantic segmentation methods which have been used in different research fields. The original U-Net have been selected as first method to implement and apply in our study. As next phase, we slightly improved the obtained result by modifying the original U-Net architecture by adding the residual mechanism together with soft attention mechanism as extension into the original U-Net. At the last step, we replaced the encoder (feature extraction) part of the U-Net with more powerful VGG16 architecture to build hybrid CNN architecture with more efficient feature extraction section and compare the obtained result with previous methods in term of performance and computational costs.

To the best knowledge, there is no similar research that has been done be-

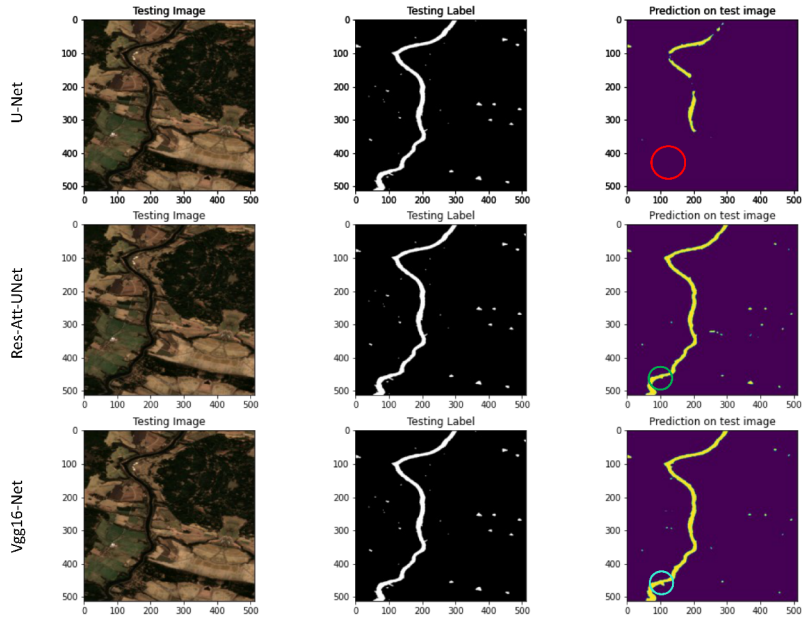

Figure 8: Result of Segmentation for the U-Net (the red circle visualises the miss-segmentation of water bodies), Residual Attention U-Net (the green circle visualises the under-segmentation issue), and the VGG16-U-Net (light blue circle visualises the accurate segmentation of the water bodies. The size of images is 512×512.

fore based on the proposed methods for detecting and segmenting inland water. However, Some researchers applied different deep learning algorithms to detect and segment the inland waters. Table 4 represent the comparison of the similar literature with the proposed methods in this study. Zhong et al. [37] proposed a noise-cancelling transformer network (NT-Net) for the automatic extraction of lake water bodies from remote sensing images and resolve the over-segmentation problem obtained by other literature. The proposed method obtained a 0.862 accuracy value in terms of the IoU metric. Zhang et al. [38] proposed a modified feature extraction network and a modified encoder-decoder network based on depth-wise separable convolution for segmenting the water bodies. The proposed method achieved 0.984 IoU metric accuracy. The authors in [39] proposed a dense pyramid pooling module (DensePPM) to extract global prior knowledge

with a dense scale distribution for Segmenting Water Bodies From Aerial Images. The proposed method obtained a 0.842 metric value in terms of the IoU metric. Chang et al [40] proposed modified U-Net with residual mechanism and attention mechanism in encoder section based on PMS1 remote sensing data of GF2 satellite. The authors achieved good result (i.e., IoU =0.9270). Ch et al. [41] used Sentinel-2 image with two Band3 (Sentinel-2 Green Channel) and Band8 (Sentinel-2 Infrared Channel) and combined these two channel by following "NWDI" formula (as described in original paper) to achieve dataset images and then applied original U-Net architecture to analyse them. The authors achieved 0.89 of Mean IoU score based on suggested method.

Table 4: comparision of the proposed CNNs with other similar literature. The highlighted Green value represent the highest segmentation accuracy achieved by proposed methods.

Models	**IoU**	**Dice**	**Acc**
prop. U-Net	0.9707	0.9849	0.9710
prop. Residual Attention-U-Net	0.9848	0.9923	0.9852
prop. VGG16-U-Net	0.9850	0.9924	0.9855
NT-U-Net [37]	0.862	-	-
Modified Encoder-Decoder [38]	0.984	-	-
DensePPM [39]	0.842	-	-
Res2U-Net [40]	0.9270	-	-
ResNet50 [18]	0.9781	-	-
U-Net [41]	0.89	-	-

4. Conclusions

The efficiency and quality of the segmentation of orbital remote sensing images are the fundamental elements influencing the application of remote sensing for land cover/use mapping. Image semantic segmentation methods based on deep learning remarkably eliminated conventional segmentation methods' shortcomings (e.g., no distinct segmentation due to complex image background or many target instances in one image). This paper analyzed and compared three different deep learning, U-Net-based methods, including simple U-Net, Residual Attention U-Net, and VGG16-U-Net, to detect and segment inland water bodies

using high-resolution satellite images. The results of this study indicate that the U-Net-based algorithms can be employed to inventory inland water bodies fast, accurately, and inexpensively in terms of computation cost. The results of this study can pave the way for implementing precision land cover mapping based on high-resolution satellite imagery by providing an objective, fast, accurate algorithm for inventorying land covers globally. Therefore, this study can be extended further to investigate other state-of-the-art deep learning algorithms also to evaluate them for other types of land cover/use mapping. The code used in this study is publicly available on our Gitlab repository (https://git.gfz-potsdam.de/ali/remotesensing-hida).

Authors contributions

Conceptualization, A.G., M.S., and S.I.; methodology, A.G., and M.S.; validation, A.G., and M.S.; formal analysis, A.G.; resources, M.S., J.B.; data curation, A.G., and J.B.; writing—original draft preparation, A.G., and M.S.; writing—review and editing, A.G., M.S., J.B., and S.I.; visualization, A.G.; supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

FUNDING

The authors would like to thank the EU, German's Federal ministry of education and research (BMBF), and The Technology Agency of the Czech Republic (TAČR) for funding in the frame of the collaborative international consortium AIHABs financed under the ERA-NET AquaticPollutants Joint Transnational Call (GA Nº 869178). This ERA-NET is an integral part of the activities developed by the Water, Oceans, and AMR Joint Programming Initiatives. Furthermore, the authors appreciate the Helmholtz information and data science academy (HiDA) funding in the frame of the Helmholtz Visiting Researcher Grant. The authors would like to thank the European Regional Development

Fund in the frame of the project ImageHeadstart (ATCZ215) in the Interreg V-A Austria–Czech Republic programme and the project GAJU 114/2022/Z.

DECLARATION OF COMPETITING INTEREST

The authors declare no conflict of interest, or known competing financial interests, or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Zhang, S. Foerster, P. Medeiros, J. C. d. Araújo, Z. Duan, A. Bronstert, B. Waske, Mapping regional surface water volume variation in reservoirs in northeastern Brazil during 2009–2017 using high-resolution satellite images, Science of The Total Environment 789 (2021) 147711. doi:10.1016/j.scitotenv.2021.147711.

[2] S. W. Cooley, J. C. Ryan, L. C. Smith, Human alteration of global surface water storage variability, Nature 591 (7848) (2021) 78–81. doi:10.1038/s41586-021-03262-3.

[3] D. Dudgeon, A. H. Arthington, M. O. Gessner, Z. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A. Prieur-Richard, D. Soto, M. L. J. Stiassny, C. A. Sullivan, Freshwater biodiversity: importance, threats, status and conservation challenges, Biological Reviews 81 (2) (2006) 163–182. doi:10.1017/s1464793105006950.

[4] R. P. Bukata, Retrospection and introspection on remote sensing of inland water quality: "like déjà vu all over again", Journal of Great Lakes Research 39 (2013) 2–5, remote Sensing of the Great Lakes and Other Inland Waters. doi:https://doi.org/10.1016/j.jglr.2013.04.001.

[5] S. C. Palmer, T. Kutser, P. D. Hunter, Remote sensing of inland waters: Challenges, progress and future directions, Remote Sensing of Environment 157 (2015) 1–8, special Issue: Remote Sensing of Inland Waters.

doi:https://doi.org/10.1016/j.rse.2014.09.021.
URL https://www.sciencedirect.com/science/article/pii/S0034425714003666

[6] Y. Xu, L. Yu, D. Feng, D. Peng, C. Li, X. Huang, H. Lu, P. Gong, Comparisons of three recent moderate resolution african land cover datasets: Cgls-lc100, esa-s2-lc20, and from-glc-africa30, International Journal of Remote Sensing 40 (16) (2019) 6185–6202. doi:10.1080/01431161.2019.1587207.

[7] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, M. Ranagalage, Sentinel-2 data for land cover/use mapping: A review, Remote Sensing 12 (14). doi:10.3390/rs12142291.

[8] G. L. Feyisa, H. Meilby, R. Fensholt, S. R. Proud, Automated water extraction index: A new technique for surface water mapping using landsat imagery, Remote Sensing of Environment 140 (2014) 23–35. doi:https://doi.org/10.1016/j.rse.2013.08.029.
URL https://www.sciencedirect.com/science/article/pii/S0034425713002873

[9] Z. Zou, J. Dong, M. A. Menarguez, X. Xiao, Y. Qin, R. B. Doughty, K. V. Hooker, K. David Hambright, Continued decrease of open surface water body area in oklahoma during 1984–2015, Science of The Total Environment 595 (2017) 451–460. doi:https://doi.org/10.1016/j.scitotenv.2017.03.259.
URL https://www.sciencedirect.com/science/article/pii/S0048969717307908

[10] J. Worden, K. M. de Beurs, J. Koch, B. C. Owsley, Application of spectral index-based logistic regression to detect inland water in the south caucasus, Remote Sensing 13 (24). doi:10.3390/rs13245099.

[11] T. Bangira, S. M. Alfieri, M. Menenti, A. van Niekerk, Comparing thresholding with machine learning classifiers for mapping complex water, Remote Sensing 11 (11). doi:10.3390/rs11111351.

[12] P. Ghasemigoudarzi, W. Huang, O. D. Silva, Q. Yan, D. Power, A Machine Learning Method for Inland Water Detection Using CYGNSS Data, IEEE Geoscience and Remote Sensing Letters 19 (2020) 1–5. doi:10.1109/lgrs.2020.3020223.

[13] J. Worden, K. M. de Beurs, Surface water detection in the caucasus, International Journal of Applied Earth Observation and Geoinformation 91 (2020) 102159. doi:https://doi.org/10.1016/j.jag.2020.102159.

[14] F. Pan, X. Xi, C. Wang, A comparative study of water indices and image classification algorithms for mapping inland surface water bodies using landsat imagery, Remote Sensing 12 (10). doi:10.3390/rs12101611.

[15] W. Zhao, S. Du, Q. Wang, W. J. Emery, Contextually guided very-high-resolution imagery classification with semantic segments, ISPRS Journal of Photogrammetry and Remote Sensing 132 (2017) 48–60. doi:https://doi.org/10.1016/j.isprsjprs.2017.08.011.

[16] Z. Lv, T. Liu, J. A. Benediktsson, N. Falco, Land cover change detection techniques: Very-high-resolution optical images: A review, IEEE Geoscience and Remote Sensing Magazine 10 (1) (2022) 44 – 63. doi:10.1109/MGRS.2021.3088865.

[17] W. Zhang, P. Tang, L. Zhao, Fast and accurate land-cover classification on medium-resolution remote-sensing images using segmentation models, International Journal of Remote Sensing 42 (9) (2021) 3277–3301. doi:10.1080/01431161.2020.1871094.

[18] S. An, X. Rui, A high-precision water body extraction method based on improved lightweight u-net, Remote Sensing 14 (17). doi:10.3390/rs14174127.

[19] ZABAGED map czech republic, https://geoportal.cuzk.cz/(S(3nw2c4xgeg3kzypmcedojne3))/Default.aspx?mode=TextMeta&text=dSady_zabaged&side=zabaged&menu=24/.

[20] Pachify lib python, https://pypi.org/project/patchify.

[21] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol. 9321, Springer, Cham, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

[22] A. Ghaznavi, R. Rychtáriková, M. Saberioon, D. Štys, Cell segmentation from telecentric bright-field transmitted light microscopy images using a residual attention u-net: A case study on hela line, Computers in Biology and Medicine 147. doi:10.1016/j.compbiomed.2022.105805.

[23] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning where to look for the pancreas, in: 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), 2018, pp. –.

[24] K. Nishimura, C. Wang, K. Watanabe, D. F. E. Ker, R. Bise, Weakly supervised cell instance segmentation under various conditions, Medical Image Analysis 73. doi:10.1016/j.media.2021.102182.

[25] Z.-L. Ni, G.-B. Bian, X.-H. Zhou, Z.-G. Hou, X.-L. Xie, C. Wang, Y.-J. Zhou, R.-Q. Li, Z. Li, Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments, in: IEEE conference on computer vision and pattern recognition, 2019. doi:10.48550/arXiv.1909.10360.

[26] A. Z. Karen Simonyan, Very deep convolutional networks for large-scale image recognition, ICLR Conferencedoi:https://doi.org/10.48550/arXiv.1409.1556.

[27] W. A. Hamwi, M. M. Almustafa, Development and integration of vgg and dense transfer-learning systems supported with diverse lung images for discovery of the coronavirus identity, Informatics in Medicine Unlocked 32 (2022) 101004. doi:https://doi.org/10.1016/j.imu.2022.101004.

[28] D. L. I. Wahyuni, W. J. Wang, C. C. Chang, Rice semantic segmentation using unet-vgg16: A case study in yunlin, taiwan, IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien City, Taiwandoi:10.1109/ISPACS51563.2021.9651038.

[29] C. Balakrishna, S. Dadashzadeh, S. Soltaninejad, Automatic detection of lumen and media in the ivus images using u-net with vgg16 encoder, in: Arxiv, 2018. doi:10.48550/arXiv.1806.07554.

[30] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, Tensorflow: large-scale machine learning on heterogeneous distributed systems, in: OSDI'16: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, 2016, pp. 265–283.

[31] Google, System spec, Retrieved on 12/12/2021.
URL https://research.google.com/colaboratory/faq.html

[32] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2020) 318–327. doi:10.1109/TPAMI.2018.2858826.

[33] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, Y. Fan, Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks, Neurocomputing 229 (2017) 88–99. doi:10.1016/j.neucom.2016.08.103.

[34] G. Csurka, D. Larlus, F. Perronnin, What is a good evaluation measure for semantic segmentation?, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2013, pp. 32.1–32.11. doi:10.5244/C.27.32.

561 [35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for se-
562 mantic segmentation, in: IEEE conference on computer vision and pattern
563 recognition, 2015, pp. 3431–3440. doi:10.1109/CVPR.2015.7298965.

564 [36] B. Vijay, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-
565 decoder architecture for image segmentation, IEEE Trans. Pattern Anal.
566 Mach. Intell. 39 (12) (2015) 228–233. doi:10.1109/TPAMI.2016.2644615.

567 [37] S. H. M. e. a. Zhong H F, Sun Q, Nt-net: A semantic segmentation network
568 for extracting lake water bodies from optical remote sensing images based
569 on transformer [j], IEEE Transactions on Geoscience and Remote Sensing
570 60 (2022) 1–13. doi:10.1109/TGRS.2022.3197402.

571 [38] P. Zhang, G. Wang, The modified encoder-decoder network based on depth-
572 wise separable convolution for water segmentation of real sar imagery, 2019
573 International Applied Computational Electromagnetics Society Symposium
574 60 (2019) 1–2. doi:10.23919/ACES48530.2019.9060500.

575 [39] W. W. D. Xiang, X. Zhang, H. Liu, Denseppmunet-a: A robust deep
576 learning network for segmenting water bodies from aerial images, IEEE
577 Transactions on Geoscience and Remote Sensing 61 (2023) 1–11. doi:
578 10.1109/TGRS.2023.3251659.

579 [40] B. Z. e. a. Chang X, Fei Y, High-resolution remote sensing water extraction
580 based on improved u-net, in: 7th International Conference on Information
581 Science, Computer Technology and Transportation, 2022, pp. 1–5.

582 [41] G. S. I. C. G. T. D. I. Ch A., Ch R., Ecdsa-based water bodies prediction
583 from satellite images with unet, Waterdoi:10.3390/w14142234.

Paper 4

Estimation of rheological parameters for unstained living cells

Authors: Lonhus, K., Rychtáriková, R., **Ghaznavi, A.**, and Štys, D.

Eur. Phys. J. Spec. Top. (2021) 230:1105–1112
https://doi.org/10.1140/epjs/s11734-021-00084-2

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Estimation of rheological parameters for unstained living cells

Kirill Lonhus[a], Renata Rychtáriková[b], Ali Ghaznavi, and Dalibor Štys

Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Kompetenzzentrum MechanoBiologie in Regenerativer Medizin, Institute of Complex Systems, University of South Bohemia in České Budějovice, Zámek 136, 373 33 Nové Hrady, Czech Republic

Received 25 May 2020 / Accepted 18 January 2021 / Published online 19 April 2021

Abstract In video-records, objects moving in intracellular regions are often hardly detectable and identifiable. To squeeze the information on the intracellular flows, we propose an automatic method of reconstruction of intracellular flow velocity fields based only on a recorded video of an unstained cell. The basis of the method is detection of speeded-up robust features (SURF) and assembling them into trajectories. Two components of motion—direct and Brownian—are separated by an original method based on minimum covariance estimation. The Brownian component gives a spatially resolved diffusion coefficient. The directed component yields a velocity field, and after fitting the vorticity equation, estimation of the spatially distributed effective viscosity. The method was applied to videos of a human osteoblast and a hepatocyte. The obtained parameters are in agreement with the literature data.

1 Introduction

A typical bright-field microscopy experiment is time-lapse recording of a sequence of images. In case of living unstained samples, it is little known about structure of the observed objects. It is usually possible to discriminate a cell from its background, find its nucleus, but not more [1]. However, the microscopy image is much more complicated and one can see motion of some intracellular structures and movement of small 'particles' inside the cell. These objects are extremely diverse in texture and shape, frequently do not have sharp boundaries, and are mostly too small for identification.

In this article, we aim to investigate cell rheological and microfluidic properties without any a priori information about cell structure or composition. There are approaches aimed specifically at investigation cell flows, e.g., [2], but they require fluorescent labeling and a mathematical model of the studied cell. There are model-free approaches as well. These are based on correlation computations, e.g., [3], have a solid mathematical background, and at good conditions and for well-behaved objects, can deliver good results. But these correlation methods suffer from the fact that they cannot distinguish the points and rely on proximity based assignment. As a result, these methods inevitably suffer from error propagation during tracking. Another way is to segment some sufficiently large objects and then track them until they are overlapping, e.g., [4]. These methods do not suffer from the error propagation so much, but require segmentable entities in the cell image. Even then, the count of followed objects can be too small for flow reconstruction. Moreover, all methods described above do not address the fact that small particles can be susceptible to the Brownian motion. All the methods also often assume that the random component of motion can be safely neglected.

The main idea of the method proposed here is tracking of identifiable spots inside a cell followed by reconstruction of local properties of media and fields of velocities. This approach is similar to two well-known model-free approaches to the velocity reconstruction such as the Particle Image Velocimetry (PIV) [5] and the Particle Tracking Velocimetry (PVT) [6]. After that, the nonlinear optimization of minimum covariance, alternating likelihood fitting, enables us to separate the observed motion to components of the Brownian and direct flow, respectively, yielding both rectified flows and local media properties.

2 Materials and methods

To show capacity of the method, we applied it to microscopic image data from time-lapse experiments on live human cells of lines MG63 and HepG2.

[a] e-mail: lonhus@jcu.cz
[b] e-mail: rrychtarikova@frov.jcu.cz (corresponding author)

2.1 Cell sample preparation

A MG63 (human osteosarcoma, Sigma-Aldrich, cat. No. 86051601) and a HepG2 (human hepatocellular carcinoma, Sigma-Aldrich, cat. No. 85011430) cell lines were grown at low optical density overnight at 37 °C, 5% CO_2, and 90% RH. The nutrient solution consisted of DMEM (87.7%) with high glucose (> 1 g L^{-1}), fetal bovine serum (10%), antibiotics and antimycotics (1%), L-glutamine (1%), and gentamicin (0.3%; all purchased from Biowest, Nuaillé, France).

During the microscopy experiments, the MG63 cells were maintained in a Petri dish with a cover glass bottom and lid at room temperature of 37 °C. The HepG2 cells were cultivated in a Bioptechs FCS2 Closed Chamber System at 37 °C (Table 1).

2.2 Bright-field wide-field video-enhanced microscopy

The living cells were captured using a custom-made inverted high-resolved bright-field wide-field light microscopes enabling observation of sub-microscopic objects (ICS FFPW, Nové Hrady, Czech Republic): The HepG2 line was captured by an older type of microscope (so-called nanoscope, built 2011), whereas the MG63 cell line was scanned using a newer type of microscope (so-called superscope, built 2020).

The optical path of the both microscopes is very simple and starts by a light emitting diode(s) which illuminate(s) the sample by series of light flashes (synchronized with a microscope digital camera exposure and image saving speed) in a gentle mode and enable the video enhancement [4]. In the case maybe, a light filter is applied to protect the sample from undesirable intensities. After passing through a sample, light reaches a Nikon objective. In the nanoscope, a Mitutoyo tube lens magnifies and projects the image on a high-resolved rgb digital camera. At this total magnification, the size of the object projected on the camera pixel is under the Abbe diffraction limit, i.e., 32 and 23 nm, respectively. The process of capturing the primary signal was controlled by a custom-made control software. In both cases, we performed a time-lapse experiment from a compromise focal plane of the cell. The microscope setups differ as written in Table 1.

2.3 Image preprocessing

To suppress the image distortions, the microscope optical path and camera chip was calibrated and the obtained time-lapse micrographs were corrected by a radiometric approach described in detail in [7].

The raw images were recorded in the color preserving RGB mode when three intensity values (in the red, green, and blue image channel) are assigned to each image point (pixel). In this color-preserving image representation, four camera pixels are always merged in a way that the resulting number of the RGB image pixels is a quarter (see [8] for details). In other words, the resulting pixel size is doubled, i.e., 64 nm and 46 nm, respectively (cf. Table 1). Since all examined feature detectors work on single-channel images, the RGB images were converted to grayscale in the standard way ($0.2989{\cdot}R + 0.5870{\cdot}G + 0.1140{\cdot}B$, where R, G, and B are intensities of pixels in the red, green, and blue raw image channel, respectively) [9]. To eliminate subtle changes in illumination, the images were robustly rescaled to [0..1], after saturating 1% of both the darkest and the brightest pixels simultaneously.

Prior to any tracking, the objects of interest (live cells) have to be robustly detected and segmented from image background. Therefore, we annotated a few (usually 1%) images from the sequence visually to interpolate contours of the observed cell in the unannotated images. For interpolation of the contours, we used a weighted mean of strings [10]. After contours were interpolated, we applied a non-parametric image deformation registration [11]. The obtained displacement field was employed to compensate position shift between the images.

3 Estimation of intracellular flows

The algorithm for the estimation of the flows and rheological parameters in the intracellular environment of the unstained cells is showed in Fig. 1 and described in detail in the following subsections. The Matlab codes and the input and output data are available at the Dryad data depository [12].

3.1 Feature extraction and tracking

There are numerous methods, e.g., [13,14], for tracking local image features, i.e., feature vectors describing special, well-distinguishable image points. These methods are usually designed to match the same object from different views. Our problem is opposite—to match different (but similar) objects from the same view. We tested BRISK [15], ORB [16], MSERF [17], KAZE [18], MinEig [19], and SURF [20] image features to estimate their efficacy (Fig. 2b; see Sect. 3.2 for determination of the error in separation of the direct motion from the random walk). The SURF performs the best, followed by the MinEig. The further analysis showed that the SURF output is much more robust to small changes in the image. The SURF method is based on calculation of the Hessian matrix for each pixel of the smoothed (via approximated Gaussian smoothing; a box filter with kernel $9{\times}9$ px and $\sigma = 1.2$) image separately. The pixels whose matrix determinants were maximal were treated as the 'points'. An image pyramid with 3 scales was further used. The descriptors themselves were oriented Haar wavelets [20].

The next step was to track a point through consecutive frames. To avoid a computationally intensive $O(n^2)$ point match (where n is a number of points in an image), we used a heuristic approach—the same points in consecutive frames should be nearby. A small, ran-

Eur. Phys. J. Spec. Top. (2021) 230:1105–1112

Table 1 Bright-field wide-field microscopy constructions and setups

Microscope (cell)	Nanoscope (HepG2)	Superscope (MG63)
LEDs	2 × Luminus CSM-360, 4500 mA (59.625 W)	1 × Luminus CFT-90-W, 40% of max. intensity
Light pattern	Light 226.1 ms–dark 96.9 ms	light 0.2 ms-dark 199.8 ms
Light filters	Edmund optics, i.r. 775 nm short-pass, u.v. 450 nm long-pass	No
Objective	Nikon LWD 40 ×, Ph1 ADL, 1/1.2, N.A. 0.55, W.D. 2.1 mm	Nikon CFI Achromat 60 ×, N.A. 0.80, W.D. 0.30 mm
Tube lens	Mitutoyo, 4 ×	No
Camera	JAI, rgb Kodak KAI-16000 chip, 4872 × 3248 px	Ximea MX500-CG-CM-X4 G2-FL rgb, 7920 × 6004 px
Camera Bayer mask	GBRG	BGGR
Camera exposure	293.6 ms (gain 0, offset 300)	0.2 ms
Pixel size	32 nm	23 nm
Scanning frequency	0.2 fps	5 fps
Experiment length	2446.869 s	83.2 s
Cell cultivation	Bioptechs FCS2 closed chamber system	Ibidi μ-dish 35 mm, high glass bottom, DIC lid
No. of px per cell	$(2.137 \pm 0.048) \times 10^6$	$(5.623 \pm 0.084) \times 10^5$
No. of images	473	416

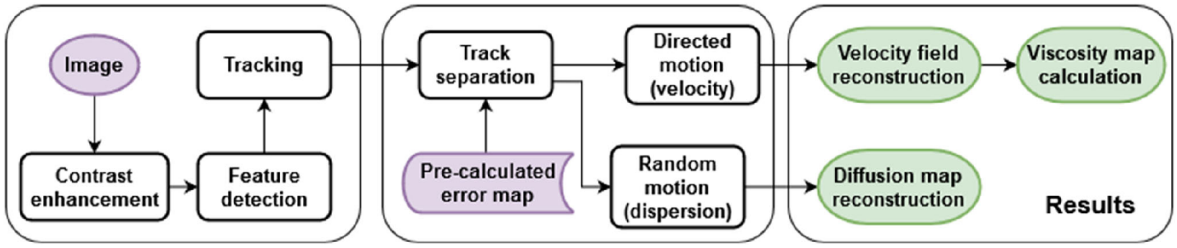

Fig. 1 Algorithm of the method for calculation of the viscosity map and diffusion map of the intracellular environment

dom, subset of (∼ 10) pairs of consecutive images was used to estimate the maximal point displacement in two images: For each pair of the consecutive frames, we found a median of the minimal distances between each two points. Then, the resulted effective displacement ED was calculated as a mean from all medians of the minimal distances. Finally, we assume that the match between the points is possible if the distance is smaller than $3 \cdot ED$. In this way, each point obtained typically 10–15 possible candidates for tracking in the following image, and thus, we effectively reduced feature matching complexity to $O(n)$ and eliminated the long-range matching error.

The tracking process itself is iterative. At each step we classified all detections into two sets: assigned and unassigned. To be assigned, a detection in any track had to fulfill two criteria—to be spatially close (closer than 3 average offsets) and feature-wise close (the Euclidean distance between the last and the current vector of the track has to be smaller than 1). The unassigned detection created new tracks. The tracks which were not assigned for a longer period than K frames were removed. Since the influence of K on quality of the final result has not been investigated, we used the safest choice of $K = 1$.

3.2 Decomposition to direct and Brownian motion

The segmented trajectories are sets of points in $\mathbb{R}^2$, usually 10–300 points. We assume that the trajectories exhibit two simultaneous types of motion—Brownian and direct. As widely accepted (the Einstein model), the Brownian motion of small particles can be described as a Gaussian process with zero mean. To separate the components of motion, we used the minimization of a maximum differential entropy, which for a multivariate normal distribution follows $h(x) \leq \frac{1}{2} \log \det \mathrm{cov}(\mathbf{X})$. In this way we proposed a formulation of the separation problem as

$$\mathbf{V}_d = \min_{\mathbf{V} \in \mathbb{R}^2} \log |\mathrm{cov}(\mathbf{P}_n - n\mathbf{V})|, \tag{1}$$

where $\mathbf{P}_n$ is a position of the tracked point in time step n and $\mathbf{V}_d$ is the searched velocity. Equation 1 can be also viewed as direct usage of the minimum covariance approach.

This optimization also gives a corrected (with a compensated drift) set of points from which 'normal' covariance and mean value can be estimated. We chose a nonlinear optimization—sequential-quadratic programming [21]—which, in the vicinity of a current point,

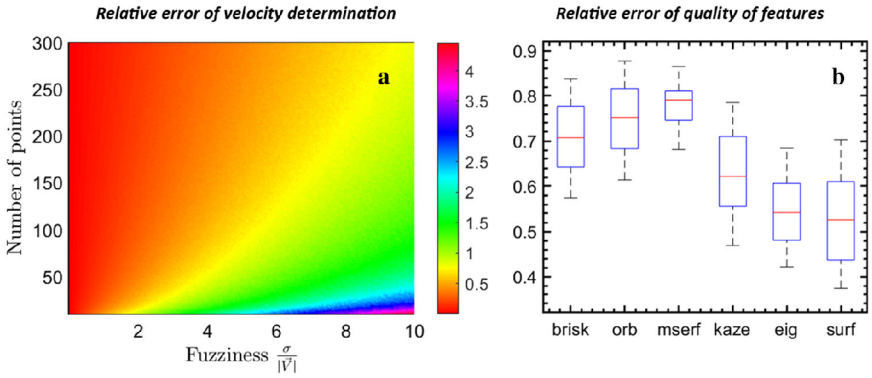

Fig. 2 **a** Relative error of velocity determination as a function of number of points in trajectory and ratio between standard deviation σ and norm of the velocity V. **b** Relative error of quality of features for feature extraction methods

iteratively approximates a nonlinear problem by a quadratic one and solves this simpler problem by a QR decomposition. This method is not global and relies on the initial guess. We used the safest guess—the zero velocity—which coincides with the null hypothesis.

To verify this approach, we performed the following numerical experiment (simulation): the most straightforward way how to mimic the Brownian motion is the random walk, where the steps are drawn from the Gaussian distribution. The simulation itself has two main parameters: a number of points N in a track and fuzziness $\frac{\sigma}{|\mathbf{V}|}$, where σ is a standard deviation of the Gaussian process $\mathcal{N}$ and $\mathbf{V}$ is a drift velocity vector. Then, the position of the tracked point in time step $(n+1)$ is

$$\mathbf{P}_{n+1} = \mathbf{P}_n + \mathbf{V} + \mathcal{N}(0, \sigma). \tag{2}$$

After that, for any random walk with drift, it is possible to apply the resulted components of the method of separation of the direct motion from a random walk and evaluate the error $\mathrm{Err} = \frac{|\mathbf{R}-\mathbf{V}|}{|\mathbf{V}|}$, where $\mathbf{R}$ and $\mathbf{V}$ is the reconstructed and real velocity, respectively.

Using Eq. 2, we simulated numerous tracks varying in the number of time steps (from 8 to 300) and in the fuzziness (from 0.01 to 10 discretized into 500 steps). The data along all 500 trials were averaged and saved as a table (Fig. 2a). By a 2D bilinear interpolation, it was allowed to calculate the error of velocity extraction Err from a non-synthetic data. It requires that the velocity is both spatially and temporarily constant (along the given track) and the observed random motion obeys the Gaussian distribution.

If the data variation is not too high ($\sigma/|\mathbf{V}| < 0.1$), we can carry out a reliable (relative error $\mathrm{Err} < 0.01$) extraction of the drift velocity from sets of down to 10 points. For a higher number of points, the drift velocity extraction gives a quite reliable estimation even if the standard deviation is much greater than the norm of the drift velocity vector.

Due to absence of the ground truth, there is no way how to evaluate quality of the reconstructed flows. But quality of the tracks can be evaluated as the mean separation error of the tracks. In this way, we compared the different feature detectors, defining that a lower reconstruction error means a better detector (Fig. 2b, more above in Sect. 3.1).

3.3 Reconstruction and analysis of intracellular flows

The velocities were defined for the most of the tracks. Some of the tracks were excluded from the future analysis due to a high separation error (the threshold value was chosen 1). There was no way how to attribute the given velocity to the specific position, because we estimated the drift for the whole trajectory. We assumed that the drift is constant along the observed positions in the trajectory. All tracks' velocities were imprinted in a single global image of the cell.

The particles passing through the same point (in 2D projection) at the same time can exhibit completely different velocities. These velocities have to be separated. Since we calculate velocities along the time window, for each pixel we obtain as many estimations of velocities as length of the time window. From these different estimations of velocities, we can calculate the error of velocity separation Err (see Sect. 3.2). In following statistical analysis, we will assign weights to the velocities estimated in this time window. Each of this weight is complementary to the error of separation, i.e., $weight = 1 - \mathrm{Err}$.

The resulted vector field is sparse. To reconstruct it, we used robust splines [22] which minimize the Generalized Cross-Validation (GCV) score. This method was designed to handle the PIV-type data specifically [23].

Eventually, this part of the algorithm produces a global velocity field through the whole image series. In view of the fact that it is not possible to do any real time series analysis, we carried out a quasi-stationary window analysis. The reconstruction was performed on subsets of frames defined by the time window of the size $wsize$ sliding along the whole image sequence. The time window is usually too short to give a reliable reconstruction, and thus, the global flows are used as a guess (with dampened weights) proportional to the ratio between the window size and the total number of images in the series. The resulted velocity field (as a function of the sliding window size) is the closest form how we can

Eur. Phys. J. Spec. Top. (2021) 230:1105–1112

approximate the real time dependence of the velocity field.

We applied the method to two types of objects—a human osteoblast and human hepatocyte observed with bright-field microscopy (see Sect. 2). The main output of the method is a velocity field and distribution of flow speeds (Fig. 3). It is predictable that the intracellular flows in the hepatocyte (a cell with high metabolic activity) are much more intense than in the osteoblast.

3.4 Diffusion and viscosity estimation

The velocity is informative enough, but it does not characterize the intracellular medium itself. To characterize the structure and composition of the medium, some hydromechanical constants, namely space-resolved diffusion coefficient and viscosity, must be extracted.

The separation procedure resulted in the drift-compensated trajectory (see Sect. 3.2). The most straightforward way how to estimate the diffusion coefficient is to use the covariance of derivatives in the random walk:

$$D = \frac{1}{4T}\left\langle \text{diag cov}\frac{\mathrm{d}\mathbf{P}_n}{\mathrm{d}n}\right\rangle, \tag{3}$$

where T is the time interval between consecutive images. Due to presence of derivative in Eq. 3, the diffusion coefficient is invariant to the drift velocity as it was supposed to. These diffusion coefficients were computed for all eligible (Err $<$ 1) tracks. The field of diffusion coefficients was reconstructed in the same way as the velocity field, i.e., by a spline minimizing the GCV score. The reconstructed diffusion fields and distributions can be seen in Fig. 4b, c, f. The values of diffusion coefficients are relatively high, presumably because both the active and passive diffusion happen in the same time and are mutually indistinguishable. Essentially, we deal with effective diffusion, and thus, the comparison with classical molecular diffusion coefficients should be done with caution. Since we work with a 2D slice of a 3D volume, the value of the derived diffusion coefficient should be accurate, assuming its isotropy. No additional smoothing of the final data was used, except removing 5% of points with the least and most intensities, respectively, before reconstruction (to eliminate possible influential errors).

Estimation of the viscosity coefficient is less model-free and based solely on the quasi-stationary velocity field. The kinematic viscosity [24] can be found from the vorticity equation for an incompressible, isotropic, Stokesian fluid in 2D as

$$\nu = \frac{\mathrm{d}\omega}{\mathrm{d}t}\cdot\frac{1}{\nabla^2\omega}, \tag{4}$$

where $\omega = \nabla\times\mathbf{V}$ is the vorticity of the velocity field. One issue of this approach is a high, namely the 3rd, order of derivatives in the spatial domain. This leads to the fact that the calculations will be thus over-susceptible to small errors. The second issue is presence of the time derivative that is absent in the results because the analysis is quasi-stationary and the intracellular flows thus depend on the time window. The window, which we used in the analysis and was the closest to zero, was 7. With decreasing size of the time window, the absolute error is increasing due to less rich statistics. For all windows from 7 to 71 images (only odd numbers are valid as the window size), we calculated the mean velocity field and mean time derivative. The distances between windows $[w, w+wsize]$ and $[w+1, w+wsize+1]$ were assumed 1 frame. But this is strictly true only for $wsize = 0$ and diverges with increasing size of $wsize$. Thus, Eq. 4 was applied to each window and then extrapolated to $wsize = 0$. Due to the higher-derivative noise, the ordinary linear fitting was not sufficient for the extrapolation. Therefore, we had to apply a robust linear fitting [25] with bi-square weights, which gave stable results without necessity of any additional data smoothing (Fig. 4a, d, e).

The obtained values of viscosity are in agreement with some literature data [26]. Nevertheless, some literature sources report much lower viscosities [27]. It can be explained by the fact that the definitions of viscosity at the microlevel are very vague, the relevant values of viscosity then depend frequently on the method of their acquisition, and thus, the real values of viscosity can vary. Again, we work with a single plane of a 3D object, and thus, diffusion and convection along the z axis is neglected. Therefore, it is more correct to call the variable derived here as effective viscosity.

4 Discussion

In this paper, we deal with the total, complex, evaluation of the intracellular flows but the origin of the intracellular flows remains an open question. We can observe visually that these flows do not coincide with specific object motions. In most cases, it is nearly shapeless disturbance in the intracellular medium which is moving, sometimes we deal with small particles or vesicles. We do not speculate nature of these objects or nature of their motion and rather try to analyze it.

The main assumption for the flow analysis is that the tracked entities are driven by two forces—the Brownian and direct motion—which are related to both some global intracellular flow (if exists) and a specific locomotion. The reconstructed flows seem not to be any consequence of the changes in the cell borders but rather some intrinsic phenomena. In an effort to interpret the results from the biological point of view, we chose two very mutually different kinds of cells—osteoblast (bone cell, low mobility, and low metabolism) and hepatocyte (liver cell, medium mobility, and intense metabolism).

There are no literature data about such intracellular velocities but, at least, their distributions follow a general meaning of cell physiology—more intense metabolism coincides with a higher mean and median of the velocity (Fig. 3). To compare the results of the described method with other methods, we estimated

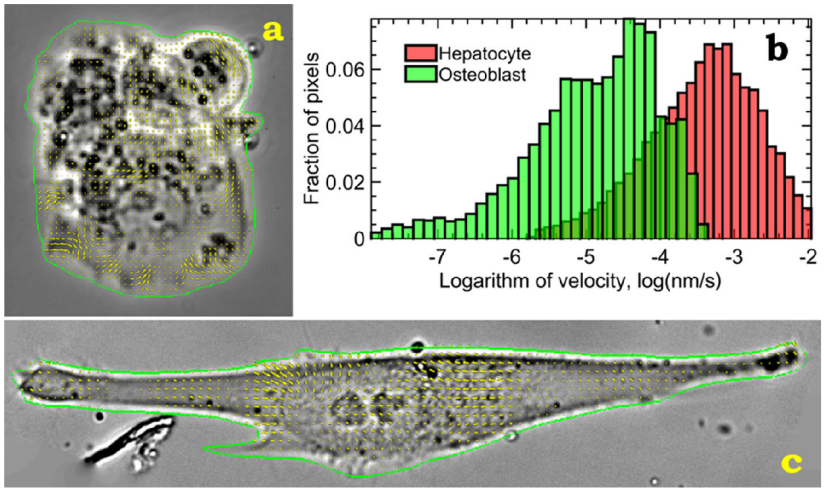

Fig. 3 The reconstructed global velocity field for a hepatocyte (**a**) and osteoblast (**c**). The corresponding velocity frequency histograms are shown in panel (**b**)

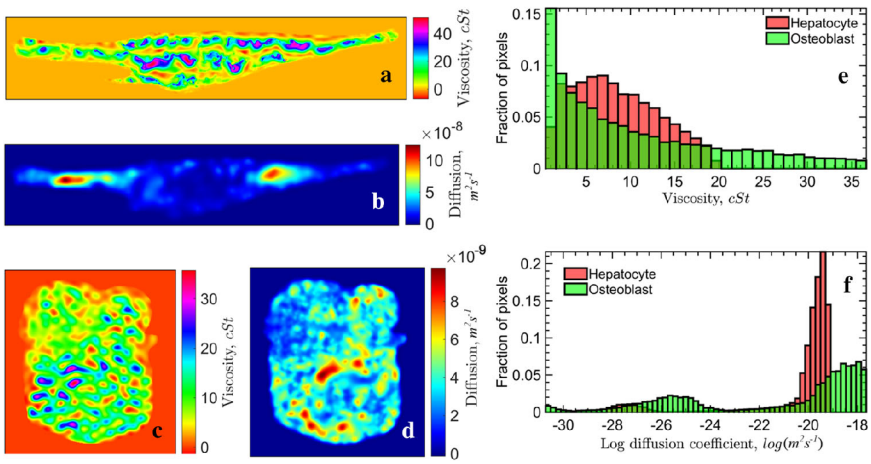

Fig. 4 The maps of intracellular effective diffusion and viscosity coefficients for a hepatocyte (**c**, **d**) and osteoblast (**a**, **b**). The relevant frequency histograms of the viscosity and diffusion coefficients are in panels (**e**, **f**)

the hydromechanical parameters of the intracellular medium. The proposed separation procedure yields a local standard deviation of the random walk-like process, which can be naturally converted to a effective diffusion coefficient (Fig. 4b, c). But any comparison with other results is complicated, because most of the diffusion coefficients are determined for molecules but we presumably observe motion of larger intracellular structures.

The obtained effective diffusion coefficients are in the range 10^{-10}–10^{-8} m^2 s^{-1} and correspond to values for particles in liquids [28]. The resulted coefficients may be related to both active and passive diffusion. Namely, the diffusion map of the osteoblast is very inhomogeneous but this has no relation to the velocity distribution (cf. hepatocyte in Figs. 3b and 4f). In the osteoblast's interior, there are two sites with very high diffusion coefficients (likely active diffusion) and the central region of low diffusion. This central region roughly corresponds to the position of nucleus (as guessed from the typical structure of osteoblasts; in the raw images, nucleus is not observed at all, because the microscope was focused on the cell surface).

The kinematic viscosities for both cells are in the range 5–50 cSt, which is comparable with palm oil and other viscous substances. The dispersion of viscosity for the osteoblast is much higher, but there is no much explanation for this. The resulted viscosity fields are quite noisy, since the numerical estimation of the 3rd derivative is a quite sensitive process. Surprisingly, the values are meaningful even without advanced smoothing. However, for in-depth analysis of the maps, we definitely need a more sophisticated processing. However, we observe only a planar slice of a 3D system and the equations here were derived for 2D. Thus, the obtained viscosity is rather effective than true, physical. Nevertheless, it is possible to compare the values of this quasi-viscosity between similar experiments; or do extensive validation and find a correction factor to obtain real kinematic viscosity and conditions, where such a explicit continuous mapping exists. Despite all the facts, a single plane derived viscosity has a reason-

able scaling, and thus, may be compared with other viscosities, but with caution.

The main advantage of the intracellular rheology estimation method described in this paper is its simplicity. As seen in this paper, the algorithm works with time-lapse image series of unstained living cells in any bright-field microscope (we show independent results for time-lapse series from two different bright-field microscopes, see Sect. 2). Nevertheless, let us note that this method can be applied in analysis of fluorescent image data. If applied, the complete analysis of flows in the stained living cells would be simplified compared to the bright-field data (due to a lower number of the possibly detected and tracked points and their identification). However, the biological relevance of such results is debatable, since the fluorophores can be cytotoxic and can completely change cell metabolism and dynamics. Thus, only autofluorescence plays an important and obvious role in interpretation of the intracellular dynamics.

In addition, the algorithm described here does not require any a priori given constant or assumptions about processes in the sample. Moreover, we have studied only one semi-tomographic slice of an active, unstained, 3D object, which can make the biologically relevant interpretation even more tricky. At least we know that the described values are sufficiently stable, and therefore, can be used for cell characterization. The conducted experiments are rather illustrative than explorative. We have not so far dealt with linking the results to biology but, compared with the literature, e.g., [27,29,30], they seem to be promising.

5 Conclusions

Better understanding of a cell behavior is one of the major task of modern biology and key to very important technologies such as growing artificial tissues and organs, or fighting against cancer. In such challenging tasks, biologists will need as many reinforcements as possible. In addition, this method, among others, is aimed to bring physicists, data scientists, and mathematicians to life sciences; and make a shortcut between classical, wet, biology and formidable machinery of modern data explanatory analysis and machine learning. Therefore, the approach is quite minimalistic. For application, one needs only a video with living cells and knowledge of a camera sensor geometrical size. The outputs of the method are physically understandable and interpretable parameters. But the origin of such flows and the overall cell fluid dynamics is a different story, and hopefully, will be solved in the meantime.

Acknowledgements This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic project CENAKVA (LM2018099), GAJU project 013/2019/Z, and from the European Regional Development Fund in frame of the projects Kompetenzzentrum MechanoBiologie (ATCZ133) and ImageHeadstart (ATCZ215) in the Interreg V-A Austria-Czech Republic programme. The authors would like to thank Petr Macháček (Image Code company, Brloh, Czech Republic) for software development, Petr Tax (Optax company, Prague, Czech Republic) for custom microscope development, and Miroslav Slivoně (a USB student) for the HepG2 microscopy data acquisition.

References

1. F. Buggenthin, C. Marr, M. Schwarzfischer, P.S. Hoppe, O. Hilsenbeck, T. Schroeder, F.J. Theis, BMC Bioinf. **14**, 297 (2013)
2. A. Boquet-Pujadas, T. Lecomte, M. Manich, R. Thibeaux, E. Labruyère, N. Guillén, J.C. Olivo-Marin, A.C. Dufour, Sci. Rep. **7**, 9178 (2017)
3. J.C. Crocker, B.D. Hoffman, in *Methods in Cell Biology* (Elsevier, 2007), pp. 141–178
4. R. Rychtáriková, D. Štys, Observation of dynamics inside an unlabeled live cell using a bright-field photon microscopy: Evaluation of organelles' trajectories, in Bioinformatics and Biomedical Engineering (IWWBIO 2017) (Springer International Publishing, 2017), pp. 700–711
5. A. Melling, Meas. Sci. Technol. **8**, 1406 (1997)
6. B. Lüthi, A. Tsinober, W. Kinzelbach, J. Fluid Mech. **528**, 87 (2005)
7. K. Lonhus, R. Rychtáriková, G. Platonova, D. Štys, Sci. Rep. **10**, 18346 (2020)
8. D. Štys, T. Náhlík, P. Macháek, R. Rychtáriková, M. Saberioon, *Least Information Loss (LIL) conversion of digital images and lessons learned for scientific image inspection*, in *Bioinformatics and Biomedical Engineering (IWBBIO 2016)* (Springer International Publishing, 2016), pp. 527–536
9. *Recommendation ITU-R BT.601-7 (2/2011): Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios* (2017). https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDFE.pdf
10. X. Jiang, H. Bunke, K. Abegglen, A. Kandel, *Curve morphing by weighted mean of strings, in Object recognition supported by user interaction for service robots*, vol. 4 (2002), pp. 192–195
11. J.P. Thirion, Med. Image Anal. **2**, 243 (1998)

12. Matlab code and image data to "Estimation of rheological parameters for unstained living cells" (2020). https://doi.org/10.5061/dryad.v15dv41t8
13. J. Li, N. Allinson, Neurocomputing **71**, 1771 (2008)
14. A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N.I. Ratyal, B. Zafar, S.H. Dar, M. Sajid, T. Khalil, Math. Probl. Eng. **2019**, 1 (2019)
15. S. Leutenegger, M. Chli, R.Y. Siegwart, *BRISK: Binary Robust invariant scalable keypoints*, in *2011 International Conference on Computer Vision* (IEEE, 2011)
16. E. Rublee, V. Rabaud, K. Konolige, G. Bradski, *ORB: An efficient alternative to SIFT or SURF*, in *2011 International Conference on Computer Vision* (IEEE, 2011)
17. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L.V. Gool, Int. J. Comput. Vis. **65**, 43 (2005)
18. P.F. Alcantarilla, A. Bartoli, A.J. Davison, *Computer Vision—ECCV 2012* (Springer, Berlin Heidelberg, 2012), pp. 214–227
19. J. Shi, *Tomasi, Good features to track, in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94* (Press, IEEE Comput. Soc, 1994)
20. H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, Comput. Vis. Image Underst. **110**, 346 (2008)
21. J.V. Burke, S.P. Han, Math. Program. **43**, 277 (1989)
22. D. Garcia, Comput. Stat. Data Anal. **54**, 1167 (2010)
23. D. Garcia, Exp. Fluids **50**, 1247 (2010)
24. E. Rossi, A. Colagrossi, G. Graziani, Comput. Math. Appl. **69**, 1484 (2015)
25. P.W. Holland, R.E. Welsch, Commun. Stat. Theory Methods **6**, 813 (1977)
26. M.K. Kuimova, S.W. Botchway, A.W. Parker, M. Balaz, H.A. Collins, H.L. Anderson, K. Suhling, P.R. Ogilby, Nat. Chem. **1**, 69 (2009)
27. W.C. Parker, N. Chakraborty, R. Vrikkis, G. Elliott, S. Smith, P.J. Moyer, Opt. Express **18**, 16607 (2010)
28. M. He, S. Zhang, Y. Zhang, S.G. Peng, Opt. Express **23**, 10884 (2015)
29. E.O. Puchkov, Biochem. (Mosc.) Suppl. Ser. A Membr. Cell Biol. **7**, 270 (2013)
30. J. Dench, N. Morgan, J.S.S. Wong, Tribol. Lett. **65**, 25 (2016)

CHAPTER 5

Curriculum vitae

Ali GHAZNAVI
Artificial Intelligence Engineer, Data Scientist

www.linkedin.com/in/ali-ghaznavi-727297145/
+420 775 698 858 a.ghaznavi@outlook.com
798/19, Studentska 20, Ceske Budejovice, Czech Republic
github.com/AliGhaznavi1986 Researchgate

Data scientist and computer programmer, with a various experience in predictive modelling and data analysis in business and scientific domain. I have leverage knowledge in image analytic based on my PhD research and studies in AI. Highly skilled in in different disciplines including deep neural network, machine learning, image processing, remote sensing and data visualization. Very eager to expand my knowledge in artificial intelligence fields to pursue my professional career by researching and working in this interesting fields.

EDUCATION

2019–2023 **PhD student in Biophysics**, University of South Bohemia, Czech Republic – Will graduate till 26 of June 2023
Thesis Title : Cell segmentation from wide-field light microscopy images using different variant of CNNs.
Supervisor : Prof. Dalibor stys

2013–2016 **M.Sc. in Artificial Intelligence**, Azad Qazvin University, Qazvin, Iran
Total GPA : 15.60 /20
Thesis Title : Image object retrieval based on optimized representation extracted from region base visual and textual feature – **Grade :** 17.5 /20
Supervisor : Dr. Amir Masoud Eftekhari

2006–2012 **B.Sc. in Computer Software Engineering**, Payam Noor University, Parand, Iran,
Total GPA : 16.74 /20
Thesis Title : Research based on RFID systems – **Grade :** 20 /20
Supervisor : Dr. Mostafa Kishani

PROFESSIONAL EXPERIENCE

May 2022 – February 2022 **Data analysis, BOSCH COMPANY, Ceske Budejocie, Czech Republic**
- Data analysis with regression methods
- Binary classification
- Applied deep learning methods for regression and classification model training
- Develop and implement algorithms based on Python platform with Keras and Tensorflow

Machine learning | Deep learning | AI | logistic regression | TensorFlow | Keras | Scikit-learn | data transforming

December 2022 – September 2022 **Visiting Researcher under HiDA data science fellowship program, GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES, POTSDAM, Germany**
- Satellite data analysis
- Remote Sensing data validation
- Applied Machine/hybrid deep learning methods for mapping global inland waters studies
- Develop and implement algorithms based on Python platform with Keras and Tensorflow

Image processing | Machine learning | Deep learning | CNN | AI | Inland Water detection and segmentation | TensorFlow | Keras | Scikit-learn | OpenCV | SQL | Google Earth engine

January 2022 – October 2021 **PhD Internship as Researcher, GFZ GERMAN RESEARCH CENTRE FOR GEOSCIENCES, POTSDAM, Germany**
- Principal Investigator in EJP-STEROPES
- Remote sensing data analysis
- Quantification of soil organic carbon using stacked auto-encoder feature extraction and deep learning techniques
- Develop and implement algorithms based on Python platform with Keras and Tensorflow

Signal processing | Soil Organic Carbon Monitoring | Machine learning | Deep learning | AI | TensorFlow | Keras | FCN | Auto Encoder | CNN | svm | random forest

Present
February 2019

Research assistant and lab technician – part time | Institute of Complex systems , University of South Bohemia in Ceske Budejovice, Czech Republic

- Application of image processing and machine learning in transmitted bright–field microscopy images
- Cell and tissue detection and semantic segmentation
- Applied Deep learning methods in bright field microscopy images
- Unique bright field microscopy dataset labeling and preparation
- Develop and implement method for single class semantic and instance Hela living cell segmentation from transmitted bright–field microscopy images

Image processing | Machine learning | Deep learning | Model development | U-Net | Data handling | Residual Attention U-Net | TensorFlow | Keras | Google Colab

July 2022
May 2022

Summer School supervisor | Institute of Complex systems , University of South Bohemia in Ceske Budejovice, Czech Republic

- application of Deep learning methods in reflective bright–field microscopy images
- Categorical cell segmentation
- Multi class data set labeling and preparation
- Develop and implement deep learning method for Multi class MG63 living cell segmentation from reflective bright–field microscopy images

Machine learning | Deep learning | Model development | Data handling | ResNet | U-Net | Vgg19 | Inception | Python | Keras | TensorFlow

October 2018
September 2016

Data Specialist, Mando Company, Tehran, Iran

- Classifying and analysing datasets related with Auto Industry companies with Machine Learning and Data Mining Modeling, Regression and Classification methods.

Data Mining | Regression | Machine learning | Data handling | SPSS | Matlab

Januaray 2016
Januaray 2013

Computer Software Engineer | Paliz Sanat Pars Company, Tehran, Alborz, Iran

- Collaborating with senior engineers to establish projects goal and deadlines.
- Programming solution, troubleshooting and developing and debugging the scripts based on the Python and MATLAB programming language

Image processing | Matlab | Programming | Supervise and unsupervise learning | Data mining | IBM SPSS

Publications

2022 **Ghaznavi, A.**, Rychťariková, R.,Saberioon, M., Stys, D.:Cell segmentation from telecentric bright-field transmitted light microscopic images using a Residual Attention U-Net : a case study on HeLa line. Computers in Biology and Medicine. 10.1016/j.compbiomed.2022.105805

2020 Lonhus, K., Rychtáriková, R., **Ghaznavi, A.**, Stys, D : Estimation of rheological parameters for unstained living cells. The European physical journal special topics – 2021. 10.1140/epjs/s11734-021-00084-2

Per-Review **Ghaznavi, A**, Rychťariková, R., Cisar P., Ziaei M.M., Stys, D .:Hybrid deep-learning multi-class segmentation of HeLa cells in reflected light microscopy images. Under review at Biomedical Signal Processing and Control.

Per-Review **Ghaznavi, A**, Saberioon, M, Brom j, Itzerott, S .:Comparative Performance Analysis of simple U-Net, Residual Attention U-Net, and VGG16-U-Net for Inventory Inland Water Bodies. In review at Remote Sensing, MDPI.

Per-Review Mohammadmehdi Saberioon, Asa Gholizadeh, **Ali Ghaznavi**, Sabine Chabrillat, Kathrin J. Ward,:Soil organic carbon modeling using open-access soil spectroscopy libraries and machine learning algorithms. Under review at Computers and Electronics in Agriculture.

Publication available : Researchgate

Languages

Persian	●●●●●
Turkish	●●●●●
English	●●●●○
Czech	●○○○○
German	●○○○○

Programming Languages

- Python (Since 2019)
- MATLAB (Since 2014)
- IBM SPSS (Since 2015)
- Shell (Since 2022)

Research Interests

- Machine learning
- Deep Neural Networks (DNN)
- Computer Vision
- Object detection and segmentation
- Remote Sensing data analysis
- Data Visualization
- Fuzzy Systems
- Statistical Data analysis
- Big Data Analytics
- Information and Image Retrieval
- IBM Bioinformatics
- google map engine

Skills and Package

- Python
- Matlab
- TensorFlow–Keras
- Scikit-learn
- OpenCv
- Pandas
- SciPy
- Google Colab
- PyTourch
- AWS
- Git
- Big Data

Honors and Awards

2022 Recipient of HiDA data science Helmholtz Visiting Researcher fellowship grant from Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Germany

2021 Recipient of fellowship for PhD internship from Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany

2016 Outstanding student research from Azad Qazvin University (QIAU), Iran

2013 Rank 26^{th} among 2400 in university entrance exam for Master Degree program, Qazvin Azad University (QIAU), Iran

Dataset

2022 **Ghaznavi A.**, Rychtáriková R., Saberioon M., Štys D. Telecentric bright-field transmitted light microscopic dataset.
datadryad Repo.

References

Prof. RNDr. Dalibor stys, CSc.
University of South Bohemia,
@ stys@frov.jcu.cz
+420 38 777 3843

Dr. Mohammadmehdi Saberioon
GFZ German Research Centre for Geosciences,
@ mohammadmehdi.saberioon@gfz-potsdam.de
+49 331 288-27539

ghaznavi@jcu.cz

Title: Cell segmentation from wide-field light microscopy images using CNNs.
Ph.D. Thesis Series, 2023, No. 7.

Printed in the Czech Republic by Typodesign
Edition of 10 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: sekret-fpr@prf.jcu.cz