TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies

Anomaly Detection in Galaxy Images using Deep Learning

Master Thesis

Study programme: N0714A150003 Mechatronics
Author: **Karina Batalova**
Thesis Supervisors: Ing. Karel Paleček, Ph.D.
Institute of Information Technology and Electronics

Liberec 2022

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies

Master Thesis Assignment Form

Anomaly Detection in Galaxy Images using Deep Learning

Name and surname: **Karina Batalova**
Identification number: M21000199
Study programme: N0714A150003 Mechatronics
Assigning department: Institute of Information Technology and Electronics
Academic year: **2021/2022**

Rules for Elaboration:

1. Study and theoretically compare approaches based on covolutional autoencoders and generative adversarial networks for detection of anomalies in astronomical images.
2. Design and implement a deep neural network model from one of the above classes in PyTorch library and Python programming language.
3. Empirically evaluate your model on a suitable dataset of astronomical objects, especially galaxies.
   Describe advantages and limitations of your model. If available, compare your results with existing solutions.

Scope of Graphic Work:	
Scope of Report:	
Thesis Form:	printed/electronic
Thesis Language:	English

List of Specialised Literature:

[1] Goodfellow, I., Bengio, Y., Courville, A. Deep learning. MIT Press, 2016
[2] Bishop, C. Pattern Recognition and Machine Learning. 2006. ISBN 13: 978-038731073
Karpathy, A., Johnson, J., Li, F. Convolutional neural neworks for visual recognition. dostupné online: http://cs231n.stanford.edu

Thesis Supervisors:	Ing. Karel Paleček, Ph.D. Institute of Information Technology and Electronics
Date of Thesis Assignment:	October 12, 2021
Date of Thesis Submission:	May 16, 2022

L.S.

prof. Ing. Zdeněk Plíva, Ph.D.
Dean

prof. Ing. Ondřej Novák, CSc.
head of institute

Liberec October 19, 2021

Declaration

I hereby certify, I, myself, have written my master thesis as an original and primary work using the literature listed below and consulting it with my thesis supervisor and my thesis counsellor.

I acknowledge that my master thesis is fully governed by Act No. 121/2000 Coll., the Copyright Act, in particular Article 60 – School Work.

I acknowledge that the Technical University of Liberec does not infringe my copyrights by using my master thesis for internal purposes of the Technical University of Liberec.

I am aware of my obligation to inform the Technical University of Liberec on having used or granted license to use the results of my master thesis; in such a case the Technical University of Liberec may require reimbursement of the costs incurred for creating the result up to their actual amount.

At the same time, I honestly declare that the text of the printed version of my master thesis is identical with the text of the electronic version uploaded into the IS/STAG.

I acknowledge that the Technical University of Liberec will make my master thesis public in accordance with paragraph 47b of Act No. 111/1998 Coll., on Higher Education Institutions and on Amendment to Other Acts (the Higher Education Act), as amended.

I am aware of the consequences which may under the Higher Education Act result from a breach of this declaration.

May 23, 2022 Karina Batalova

Anomaly Detection in Galaxy Images using Deep Learning

Abstrakt

Tato práce představuje výzkum v oblasti detekce anomálií pomocí generativních modelů. Hlavním účelem této práce je zhodnotit výkon metod detekce hlubokých anomálií pro obrazy galaxií.Byla nalezena metoda s dobrým výkonem detekce anomálií pro astronomické obrazy.

Klíčová slova: Detekce hlubokých anomálií, generativní sítě, obrazy galaxií

Anomaly Detection in Galaxy Images using Deep Learning

Abstract

This work presents research in the field of anomaly detection using generative models. The main purpose of this work is to evaluate performance of deep anomaly detection methods for galaxy images. A method with good anomaly detection performance for astronomical images was found.

Keywords: Deep anomaly detection, Generative Networks, Galaxy images

Acknowledgements

I would like to express my deepest appreciation to my thesis supervisor Ing. Karel Paleček, Ph.D. at the Technical University of Liberec for his patience, guidance, and support.

I am extremely grateful to all the teachers for the knowledge I have gained during my studies at the Technical University of Liberec. I am also grateful to Ing. Lukáš Hubka, Ph.D. for his support and assistance during the entire study.

I would like to extend my sincere thanks to the prof. Ing. Jaroslav Nosek, CSc. with the leadership of the Technical University of Liberec and Erasmus+ program for the exceptional opportunity to study at the great Czech university.

Contents

List of abbreviations . . . 8

Introduction **9**

1 Overview of existing solutions **10**

1.1 Generative Adversarial Networks . . . 10

1.1.1 Anomaly detection in Hyper Suprime-Cam galaxy images with generative adversarial networks (AD-HSC) . . . 10

1.1.2 Detecting outliers in astronomical images with deep generative networks . . . 11

1.2 Convolutional autoencoder . . . 11

1.2.1 Anomaly detection in Astrophysics: a comparison between unsupervised Deep and Machine Learning on KiDS data . . . 11

2 Anomaly Detection **13**

2.1 Comparison of generative models . . . 13

2.1.1 GAN . . . 15

2.1.2 CAE . . . 19

3 Tested architectures **24**

3.1 GANomaly . . . 24

3.1.1 GANomaly architecture . . . 24

3.1.2 GANomaly training . . . 25

3.1.3 GANomaly testing . . . 26

3.2 f-AnoGAN . . . 27

3.2.1 f-AnoGAN architecture . . . 27

3.2.2 f-AnoGAN training . . . 28

3.2.3 f-AnoGAN testing . . . 29

4 Data **31**

5 Experiments **34**

5.1 Evaluation metrics . . . 34

5.2 Training and Anomaly detection . . . 37

5.2.1 GANomaly . . . 38

5.2.2 f-AnoGAN . . . 38

5.2.3 Results 39

Bibliography **44**

Appendix **45**
A. Results for each class of galaxy 45

List of abbreviations

GAN	Generative Adversarial Networks
G	Generator
D	Discriminator
AE	Autoencoders
WGAN	Wasserstein Generative Adversarial Network
HSC	Hyper Suprime-Cam
CAE	Convolutional autoencoder
CNN	Convolutional Neural Network
DCA	Disentangled Convolutional Autoencoders
KiDS DR	Kilo Degree Survey Data Release
MAD	Median Absolute Deviation
RNN	Recurrent Neural Networks
PCA	Principal Component Analysis
DCGAN	Deep Convolutional Generative Adversarial Network
CGAN	Conditional Generative Adversarial Network
MSE	Mean Squared Error
SSIM	Structural Similarity Index
ReLU	Rectified Linear Unit
DECam	Dark Energy Camera
DESI	Dark Energy Spectroscopic Instrument
E	Elliptical
S	Spiral
So	Lenticular
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
AUC	Area Under The Curve
ROC	Receiver Operating Characteristics
TPR	True Positive Rate
FPS	False Positive Rate

Introduction

Many astronomical researches aim to identify outliers in the information collected by various modern telescopes. There are various methods for finding outliers also known as anomaly detection methods. These methods are based on the theory of normal data distribution. Anomalies are data that do not follow the rules of normal distribution.

The amount of data obtained by astronomical surveys increases exponentially in volume and complexity every year. To analyze such a large amount of data, it is impossible to process only with the help of human efforts. Modern methods of anomaly detection are needed. When the data that is being processing by these methods increases and becomes more complex, deep learning methods have been proposed to solve this problem.

The application of deep learning for anomaly detection is increased rapidly, providing for the study of this area and the development of new architectures of neural networks. Generative models are one of the most popular trends. The ability to learn the distribution of data by optimizing a feature learning loss function has given generative models a widespread in the problem of anomaly detection.

At the moment, there are a large number of generative models with improved architecture for detecting anomalies. Training and testing these architectures to assess their performance in specific tasks, such as detecting anomalies in galaxy images, can specify the right direction in improving and creating modern applications for solving this problem.

The purpose of this work is to test the ability of generative models to detect anomalies in galaxy images. This work theoretically compares two popular generative models for detecting anomalies in images: Generative Adversarial Networks and Convolutional autoencoder. According to the preliminary analysis, two GAN-based architectures were chosen to solve given problem: GANomaly and f-AnoGAN. These models are specialized in detecting anomalies using only normal data for training. The performance of the models is evaluated using standard evaluation metrics of anomaly detection: AUC-ROC and F1-score.

1 Overview of existing solutions

For anomaly detection in a large amount of data, it is rational to use deep learning, since classical methods of machine learning require a lot of work at the stage of data preparation. One of the most promising ways is to use generative models. In the field of astronomical data processing, this approach provides high efficiency.

The basic goal of the Generative models is to learn true data distribution in order to generate new data points with some variations. In general, the two most popular and efficient generative approaches:

- Generative Adversarial Networks (GAN);
- Autoencoders (AE).

At the moment, there are several modern solutions for detecting anomalies in astronomical images using generative models.

1.1 Generative Adversarial Networks

1.1.1 Anomaly detection in Hyper Suprime-Cam galaxy images with generative adversarial networks (AD-HSC)

Method: Wasserstein generative adversarial network (WGAN)
Data: Hyper Suprime-Cam galaxy images

Technology **AD-HSC**[13] based on an unsupervised anomaly detection method using a Wasserstein generative adversarial network (WGAN). Galaxy images obtained from the Hyper Suprime-Cam (HSC) survey (the second public data release) were used to train this model.

For the model the standard formulation of a Wasserstein GAN with gradient penalty (WGAN-GP) is used. The generator and the discriminator are convolutional neural networks.

The authors use a discriminator score based on the residual between the original and the reconstruction to detect anomalies that distinguish their approach from the classic one, which uses the generator and discriminator losses.

Among other things, this solution uses post-processing of anomalous images to find potentially scientifically interesting images. For this purpose, convolutional autoencoder (CAE) is applied in the work. Presumably, the residual images contain information about why WGAN marked the object as abnormal. In this solution CAE is used to reduce the dimension of data and highlight significant information. This network contains 4 encoding and decoding layers, uses a standard MSE loss between the real and reconstructed image.

This approach combining WGAN anomaly detection and CAE post-processing allows obtaining a number of potentially scientifically meaningful images of galaxies.

1.1.2 Detecting outliers in astronomical images with deep generative networks

Method: Wasserstein generative adversarial network (WGAN)
Data: Horizon-AGN astronomical images

This work focuses on the ability of deep generative networks for detecting outliers in astronomical image datasets [4].

For this research, the Horizon-AGN cosmological hydrodynamical simulation data and images from the CANDLES survey are used.

The solution is based on the WGAN theory with Wasserstein-1 distance as the metric to evaluate the similarity between a real and a generated distribution. The implementation is CNN architectures for the discriminator and the generator.

The anomaly detection method is similar to the previous solution. The idea is that the generator is able to create images identical to the real image from the sample. Thus it will not be possible to reconstruct a duplicate image when an image is from the distribution of abnormal images. Therefore, the abnormal image will have a bigger loss than normal images. An anomaly score is defined as the loss at the last iteration when the training has converged and the closest image has been found. The total loss is defined as the sum of a residual loss and a generator loss.

This solution was tested with isolated galaxy images. The paper showed that the WGAN defines 80 percent of the test images as anomalous and only 10 percent of the samples are falsely determined as anomalous.

1.2 Convolutional autoencoder

1.2.1 Anomaly detection in Astrophysics: a comparison between unsupervised Deep and Machine Learning on KiDS data

Method: Disentangled Convolutional Autoencoders (DCA)
Data: Kilo Degree Survey Data Release (KiDS DR4)

Two anomaly detection methods (a Disentangled Convolutional Autoencoder and an Unsupervised Random Forest) are reported in this paper to address to compare but only one of them relates to the generative model. They are considered these methods as potentially promising methods to detect peculiar sources. The DCA method has been considered potentially capable of detecting peculiar objects like interacting galaxies and gravitational lenses.

The authors performed research on Kilo Degree Survey Data Release (KiDS DR4) data.

The main idea is that the residual images obtained by subtracting the decoder output signal from the corresponding input should contain only residual noise. Thus the statistical estimators calculated on the residual images are fundamental for assessing anomalies. The authors use three statistical estimators:

- Median Absolute Deviation (MAD). In the residual image, a median of the pixel values corresponds approximately to the mean value of the background. Therefore, MAD is a valid measure of how broadly the residuals are distributed around the background. A high value could indicate the presence of substructures or artifacts.

- Skewness. Extraordinary high or low values of this statistical parameter are able to indicate that there is something odd in the image.

- Maximum. Image artifacts produce the pixels with high brightness in the residual images.

DCA showed good results in memory and computing time and was able to detect some peculiar sources, showing substructures that were hidden by nearly distant galaxy light, as well as objects with extremely small or pale close companions.

2 Anomaly Detection

Anomaly detection is the identification of data that inconsistent to the distribution of normal data, i.e. does not correspond to normal appearance, semantic content, quality or expected behavior. The scope of the anomaly search can be the detection of network intrusions, data quality monitoring, detection of abnormal behavior on video, monitoring the quality in manufacturing, detection of anomalies on X-rays and much more.

The conditions for evaluation anomalies have two main aspects:

1. The normal samples have similar feature distribution in latent space.
2. The distribution of abnormal samples has a large distance from the normal data.

2.1 Comparison of generative models

The classic way of classification using machine learning is to label a dataset and use a neural network capable of returning the probability of belonging to a particular class. In the field of processing large databases, this approach requires a lot of work of experts at the stage of data preparation and interpretation of the results of the classifier requires additional work.

An alternative approach is to use deep neural networks which have been used to detect anomalies in various industries and demonstrate good results.

- Recurrent Neural Networks (RNN)
- Convolutional Neural Networks (CNN)
- Autoencoder (AE)
- Generative Adversarial Networks (GAN)

Basically, for most anomaly detection tasks, there are no balanced datasets in which there are abnormal samples in sufficient amounts, or they are completely absent. Thus, many deep learning methods face this existing barriers. However, some types of deep learning models have features that allow them to show good performance for detecting anomalies even to data with a large imbalance between abnormal and normal data.

The ability to fit the distribution give opportunity to generative models have become one of the best methods for detecting anomalies. Generative models learn the representations of data samples by optimizing a feature learning objective function. This model is not originally developed for anomaly detection. However, they capture some key patterns underlying the data, which contributes to the detection of anomalies. The basic goal of the generative models is to learn true data distribution to generate new data points.

The two most popular and indicative generative approaches are

1. Autoencoders (AE)
2. Generative Adversarial Networks (GAN)

Autoencoders are popular in the field of anomaly detection. Autoencoders is nearly equivalent to Principal Component Analysis **(PCA)**[16]. PCA is constrained to a linear dimensionality reduction, autoencoders include linear and nonlinear transformations. It is more efficient to train multiple layers using an automatic encoder, rather than training one huge transformation using PCA. Thus, autoencoder methods show their advantages when data problems are complex and nonlinear in nature. Autoencoders represent data within hidden layers by reconstructing the input data. The autoencoders which are trained using only normal data samples produce a large reconstruction error and that is an indicator of the anomalous data samples. The types of autoencoder architectures promising results in anomaly detection are proposed in Fig.2.1 . The selection of autoencoder architecture depends on the kind of data. For images, convolution networks are preferred. Autoencoders are simple and have effective architectures for anomaly detection. However, the performance is decreased when using low-quality and noisy training data with a high degree of distortion.

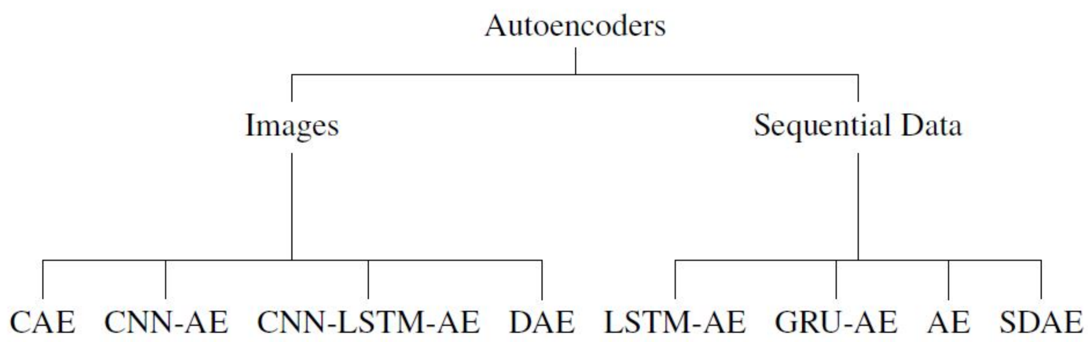

Figure 2.1: Autoencoder architectures for anomaly detection

GANs are introduced to effectively learn the input data distribution. Generative Adversarial Networks-based Anomaly Detection technologies (**AnoGAN**[19], **Efficient-GAN**[10], **GANomaly**[18],**f-AnoGAN**[20], **AD-HSC**[13], **AnoSeg**[12]) are provided efficiency for identifying anomalies on high-dimensional and complicated datasets due to the possibility of learning input data distribution. GAN exhibits good results in creating realistic images. It trains using only normal samples and learns the feature representations in a latent space which makes it possible to detect abnormal samples that have high residual errors.

2.1.1 GAN

Generative adversarial neural network **(GAN)** [11] is an architecture consisting of a generator and a discriminator configured to work against each other. Hence GAN got the name generative-adversarial. In the case of working with images, a convolutional neural network is used to design the generator and discriminator.

Standard GAN has two subnetworks: Generator and Discriminator.

- Generator. Generative algorithms model the distribution of individual classes.
- Discriminator. Discriminative algorithms attempt to classify input data. Taking into account the features of the data obtained, they try to determine the category to which they belong.

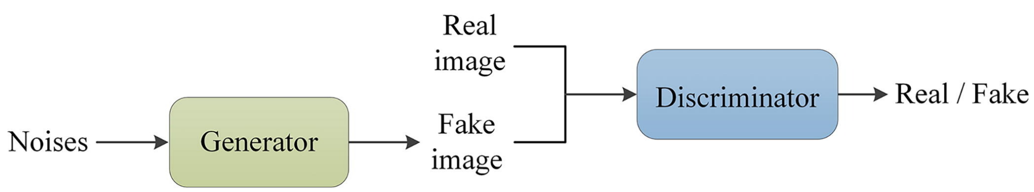

Figure 2.2: GAN structure

Generator creates new data instances, and the other, a discriminator, evaluates them for authenticity. The discriminator decides whether each data instance it considers belongs to a training data set or not.

Stages of GAN work:

- The generator gets a random query and returns an image.
- This generated image is fed into the discriminator along with a stream of images taken from the actual dataset.
- The discriminator accepts both real and fake images and returns probabilities, numbers from 0 to 1, with 1 representing a genuine image and 0 representing a fake one.

The discriminator network is a standard convolutional network that can classify images submitted to it using a binomial classifier that recognizes images as real or as fake. The generator is in some sense an inverse convolutional network: although a standard convolutional classifier takes an image and reduces its resolution to get a probability, the generator takes a random noise vector and converts it into an image. The first filters out the data using downsampling techniques such as maxpooling, and the second generates new data. Both networks are trying to optimize the target function or the loss function. For guiding the discriminator training, loss function is used:

$$\max V(D,G) = \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}(\mathbf{z})}}[\log(1-D(G(\mathbf{z})))]$$

where V denotes the output value of the loss function, G is a generator and D is a discriminator, P_data is the real data distribution, P_z is the generated data distribution, and E is the average value.

For guiding the generator training, loss function is used:

$$\min V(D,G) = \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}(\mathbf{z})}}[\log(1-D(G(\mathbf{z})))]$$

The GAN is trained with the following objective:

$$\min_G \max_D V(D,G) = \mathbf{E}_{\mathbf{x}\sim P_{\text{date}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}(\mathbf{z})}}[\log(1-D(G(\mathbf{z})))]$$

The vanilla GAN has some problems with training stability and mode collapse when the gradient of the discriminator vanishes. In order to make it possible to solve important tasks, it is necessary to compensate for the drawbacks of the vanilla GAN. Different variants of architecture and losses to decide these problems have been proposed. Several variations is suitable to decide the anomaly detection problems:

1. Deep convolutional generative adversarial networks **(DCGAN)**[3]. DCGAN is a GAN architecture with striped convolutions for the discriminator and fractional-stride convolutions for the generator. This model removes fully connected hidden layers and uses batch normalization. DCGAN employs ReLU in a generator and for output generator layer is tanh activation function. For all discriminator layers LeakyReLU activation is used.

2. Conditional Generative Adversarial Nets **(CGAN)**[14]. This is a modified version of the GAN algorithm, which can be constructed by transmitting additional data y which is a condition for the generator and discriminator. y can be any additional information, for example, a class label, an image, or data from other models, which can allow you to control the data generation process. Additional information y is supplied to the input of the generator and discriminator from GAN, for example, the condition can be represented by an additional input layer. (Fig. 2.3).

   In this case, the optimization problem will look like this:

   $$\min_G \max_D \underset{x\sim p_{\text{data}}}{E}[\log D(x \mid y)] + \underset{z\sim p_z}{E}[\log(1-D(G(z \mid y))]$$

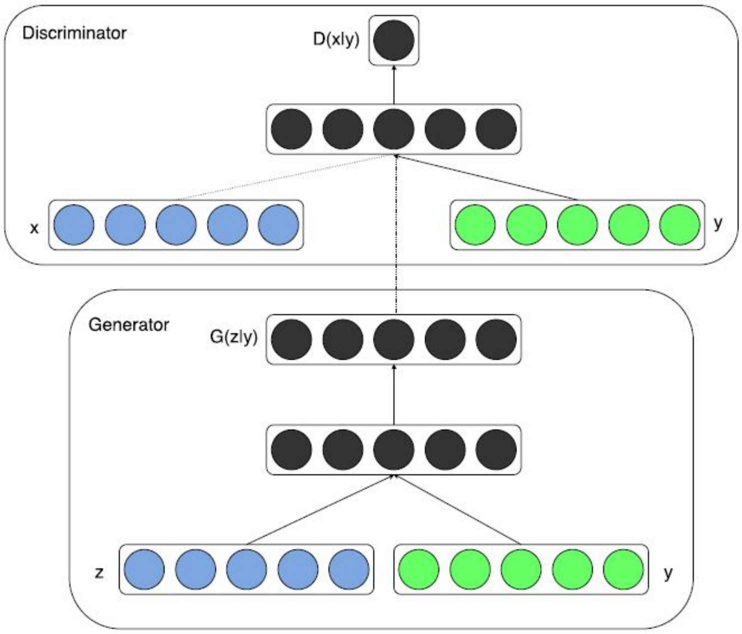

Figure 2.3: Conditional Generative Adversarial Net[14]

3. **InfoGAN**[21] devides the generator input vector z into two parts: an interpretable latent variable and an incompressible noise vector. The dimension of the latent variable corresponds to the semantics of the generated sample constraining the relation between them. InfoGAN enhances semantics control of sample generation due to entering information theory. This model adds a regularization term to the standard GAN loss function:

$$\min_G \max_D V_I(D,G) \quad = \min_G \max_D [V(D,G) - \lambda I(\mathbf{c}; G(\mathbf{z},\mathbf{c}))]$$

where $\mathbf{c}$ is the interpretable latent vector, $I(\cdot)$ means the computation of mutual information between the latent vector c and the generator output, and λ is regularization constant.

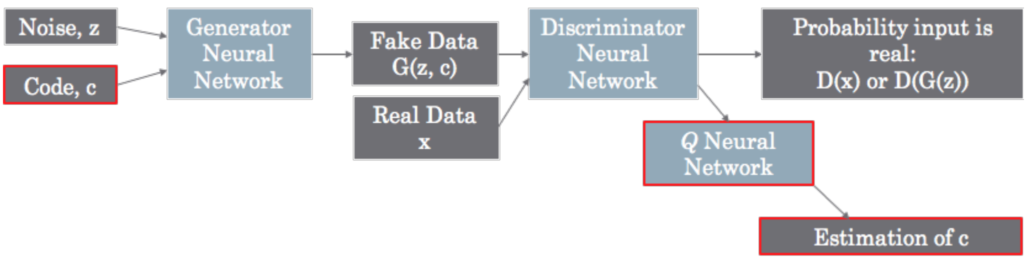

Figure 2.4: Information maximizing generative adversarial network. New components are indicated in red.

4. Wasserstein generative adversarial networks **(WGAN)**[4]. WGAN solves training instability and mode collapse problems that arise in vanilla GAN using minimizing the Wasserstein distance. There are several versions of WGAN. One of the most popular is WGAN with gradient penalty for generator.

   WGAN-GP loss function is defined as:

$$\min_G \max_D V(D,G) = \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[D(G(\mathbf{z}))] - \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[D(\mathbf{x})] + \lambda_{gp}\mathbf{E}_{\widehat{\mathbf{x}}\sim P(\widehat{\mathbf{x}})}\left[(\|\nabla_{\widehat{\mathbf{x}}}D(\widehat{\mathbf{x}})\|_2 - 1)^2\right]$$

GAN is constantly evolving especially its representational learning ability that makes this model a priority for anomaly detection. Mostly GAN-based anomaly detection methods train a network to learn the feature representation of normal samples. Deep Anomaly detection (DAD) review [8] concludes generic normality feature representation learning based on GAN (Fig. 2.5). The training data χ is forward-passed to the network to learn and derive a feature extractor ϕ. The loss function depends on the residual error or an anomaly measurement is used to detect anomalies. Since during the training period network adopts only normal instances all samples in the latent space store signs of only normal images, which means that such a network will restore normal images well, but as soon as there is an abnormal sample at the input, the optimization function will show large losses.

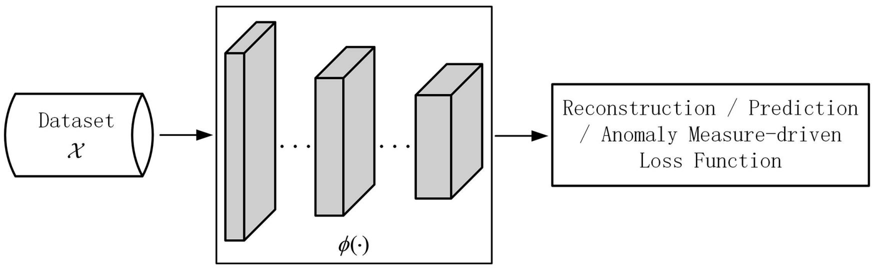

Figure 2.5: GAN-based generic normality feature representation learning. [8]

For anomaly detection, in general GAN-based method uses only normal data to train the network, learn its feature distribution and identify anomalies using a residual image to compare the reconstructed and given samples using generator, discriminator, or both. The probability output by the discriminator can be used as an anomaly score. The stability of the discriminator for anomaly detection is contentious since it can deteriorate after training.

The advantages and disadvantages of GAN-based anomaly detection approaches strongly depend on the chosen architecture and the objective functions. Chapter 3 details two popular approaches and their advantages and disadvantages, which are used by popular architectures for anomaly detection described in this section. The pros and cons are summarized here based on these approaches.

Advantages:

- GANs demonstrate excellent ability in generating realistic images, which helps to detect anomalies that are unsuccessfully reconstructed from the latent space.
- Most of the existing models and theories based on GAN can be adapted to detect anomalies.

Disadvantages:

- Vanilla GAN training is prone to problems such as failure to converge and collapse mode thus it is add difficulties and inaccuracy for anomaly detection. But there is already a solution for each of these problems and it is implemented in adapted architectures.
- The generator can be confused and generates images out of the manifold of normal data. This can happen if the data distribution is complex or outliers appear in the training data.
- Anomaly score may be suboptimal because it is based on a generator with a purpose intended for data generation rather than anomaly detection. This problem is also easily solved by making changes to the standard GAN.

2.1.2 CAE

Autoencoder (AE) is a neural network that copies input data to output. The model has similar in architecture to a perceptron. Autoencoders compress input data to represent them in latent space, and then restore output from this representation. The goal is to get the response closest to the input on the output layer. At the time when the autoencoder copies the input to the output, the latent space h obtains useful properties. Thus, autoencoders are able to extract the feature from the data. A distinctive feature of autoencoders is that the number of neurons at the input and output are the same. Autoencoders useful for various problems such as dimensionality reduction, noise cancellation, feature extraction or data generation.

Autoencoder consists of two parts:

- The encoder is responsible for compressing the input to latent-space. Transformation from original input to the latent representation is denoted by the encoding function $h = f(x)$.
- The decoder is designed to recover input from latent space. Transformation from latent representation to reconstructed output is denoted by the decoding function $r = g(h)$.

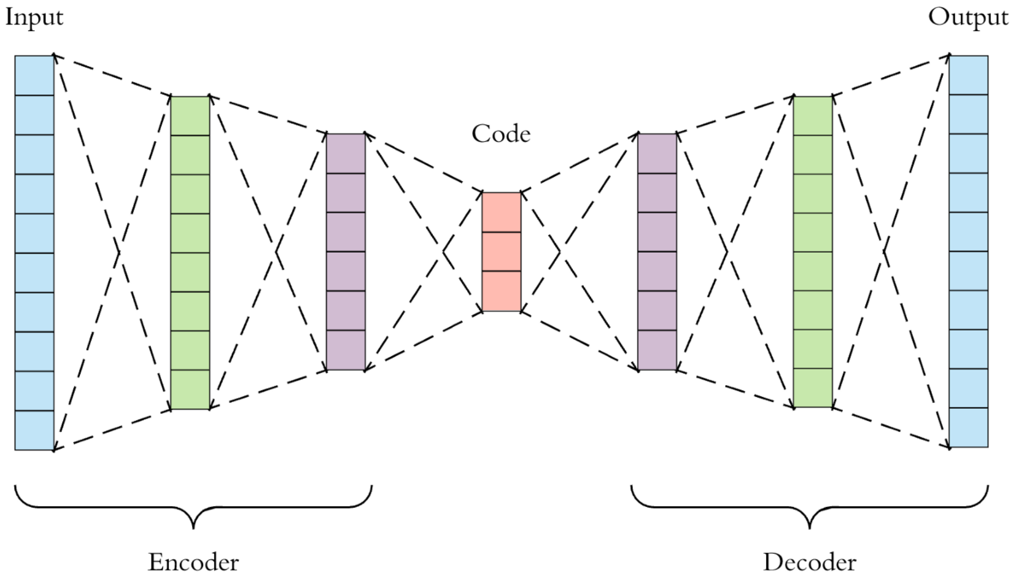

Figure 2.6: Autoencoder structure.

Encoder takes the raw data and maps to a latent or compressed representation a deterministic function. The obtained results is code which contains important information about input data. Using the code and a reverse mapping decoder reconstruct the input.

A convolutional autoencoder is a particular case of autoencoder which is employed for images and Convolutional Neural Networ (CNN) is used in the structure of the autoencoder in the encoding and decoding parts. The CNN structure allow to extract the features in the input image. This specificity gives opportunity to use autoencoders for classification and anomaly detection.

At the training stage only the normal data is used to train the CAE. Since it is necessary to optimize the CAE model to minimize the reconstruction error for the input and resulting image. Thus, the model will show high reconstruction error if the input image is anomalous. It minimizes the reconstruction error e between input $\mathbf{x}_i$ and output $\mathbf{y}_i$ by adjusting its parameters as shown by the loss function

$e(\mathbf{x}, \mathbf{y}, \mathbf{W}) = \frac{1}{2N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{y}_i\|_2^2 + \lambda \|\mathbf{W}\|_2^2,$

At the testing stage two main problems should be solved:

1. The metrics of the reconstruction error between the real and reconstructed images should be defined to compute the anomaly score.

2. The threshold (standard deviation, σ) of the reconstruction error should be determined in order to classify the input image as normal or anomaly data. (Fig. 2.8).

This several could be also used as the reconstruction error

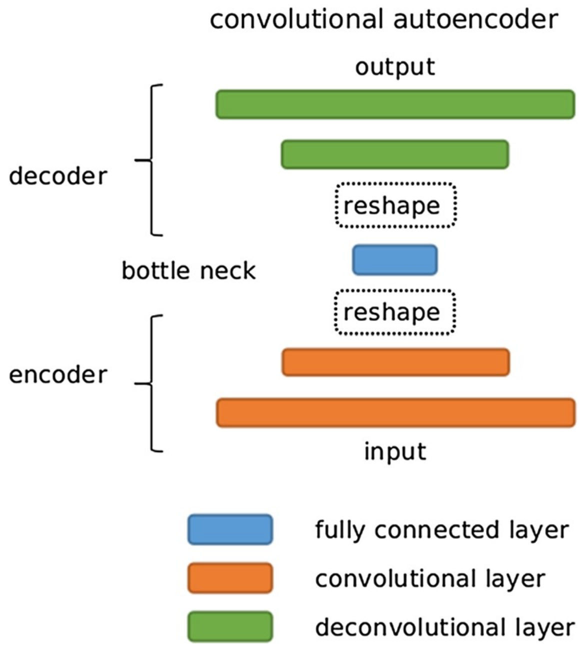

Figure 2.7: CAE stucture.

- Mean Squared Error (MSE)
  $MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$
- Structural Similarity Index (SSIM)
  $\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$
- Binary cross-entropy
  $H(P,Q) = -\sum_x P(x) \log Q(x)$
  $$L = -\sum_{i=1}^{2} t_i \log(p_i) = -[t \log(p) + (1-t)\log(1-p)]$$
  where t_i is the truth value taking a value 0 or 1 and p_i is the Softmax probability for the i^{th} class,

Advantages:

- CAE is simple and universal for various types of data.
- The CAE models and their improvements can be employed to perform anomaly detection.
- CAE is good at reconstruction of images from latent space.

Disadvantages:

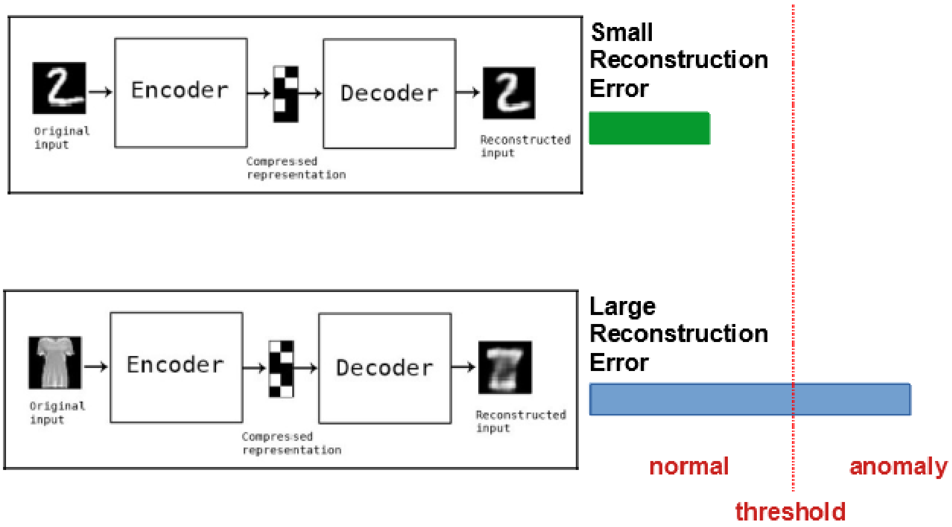

Figure 2.8: The concept of the reconstruction error for anomaly detection.

- The feature representations can be biased due to rare patterns outliers in the training samples.
- The loss function is more designed for dimension reduction, denoising, classification, or image compression than anomaly detection. For this reason, the resulting representations are a general presentation of the main patterns from which it is difficult to extract important and specific information from code. But this problem is being solved by replacing the decoder CNN with a specified function which creates a synthetic model of the input data.
- CAE produces really smooth images which absence the noise properties of the real data. This property is not suitable for astronomical images and anomaly detection.

Summarize

GAN is more appropriate for solving the problem of detecting anomalies in astronomical images compared to CAE.

1. GAN is primarily designed to learn the data distribution of the input images and it demonstrates excellent quality in the reconstruction of real images. That means during training GAN receives complete information in the normal data distribution, which allows him to identify outliers from it. CAE is designed to emphasize the most relevant and significant information from the image to compress it without considerable losses.

2. CAE smooths the input data which is a serious disadvantage for astronomical images. Since this feature of the model can erase subtle anomalies it is not suitable for anomaly detection in galaxy images.

3. At present GAN models are better suited for detecting anomalies, presumably because of the original purpose for which they were created. GAN models provide better performance and find more interesting anomalies since it allows to use of the feature space for quantitative assessment of anomalies.

3 Tested architectures

This chapter discuss two state-of-the-art methods more in depth. Two generative approach are described.These methods were selected since they are state-of-the-art in anomaly detection, at the same time have an accessible practical implementation and good performance in other problems.

3.1 GANomaly

GANomaly[18] is anomaly detection architecture based on a conditional generative adversarial network. This is a modified version of the GAN algorithm, which can be constructed by transmitting additional data, which is a condition for the generator and discriminator. The model learns the generation of high-dimensional image space and the inference of latent space.

The network is able to map the input image to a lower dimension vector, which is then used for reconstructing the generated output image fixing the distribution of training data both in the image and in the latent vector space by adversarial autoencoder within an encoder-decoder-encoder subnetwork for generator. In learning the data distribution for normal samples, it is important to minimize distances from these images and latent vectors during training. This is the reason for which, at inference time, a larger distance metric from this learned data distribution at inference time is indicative of an outlier from that distribution, that is, an anomaly. This adversarial training architecture such provides higher performance from both a statistical and computational point of view.

3.1.1 GANomaly architecture

GANomaly pipeline contains three sub-networks: autoencoder, encoder, and discriminator networks.

1. Autoencoder network that behaves as the generator G. Encoder G_E learns the input data representation and a decoder G_D reconstructs the input image.Network use convolutional layers followed by batch-norm and leaky Rectified Linear Unit (ReLU) activation function. Encoder of the generator part downscales input image x by compressing it to a latent vector $z = G_E(x)$ supposing the lowest dimension containing the best representation of input image. To upscales the vector and reconstruct image

the decoder of the generator part employs model of a **DCGAN**[3] generator which contains convolutional transpose layers, ReLU() and batchnorm and tangent activation functions. Thus Genrator generates image $\hat{x} = G_D(z)$.

2. Encoder network E to compress the image $\hat{x}$ which is reconstructed by the generator network. It has similar architectural details as encoder of the generator part with different parametrization. Encoder downscales generator image $\hat{x}$ to find the feature representation $\hat{z} = E(\hat{x})$. The encoder network explicitly learns to minimize the distance by its parametrization and this minimization is used for anomaly detection at the test time.

3. Discriminator network D to identify the input x and the output $\hat{x}$ as real or fake images. This sub-network adopts the standard discriminator network from **DCGAN**[3].

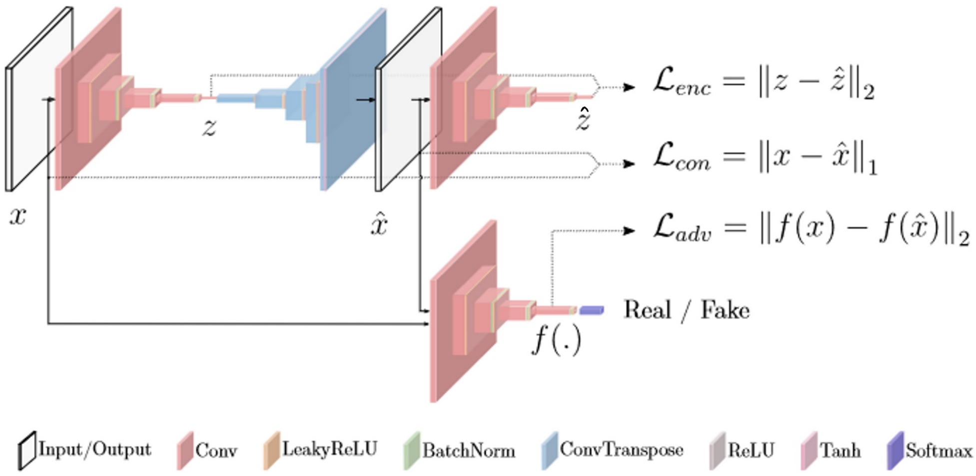

Figure 3.1: The architecture of GANomaly[18]

3.1.2 GANomaly training

The network is trained only on normal samples thus it is not able to reconstruct anomaly image by obtained parametrization. Therefore, the vectors z and z will have significant distinguish when the network is fed anomaly image since generator and encoders networks has missed abnormal feature representation.

In order to track dissimilarity within latent vector space the objective function is formed by combining the loss functions using individual subnetworks optimization.

- Adversarial Loss

The network updates generator G using the internal representation of D. Let f be a function that outputs an intermediate layer of the discriminator D for a given input x drawn from the input data distribution p_X. The feature matching computes the $L2$ distance between the feature representation of the real and the generated images.

$$\mathcal{L}_{adv} = E_{x \sim p_{\mathbf{X}}} ||f(x) - E_{x \sim p_{\mathbf{X}}} f(G(x)||_2$$

- Contextual Loss

  The generator should be optimized to process contextual information about the input images. It can be realize by contextual loss. The network employs measuring the $L1$ distance between the input image x and the generated image $G(x)$ to penalize generator G.

  $$\mathcal{L}_{con} = E_{x \sim p_{\mathbf{X}}} ||x - G(x)||_1$$

- Encoder Loss

  An additional encoder loss minimize the distance between the bottleneck features of the input $z = G_E(x)$ and the encoded features of the generated image $\hat{z} = E(G(x))$.

  $$\mathcal{L}_{enc} = E_{x \sim p_{\mathbf{X}}} ||G_E(x) - E(G(x))||_2$$

Thus, such a loss function system allows the network to produce images that are realistic and contextually justified. Otherwise minimizing the distance between the original and the generated images is unfeasible for anomaly samples due to optimization only to normal samples.

Objective function for the generator the following:

$$\mathcal{L} = w_{adv}\mathcal{L}_{adv} + w_{con}\mathcal{L}_{con} + w_{enc}\mathcal{L}_{enc}$$

where $w_{adv}, w_{con}, w_{enc}$ are the weighting parameters regulating the influence of each losses to the total objective function.

3.1.3 GANomaly testing

The network tests given sample with encoder loss equation to compute anomaly score $\mathcal{A}(\hat{x})$ for a test sample $\hat{x}$ using reconstruction error.

$$\mathcal{A}(\hat{x}) = ||G_E(\hat{x}) - E(G(\hat{x}))||_1.$$

Total evaluation of anomaly performance is define as set of anomaly scores for each test sample $\hat{x}$ in the test set $\hat{\mathcal{D}}$

$$\mathcal{S} = \{s_i : \mathcal{A}(\hat{x}_i), \hat{x}_i \in \hat{\mathcal{D}}\}$$

scaling within the probabilistic range of $[0, 1]$.

$$s_i' = \frac{s_i - in(\mathcal{S})}{max(\mathcal{S}) - min(\mathcal{S})}$$

Eventually the total evaluation of the test set $\hat{\mathcal{D}}$ is an anomaly score vector $\mathcal{S}'$.

3.2 f-AnoGAN

f-AnoGAN[20] is a fast unsupervised GAN based anomaly detection architecture based on a Wasserstein GAN network. This approach uses a generative model of normal training data and a fast mapping technique of new data to the latent space of the GAN network using trained encoder. For anomaly detection the network employs a combined anomaly score based on a discriminator feature residual error and an image reconstruction error. Discriminator does not have an output sigmoid function an outputs a scalar score not a probability.

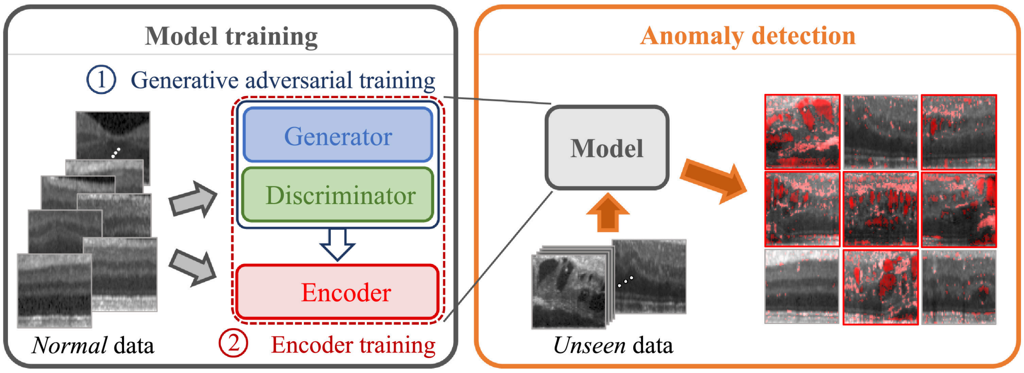

Figure 3.2: Anomaly detection by f-AnoGAN

3.2.1 f-AnoGAN architecture

f-AnoGAN contains two parts: Wasserstein GAN and Encoder.

1. The GAN model contains two networks that are generator and discriminator. This part uses the Wasserstein GAN model which uses the Wasserstein loss to keep stabilized training. The interaction to limit the weight and the cost function can lead to the vanishing or explosion of gradients. In order to this problem introduces a gradient penalty for the generator part to improve stability. The type of such model is known as a WGAN-GP.

2. Encoder is used for a fast mapping approach. That is meant fast mapping of given data to the latent space of the WGAN model. It maps images to locations in the latent space that map to the normal input image, when forward-pass as input to the trained generator. For normal image, mapping from image space to the latent space by the encoder and back to image space by the generator gives low deviation. Otherwise, for anomaly image the degree of deviation will indicate an anomaly. This degree defines anomaly score.

3.2.2 f-AnoGAN training

f-AnoGAN anomaly detection framework has two serial training stage using only normal images: WGAN and Encoder training.

1. WGAN training. At this stage only unlabeled normal data are used. WGAN is being trained for learning a nonlinear mapping function from latent space Z to the manifold X in the image space that represents the variability of normal training images. The generator G and the discriminator D are simultaneously optimized. The generator learns to reconstruct images of the training distribution capturing normal variability. The discriminator gives an evaluation of the fit of reconstructed images to the distribution of normal images. The generator and discriminator with fixed weight are used for subsequent encoder training.

   The training process of WGAN is illustrated Fig.3.3

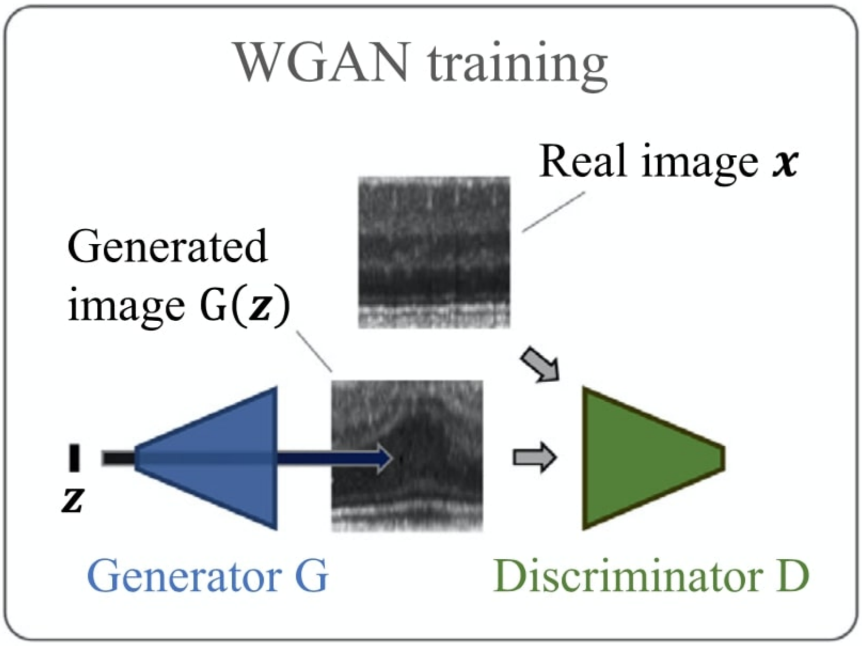

Figure 3.3: WGAN training

2. Encoder training using the trained WGAN. Inverse mapping is needed for anomaly detection. This is achieved by using training a deep encoder network E. The encoder is trained with architectures image-z-image izi_f by a discriminator. The izi_f approach simultaneously guides encoder training in the image space and in the latent space. The residual in the feature space obtained with the discriminator is a reliable basis to detect anomalous images, which is an essential term in the encoder training objective. This approach calculates image statistics of the real image and

the reconstructed image. The loss function for discriminator guided izi_f encoder training is:

$$L_{izi_f}(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G(E(\mathbf{x}))\|^2 + \frac{\kappa}{n_d} \cdot \|f(\mathbf{x}) - f(G(E(\mathbf{x})))\|^2$$

where discriminator features $f(\,)$ of an intermediate layer are used as statistics of a given input, n_d is the dimensionality of the intermediate feature representation, and is a weighting factor.

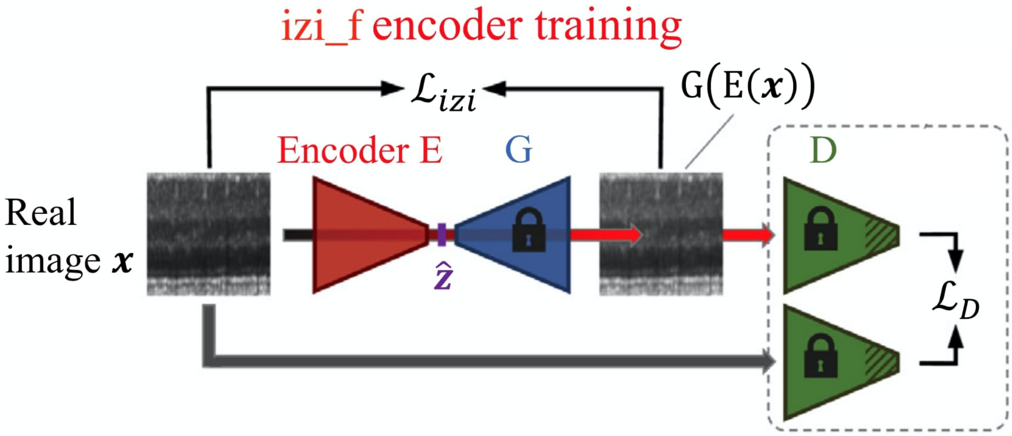

Figure 3.4: Encoder training

3.2.3 f-AnoGAN testing

The anomaly detection stage involves the evaluation of the degree of the deviation between given images and corresponding images produced by the generator. For this stage, the results from the previous two stages are used. Anomaly detection uses training WGAN and encoder to compute anomaly score.

The anomaly score explicitly depends on the loss function used to train the encoder part. f-employs the discriminator guided izi_f encoder architecture. The total anomaly score $\mathcal{A}(x)$ for input image x is defined as Reconstruction error + Discriminator error:

$$\mathcal{A}(x) = \mathcal{A}_{\mathcal{R}}(x) + \kappa \cdot \mathcal{A}_{\mathcal{D}}(x),$$

where is a weighting factor and

$$A_R(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G(E(\mathbf{x}))\|^2,$$

$$A_D(\mathbf{x}) = \frac{1}{n_d} \cdot \|f(\mathbf{x}) - f(G(E(\mathbf{x})))\|^2$$

This definition of the anomaly score include a discriminator term since the specific architecture for encoding training (Fig.3.4).

This approach provides high anomaly scores for images with anomalies and low anomaly scores for normal images. That is related to the model being only trained on normal data, thus training model is capable of only generating an image similar to the input image, that is, images without anomalies, incoming on the manifold of normal data. The ability to reconstruct an image indicates the type of image and the degree of deviation classifies normal and anomaly data.

For pixel-level anomaly localization, the absolute value of pixel-wise residuals is defined as:

$$A_R(\mathbf{x}) = |\mathbf{x} - G(E(\mathbf{x}))|$$

4 Data

For this research the Galaxy10 DECals Dataset [9] which creating due the Galaxy Zoo DECaLS [15] is used. Galaxy Zoo data is a really popular database with various versions. A lot of high-quality data allows to solve different tasks and perform experiments.

Galaxy Zoo is a scientific project on the morphological classification of large numbers of various types of galaxies. This project was created to separate galaxy images by classification tree and citizens to help in scientific research. In order to classify galaxies was used a series of questions and at least 40 volunteers for each image to obtain a probability for each parameter.

Galaxy Zoo DECals employs data from Dark Energy Camera Legacy Survey **DESI**[7].DECaLS employs the Dark Energy Camera (DECam) at the 4m Blanco telescope at Cerro Tololo Inter-American Observatory, close to La Serena, Chile.The DECaLS survey contributes to target images for the Dark Energy Spectroscopic Instrument (DESI). DECaLS observes the DESI footprint in the Southern Galactic Cap and the $\delta \leq 34$ region of the Northern Galactic Cap.

DECaLS provides galaxy data with increased resolution, image quality, and visibility of detailed morphology regarding the previous Galaxy Zoo 2 project (Fig. 4.1). The DECaLS images with r = 23.6 reveal features not previously visible. This dataset also has a high signal-to-noise ratio.

Galaxy Zoo DECals contains about 17000 galaxy images with 256x256 pixels size which was classified to 10 classes by volunteer votes and labels from Galaxy Zoo (Fig. 4.2).

Galaxy10 dataset (17736 images):

- Class 0 - Disturbed Galaxies (1081 images)
- Class 1 - Merging Galaxies (1853 images)
- Class 2 - Round Smooth Galaxies (2645 images)
- Class 3 - In-between Round Smooth (2027 images) Galaxies
- Class 4 - Cigar Shaped Smooth (334 images) Galaxies
- Class 5 - Barred Spiral Galaxies (2043 images)
- Class 6 - Unbarred Tight Spiral Galaxies (2043 images)

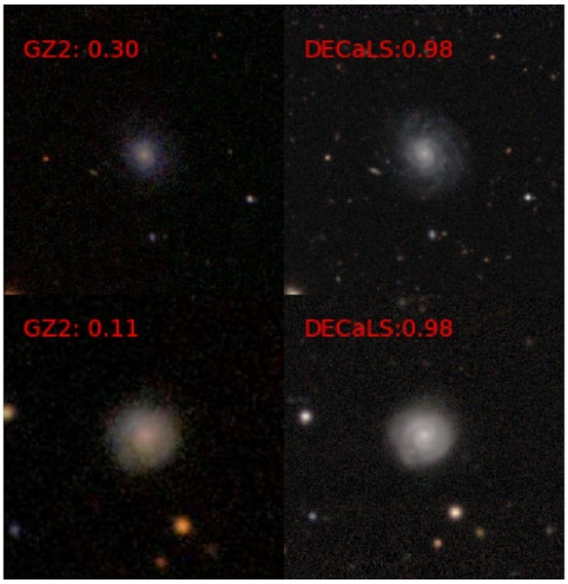

Figure 4.1: Comparison of GZ2 and GZ DECaLS images[15].

- Class 7 Unbarred Loose Spiral Galaxies (2628 images)
- Class 8 - Edge-on Galaxies without Bulge (1423 images)
- Class 9 - Edge-on Galaxies with Bulge (1873 images)

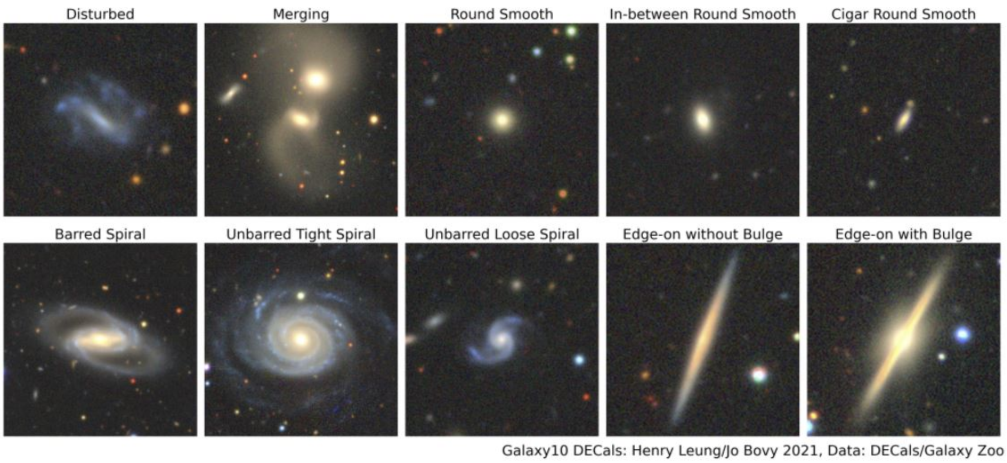

Figure 4.2: Example of each class from Galaxy10 DECals.

For this work the Hubble sequence to separate galaxies is used. It divides regular galaxies into three broad classes: elliptical, spiral and lenticular. The number of images in each class of Galaxy10 DECals Dataset is not evenly distributed. In order to balance the classes for detecting anomalies, the following separation was made:

- Elliptical (E) - Class 2 and Class 3 (Fig. 4.3.a)
- Spiral (S) - Class 5 and Class 6 (Fig. 4.3.b)
- Lenticular (S0) Class 8 and Class 9 (Fig. 4.3.c)

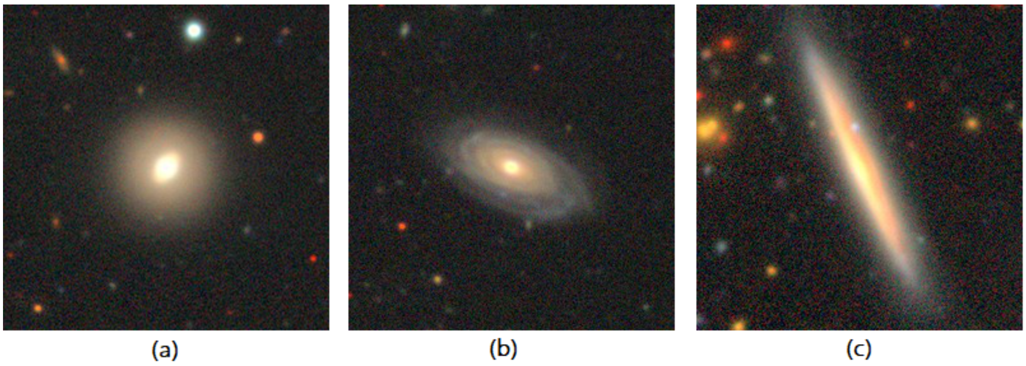

Figure 4.3: Example of each galaxy type for final dataset.

5 Experiments

This chapter describes the training of selected architectures for evaluating the performance of the anomaly detection problem for galaxy images.

Two architectures for anomaly detection were chosen for the experiment: GANomaly and f-AnoGAN. These models were trained using a dataset containing images from Galaxy 10 DECals. The dataset is divided into three classes containing images of regular galaxies: elliptical (E), spiral (S), lenticular (S0). Only normal data was used to train both models. Abnormal and normal data in a balanced ratio were used for testing. This experiment allows to evaluate the selected models for their application as detection of anomalous images in large astronomical data.

To evaluate the performance of the models and compare the work of algorithms for given task, two metrics that are common for the anomaly detection task are used. These are AUC-ROC and F1-score metrics which is decribed in the Section 5.1.

5.1 Evaluation metrics

In machine learning tasks, metrics are used to evaluate the quality of models and compare different algorithms. Before calculating performance metrics, it is necessary to obtain a Confusion matrix that contains predicted and actual values (Fig. 5.1). It is simple table to obtain the performance of algorithms for classification problem.

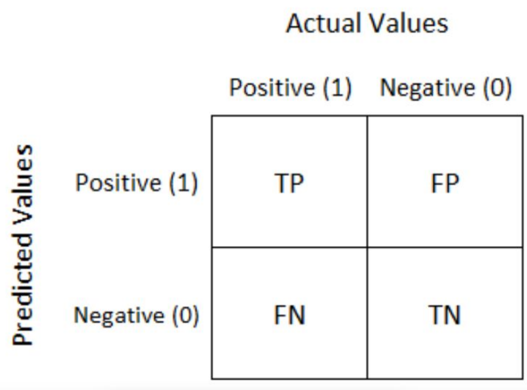

Figure 5.1: Confusion Matrix.

Terminology for confusion matrix:

- P (condition positive) is the number of positive cases;
- N (condition negative) is the number of negative cases;
- TP (true positive) is the number of images with an anomalies for which the model gives an estimate that there is an anomaly;
- FP (false positive) is the number of images without anomalies for which the model gives an estimate that there is an anomaly;
- TN (true negative) is the number of images without anomalies for which the model gives an estimate that there is no anomaly;
- FN (false negative) is the number of images with anomalies for which the model gives an estimate that there is no anomaly.

Performance metrics are accuracy, recall, precision, F1-score, AUC-ROC which are computed on the basis of the based on the above TP, TN, FP and FN values. The most commonly used performance evaluation metrics for anomaly detection problem according to these values is the area under the curve (AUC) of the receiver operating characteristics (ROC)[6] and F1-score metric.

To calculate the AUC-ROC metric, it is required to calculate true positive rates (TPR), which corresponds to the ability of the model to recognize images with an anomaly, and false positive rates (FPS), which corresponds to the ability of the model not to take a normal image for an abnormal one.

The **ROC**[17] curve is a graph that allows you to evaluate the quality of classification, displays the ratio between the proportion of objects from the total number of containing of feature correctly classified as containing of feature (true positive rate, TPR is the sensitivity of the classification algorithm), and the proportion of objects from the total number objects that do not contain a feature, mistakenly classified as containing a feature (false positive rate, FPR, the value of $1 - FPR$ is called the specificity of the classification algorithm) when the threshold of the decisive rule is varied.

- True Positive Rate is also known as recall.

  TPR $= \frac{TP}{TP+FN}$
- False Positive Rate

  FPR $= \frac{FP}{TN+FP}$

Quantitative interpretation of ROC gives the AUC metric. AUC is the area bounded by the ROC curve and the axis of the proportion of the false-positive axis. The higher the AUC is the better the model. This shows how much the model is able to recognize anomalies and at the same time not define normal data as abnormal. A value of 0.5 demonstrates the unsuitability of

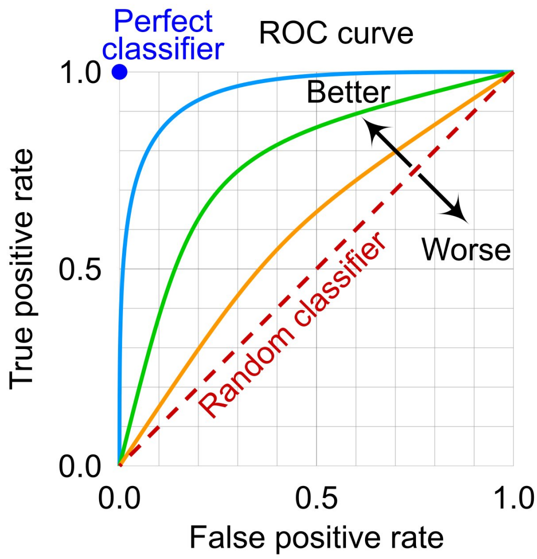

Figure 5.2: ROC curve with FPR and TPR [5].

the chosen anomaly detection method (corresponds to random prediction). A good model has an AUC value close to 1, which means a high degree of detection.

The AUC value for a anomaly detection problem with two cases (normal and abnormal) can be calculated as:

$$AUC = \frac{\sum_{i=1}^{P} r_i - P(P+1)/2}{PN}$$

where r_i is the rank of ith positive sample in the ranking table according to the probability of the normal class [17].

AUC has two main advantages :

- Scale-invariance. AUC measures not the absolute values of prediction, but rather how correct these are ranked.
- Classification-threshold-invariance. AUC measures the model's prediction quality regardless of the classification threshold choice.

F1-score is harmonic mean of the recall and precision. To calculate the F1-score it is necessary to compute two important value recall and precision which is also related with predicted and actual values.

Precision is the proportion of objects which the classifier called as positive and concurrently which is really being positive. Recall demonstrates the proportion of objects of a positive case of all objects of a positive class that was found. Recall shows the skill of the algorithm to identify the class overall. The precision presents the skill to separate one class from other classes.

$$precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

$$F1 - score = 2 * \left(\frac{\text{precision * recall}}{\text{precision + recall}}\right)$$

The F1-score reaches a maximum with precision and recall equal to one, and is close to zero if one of the arguments is close to zero. F1-score takes values in the range from 0 to 1. The precision and recall values are perfect if F1-score is equal to 1.

F1-score has problems when evaluating unbalanced classes, for this reason sometimes modifications of the dataset are necessary.

5.2 Training and Anomaly detection

The training of the models are performed using cloud service. The Google Colaboratory Pro was used to train both selected models. The Google Colaboratory allows to write and execute python code by web browser. This service is perfect for deep learning tasks it is one of the best ways to perform high performance computing since Colab Pro gives access to a graphics processing unit (GPU).

Colab Pro offers two GPUs Tesla P-100, Tesla T4 with 16 gigabytes of memory, 4 cores of CPU, huge RAM memory and CUDA version 11.2. Colab Pro limits the session time to 24 hours. This solution is suitable for a large number of tasks. Colab Pro interacts well with Google drive, which simplifies the work and makes it possible to automate the learning process.

Before the anomaly detection performing, it is necessary to determine the anomaly detection assumption. The training and test data are divided according to the normal and abnormal data. That is it is necessary to set how to divide the dataset into normal and abnormal data, which classes belong to normal data, and which to abnormal. There are several approaches to solve this problem. The different datasets to separate into normal and abnormal data can be used, it is also known as multiple sata sets. However, this approach is more often used for the out-of-distribution detection task. Another assumption is to use multi-class that means the one class of the dataset is considered as abnormal data, and others are considered as normal data. This approach is used for anomaly detection and novelty detection prblems. There is also a so called one-class assumption. This approach consider one class as normal and the other classes are anomalous. This research employs the one-class single dataset assumption.

To evaluate GANomaly and f-AnoGAN models the Galaxy Galaxy Zoo DECaLS data and Galaxy10 DECaLS dataset was used. The collected dataset consists of three classes according to the classification of regular galaxies: Elliptical (E), Spiral (S), Lenticular (S0).

A total of six experiments were performed for this dataset. Three training sessions were done for each model. Each class from the dataset was alternately selected as normal, and the rest as abnormal. Therefore, three combinations were used:

- E as normal + S, So as abnormal;
- S as normal + E, So as abnormal;
- So as normal + E, S as abnormal.

GANomaly and f-AnoGAN have been trained for every case. For training models only normal samples is used. For testing the normal and abnormal samples is used.

5.2.1 GANomaly

GANomaly is GAN-based Semi-Supervised Anomaly Detection model which using adversarial training. For evaluation GANomaly model an implementation based on the PyTorch framework was chosen [2].

The model allow set three losses for anomaly detection using weighting parameters:

- Adversarial loss (w_{adv})
- Contextual loss (w_{con})
- Encoder loss (w_{enc})

Training configuration:

- Image size: 256
- Batch size: 4
- Number of epochs: 40

The remaining settings are defined in accordance with the standard implementation. The model uses Adam optimization algorithm with initial learning rate $l_r = 0.0002$ and momentums $\beta_1 = 0.5$, $\beta_2 = 0.999$. Loss function is realized with weight values $w_{adv} = 1$, $w_{rec} = 50$ and $w_{wenc} = 1$. The dimension of the latent space is equal to 100.

5.2.2 f-AnoGAN

f-AnoGAN is GAN-based Unsupervised Anomaly Detection model which has three-component architecture that include an encoder, generator and discriminator. One of the significant traits of the model is a serial training process. The two adversarial networks (GAN) and Encoder are trained independently. The structure allows to compute of an anomaly score using discriminator residual error and an image reconstruction error. For evaluation f-AnoGAN architecture an implementation based on the PyTorch framework was selected [1].

Training configuration for WGAN:

- Image size: 256
- Batch size: 4
- Number of epochs: 150

Training configuration for encoder:

- Image size: 256
- Batch size: 4
- Number of epochs: 100

The others settings for WGAN and encoder training are defined in the same way as for GANomaly: Adam - $l_r = 0.0002$, $\beta_1 = 0.5$, $\beta_2 = 0.999$. The dimension of the latent space is 100.

5.2.3 Results

Two different methods mentioned in the previous section were tested. Each method was tested on three cases for one dataset. Alternately, one of the classes was called normal, and the rest abnormal.

In order to evaluate and compare the models the Area-Under-the-Curve (AUC) of the Receiver-Operating-Characteristic (ROC) curve was employed. The mean performance across all classes was calculated for each of the models. The results of the evaluation can be seen in Table. 5.1 and Table. tab:res1.

Table 5.1: AUC-ROC values for GANomaly and f-AnoGAN

	E	S	So	mean
GANomaly	0.85	0.64	0.61	**0.7**
f-AnoGAN	0.89	0.82	0.80	**0.84**

Table 5.2: F1-score values for GANomaly and f-AnoGAN

	E	S	So	mean
GANomaly	0.87	0.79	0.80	**0.82**
f-AnoGAN	0.74	0.71	0.71	**0.72**

The tables show the results for 6 cases for 2 models and 3 classes.

The GANomaly model has well reconstructed a class with elliptical galaxies. But for the other two classes, the model could not correctly determine the distribution of features. The distribution of abnormal and normal images for the test data strongly overlap.

The f-AnoGAN turned out to be more responsive to anomalies. It showed good results in detecting abnormal images for all classes. This model reconstructed all three classes well and showed the correct distribution of data between them for all classes. Some graphs are presented in appendix A.

The f-AnoGAN architecture has provided better results for detecting anomalies than the GANomaly architecture. Presumably it related with structure of f-AnoGAN which allows to compute of an anomaly score using discriminator residual error and an image reconstruction error. This improves its performance for this task and for many others.

Conclusion

The main goal of the current research was to estimate specialized deep learning methods for astronomical images. In this work, generative methods were investigated as a way to detect anomalous behavior in images of galaxies without labels.

In the course of the work, an overview of existing solutions was performed. The overview showed that the existing solutions for anomaly detection in astronomical images do not apply adapted deep anomaly detection architectures. In this regard, a theoretical comparative analysis of two generative models was conducted , which demonstrated that Generative Advesireal Networks are more appropriate for anomaly detection in astronomical images than Convolutional Autoencoders.

The study showed that currently there are a large number of GAN-based architectures to solve given problem. In this work, two popular models for detecting anomalies were studied and trained: GANomaly and f-AnoGAN. GANomaly is developed on the basis of DCGAN, f-AnoGAn is developed on the basis of WGAN-GP. These models have been trained using Galaxy Zoo DECals Dataset.

Testing of models has shown that f-AnoGAN gives performance results better than GANomaly. Perhaps this is due to the GANomaly architecture based on DCGAN, which is prone to colaps mode and also because of the architectural features of f-AnoGAN that allow anomalies to be analyzed using reconstruction error and discretionary error.

This study has found that f-AnoGAN is able to perfectly reconstruct astronomical images and learn the true distribution of data in given samples. This model is appropriate for anomaly detection in galaxy images.

Bibliography

[1] Implementation of f-anogan with pytorch. https://github.com/A03ki/f-AnoGAN, 2020.

[2] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, and Utku Genc. Anomalib: A deep learning library for anomaly detection, 2022.

[3] Soumith Chintala Alec Radford, Luke Metz. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434, November 2015.

[4] Tom Charnock Carla Margalef-Bentabol Mariangela Bernardi Yohan Dubois Kate Storey-Fisher Lorenzo Zanisi Berta Margalef-Bentabol, Marc Huertas-Company. Detecting outliers in astronomical images with deep generative networks. Monthly Notices of the Royal Astronomical Society, Volume 496, Issue 2, page 2346–2361, June 2020.

[5] MartinThoma Cmglee. Receiver operating characteristic (roc) curve with false positive rate and true positive rate. Roc-draft-xkcd-style.svg, CC BY-SA 4.0,, 2018.

[6] H. Zhang C.X. Ling, J. Huang. Auc: a statistically consistent and more discriminating measure than accuracy. Ijcai, 3, pages 519–524, 2003.

[7] Dey A. et al. Overview of the desi legacy imaging surveys. AJ, 157, 168, 2019.

[8] L. Cao A.V.D. Hengel G. Pang, C. Shen. Deep learning for anomaly detection: A review. ACM Computing Surveys (CSUR), 54, pages 1–38, 2021.

[9] J. Bovy H. Leung. Galaxy10 decals dataset. https://github.com/henrysky/Galaxy10, 2021.

[10] Bruno Lecouat Gaurav Manek Vijay Ramaseshan Chandrasekhar Houssam Zenati, Chuan Sheng Foo. Efficient gan-based anomaly detection. arXiv:1802.06222, February 2018.

[11] M. Mirza B. Xu D. Warde-Farley S. Ozair A. Courville Y. Bengio I.J. Goodfellow, J. Pouget-Abadie. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, MIT Press, pages 2672–2680, 2014.

[12] Ye-In Park Seong-Gyun Kim Suk-Ju Kang Jouwon Song, Kyeongbo Kong. Anoseg: Anomaly segmentation network using self-supervised learning. arXiv:2110.03396, October 2021.

[13] Nesar Ramachandra-Francois Lanusse Alexie Leauthaud Yifei Luo Song Huang J. Xavier Prochaska Kate Storey-Fisher, Marc Huertas-Company. Anomaly detection in hyper suprime-cam galaxy images with generative adversarial networks. Monthly Notices of the Royal Astronomical Society, Volume 508, Issue 2, pages 2946–2963, December 2021.

[14] S. Osindero M. Mirza. Conditional generative adversarial nets. Computer Science, pages 2672–2680, 2014.

[15] Tobias Geron Sandor Kruk-Coleman Krawczyk Kyle W. Willett Steven Bamford Lee S. Kelvin Lucy Fortson Yarin Gal William Keel Karen L. Masters Vihang Mehta Brooke D. Simmons Rebecca Smethurst Lewis Smith Elisabeth M. Baeten Christine Macmillan Mike Walmsley, Chris Lintott. Galaxy zoo decals: Detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies. Monthly Notices of the Royal Astronomical Society, Volume 509, Issue 3, pages 3966–3988, January 2022.

[16] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901.

[17] Z.-C. Qin. Roc analysis for predictions made by probabilistic classifiers. International Conference on Machine Learning and Cybernetics, volume 5, pages 3119–3124, 2005.

[18] Toby P. Breckon Samet Akcay, Amir Atapour-Abarghouei. Ganomaly: Semi-supervised anomaly detection via adversarial training. arXiv:1805.06725, May 2018.

[19] Sebastian M. Waldstein Ursula Schmidt-Erfurth Georg Langs Thomas Schlegl, Philipp Seeböck. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. The international conference on Information Processing in Medical Imaging (IPMI), March 2017.

[20] Sebastian M.Waldstein Georg Langs Ursula Schmidt-Erfurth Thomas Schlegl, Philipp Seeböck. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis, Volume 54, pages 30–44, January 2019.

[21] R. Houthooft J. Schulman I. Sutskever-P. Abbeel X. Chen, Y. Duan. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. Proceedings of the 30th International Conference on Neural Information Processing Systems, pages 2180–2188, 2016.

Appendix

A. Results for each class of galaxy

In the images Fig.1-6 is illustrated results for the f-AnoGan model.

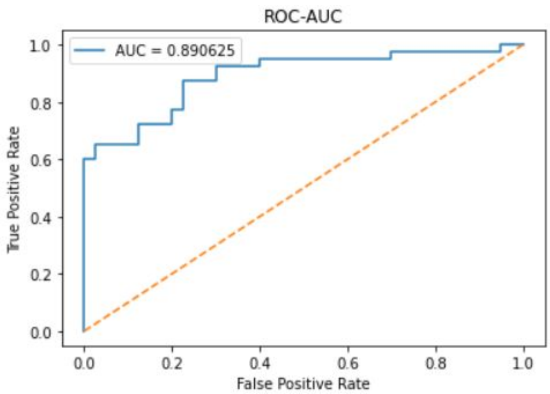

Figure 1: ROC curve for Elliptical class

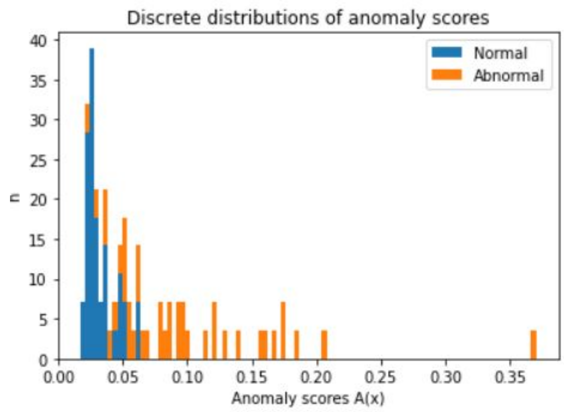

Figure 2: The distributions of normal and abnormal data for Elliptical class

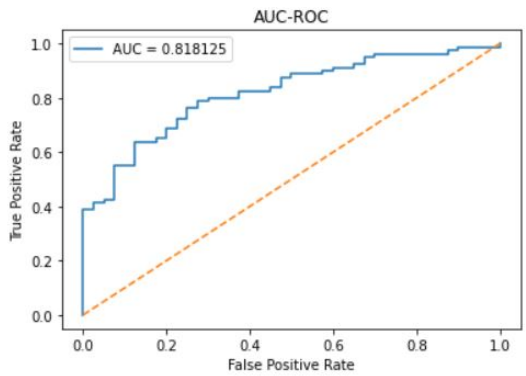

Figure 3: ROC curve for Spiral class

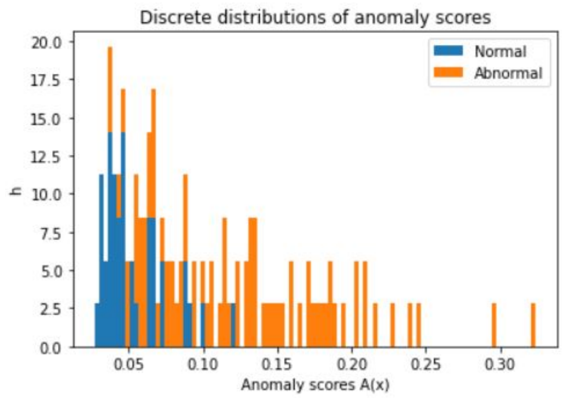

Figure 4: The distributions of normal and abnormal data for Spiral class

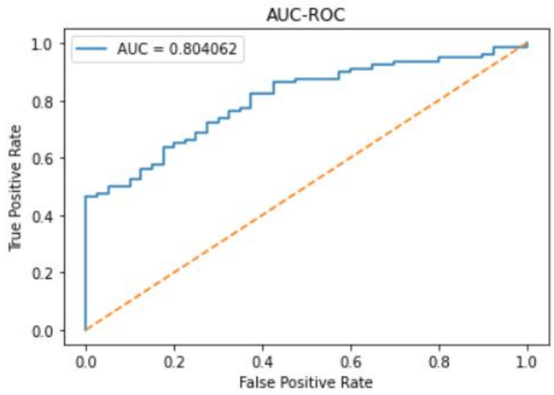

Figure 5: ROC curve for Lenticular class

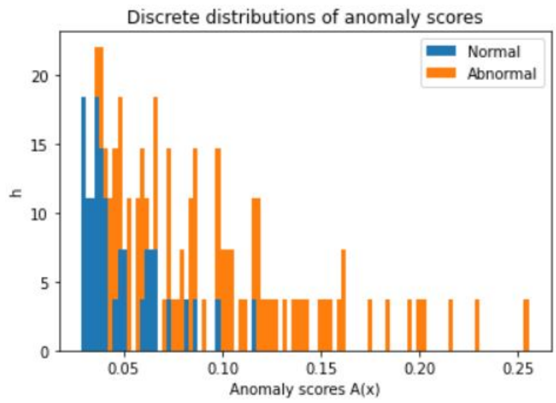

Figure 6: The distributions of normal and abnormal data for Lenticular class