

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE



# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

### ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

## ZVYŠOVÁNÍ KVALITY OBRAZOVÝCH SNÍMKŮ PRO BEZPEČNOSTNÍ SLOŽKY STÁTU

ENHANCEMENT OF IMAGE QUALITY FOR SECURITY FORCES

### BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

### AUTOR PRÁCE

AUTHOR

Adam Varga

### VEDOUCÍ PRÁCE

SUPERVISOR

doc. Ing. Radim Burget, Ph.D.

BRNO 2019

# Bakalářská práce

bakalářský studijní obor **Informační bezpečnost**  
Ústav telekomunikací

**Student:** Adam Varga

**ID:** 156107

**Ročník:** 3

**Akademický rok:** 2018/19

**NÁZEV TÉMATU:**

## Zvyšování kvality obrazových snímků pro bezpečnostní složky státu

**POKYNY PRO VYPRACOVÁNÍ:**

Seznamte se s problematikou super-rozlišení a odstraňování šumu z obrazu s využitím hlubokých konvolučních neuronových sítí za účelem dosažení vyšší kvality obrazových záznamů z kamerových zařízení. S pomocí databáze NTIRE 2018 <https://goo.gl/qBCHHU> ověřte možnosti zvyšování kvality z obecných snímků obrazu. Vhodným způsobem srovnajte vybrané přístupy včetně ESRRGAN (<https://github.com/xinntao/ESRRGAN>) a dosažené výsledky vhodně prezentujte.

**DOPORUČENÁ LITERATURA:**

- [1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
- [2] Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2014.

**Termín zadání:** 1.2.2019

**Termín odevzdání:** 27.5.2019

**Vedoucí práce:** doc. Ing. Radim Burget, Ph.D.

**Konzultant:**

**prof. Ing. Jiří Mišurec, CSc.**  
*předseda oborové rady*

**UPOZORNĚNÍ:**

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## **ABSTRAKT**

Táto bakalárska práca sa zaoberá zvyšovaním kvality obrazových snímok pre bezpečnostné zložky štátu. Zvyšovanie kvality v tomto prípade znamená zväčšenie rozlíšenia obrazových dát využitím techniky super-rozlíšenia za pomoci modelov hlbokých konvolučných neurónových sietí. Práca vo svojej teoretickej časti približuje princípy fungovania tejto techniky a v jej praktickej časti je priblížená práca s vybranými súčasnými modelmi v oblasti super-rozlíšenia.

## **KĽÚČOVÉ SLOVÁ**

Neurónové siete, Hlboké učenie, Zvyšovanie kvality obrazu, Techniky rekonštrukcie obrazu, Super-rozlíšenie, Odstraňovanie šumu z obrazu, Metódy v oblasti super-rozlíšenia, Konvolučné neurónové siete, Reziduálne neurónové siete, TensorFlow

## **ABSTRACT**

This bachelor thesis deals with image quality enhancement for security forces. Image quality enhancement in this case means increasing the resolution of image data by using super-resolution techniques using models of deep convolutional neural networks. The thesis in its theoretical part describes the principles of the operation of this technique and in its practical part is presented the work with selected state-of-the-art models in the area of super-resolution.

## **KEYWORDS**

Neural networks, Deep learning, Image quality enhancement, Image reconstruction techniques, Super-resolution, Image noise reduction, Super-resolution methods, Convolutional neural networks, Residual neural networks, TensorFlow

VARGA, Adam. *Zvyšovanie kvality obrazových snímok pre bezpečnostné zložky štátu*. Brno, Rok, 51 s. Bakalárska práca. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačných technológií, Ústav telekomunikácií. Vedúci práce: doc. Ing. Radim Burget, Ph.D.



## VYHLÁSENIE

Vyhlasujem, že som svoju bakalársku prácu na tému „Zvyšovanie kvality obrazových snímok pre bezpečnostné zložky štátu“ vypracoval samostatne pod vedením vedúceho bakalárskej práce, využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej bakalárskej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto bakalárskej práce som neporušil autorské práva tretích osôb, najmä som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a/alebo majetkových a som si plne vedomý následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona Českej republiky č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon), v znení neskorších predpisov, vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákoníka Českej republiky č. 40/2009 Sb.

Brno .....

.....

podpis autora

## POĎAKOVANIE

Rád by som sa poďakoval vedúcemu bakalárskej práce pánovi doc. Ing. Radimovi Burgetovi, Ph.D. za odborné vedenie, konzultácie, trpezlivosť a podnetné návrhy k práci. Svoju vďačnosť vyjadrujem taktiež rodine, priateľom a všetkým tým, ktorí má akýmkoľvek spôsobom motivovali počas práce na tomto diele.

Brno .....

.....

podpis autora

Tato práce vznikla jako součást klíčové aktivity KA6 - Individuální výuka a zapojení studentů bakalářských a magisterských studijních programů do výzkumu v rámci projektu OP VVV Vytvoření double-degree doktorského studijního programu Elektronika a informační technologie a vytvoření doktorského studijního programu Informační bezpečnost, reg. č. CZ.02.2.69/0.0/0.0/16\_018/0002575.



EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

Projekt je spolufinancován Evropskou unií.

# Obsah

|  |           |
|--|-----------|
| Úvod   | 10        |
| <b>1 Neurónové siete</b>   | <b>11</b> |
| 1.1 Model neurónu  | 11        |
| 1.2 Vrstvy neurónovej siete  | 12        |
| 1.3 Učenie neurónovej siete  | 13        |
| 1.3.1 Dataset - dátová množina   | 13        |
| 1.3.2 Backpropagation - učenie s algoritmom spätného šírenia                             | 13        |
| 1.4 Konvolučné neurónové siete   | 14        |
| 1.4.1 Konvolučná vrstva  | 14        |
| 1.4.2 Združovacia vrstva   | 15        |
| 1.4.3 Plne prepojená vrstva  | 16        |
| 1.5 Reziduálne neurónové siete   | 16        |
| 1.5.1 Reziduálny blok  | 17        |
| 1.6 Sieť husto prepojených blokov  | 17        |
| 1.6.1 Husto prepojený blok   | 18        |
| <b>2 Techniky rekonštrukcie obrazu</b>   | <b>19</b> |
| 2.1 Super-rozlíšenie   | 19        |
| 2.1.1 Hlavné prínosy v oblasti super-rozlíšenia  | 20        |
| 2.1.2 NTIRE - výzva nových techník v oblasti super-rozlíšenia                            | 21        |
| 2.2 Odstraňovanie šumu z obrazu  | 22        |
| 2.2.1 Digitálny šum v obraze   | 22        |
| 2.2.2 Odstraňovanie šumu pomocou hlbokého učenia   | 23        |
| <b>3 Modely hlbokého učenia pre zvyšovanie kvality obrazu technikou super-rozlíšenia</b> | <b>25</b> |
| 3.1 Použité nástroje   | 25        |
| 3.2 Vybrané datasety   | 25        |
| 3.3 Super-Resolution using Generative Adversial Network (SRGAN)                          | 27        |
| 3.3.1 Chybové funkcie modelu SRGAN   | 27        |
| 3.3.2 Architektúra siete   | 28        |
| 3.4 Enhanced Super-Resolution using Generative Adversial Network (ESR-GAN)               | 30        |
| 3.4.1 Architektúra siete   | 30        |
| 3.4.2 Popis implementácie  | 33        |
| 3.5 Enhanced Deep Super-Resolution (EDSR)  | 34        |

|          |   |           |
|----------|---|-----------|
| 3.5.1    | Architektúra siete . . . . .                        | 34        |
| 3.5.2    | Popis implementácie . . . . .                       | 35        |
| 3.6      | Residual Dense Network for Image SR (RDN) . . . . . | 36        |
| 3.6.1    | Architektúra siete . . . . .                        | 36        |
| 3.6.2    | Popis implementácie . . . . .                       | 38        |
| <b>4</b> | <b>Výsledky práce</b>                               | <b>40</b> |
| 4.1      | Metriky hodnotenia . . . . .                        | 40        |
| 4.1.1    | Stredná kvadratická chyba (MSE) . . . . .           | 40        |
| 4.1.2    | Špičkový odstup signálu od šumu (PSNR) . . . . .    | 40        |
| 4.1.3    | Structural similarity index (SSIM) . . . . .        | 41        |
| 4.1.4    | Subjektívne metriky . . . . .                       | 42        |
| 4.2      | Hodnotenie vybraných modelov . . . . .              | 42        |
| <b>5</b> | <b>Záver</b>  | <b>44</b> |
|          | <b>Literatúra</b>                                   | <b>45</b> |
|          | <b>Zoznam symbolov, veličín a skratiek</b>          | <b>50</b> |

# Zoznam obrázkov

|     |   |    |
|-----|---|----|
| 1.1 | Grafická reprezentácia vrstiev perceptrónu . . . . .  | 11 |
| 1.2 | Zobrazenie konvolučnej neurónovej siete LeNet5, zdroj [10] . . . . .  | 14 |
| 1.3 | Násobenie matíc ako príklad 2D diskkrétnej konvolúcie . . . . .   | 15 |
| 1.4 | Ukážka algoritmu Max pooling zobrazuje dvojnásobné zmenšenie vstupného obrazu reprezentovaného maticou o veľkosti 4x4 . . . . . | 16 |
| 1.5 | Reziduálny blok uvedený v práci ResNet [8] . . . . .  | 17 |
| 1.6 | Husto prepojený (Dense) blok uvedený v práci DenseNet [14] . . . . .  | 18 |
| 2.1 | Zvyšovanie kvality obrazových záznamov, model <i>FSRGAN</i> [18] . . . . .  | 19 |
| 2.2 | Ukážka architektúry modelu <i>FSRNet</i> [18] . . . . .   | 21 |
| 2.3 | Architektúra modelu <i>FFDNet</i> uvedená v práci [33] . . . . .  | 23 |
| 3.1 | Štruktúra použitých datasetov . . . . .   | 26 |
| 3.2 | Generátor a diskriminátor modelu SRGAN . . . . .  | 29 |
| 3.3 | Reziduálny blok reziduálnych husto prepojených blokov RRDB použitých v modeli ESRGAN, publikovaný v práci [4] . . . . .         | 30 |
| 3.4 | Architektúra siete generátora modelu ESRGAN . . . . .   | 31 |
| 3.5 | Časť repozitára <a href="https://github.com/xinntao/BasicSR">https://github.com/xinntao/BasicSR</a> . . . . .                   | 33 |
| 3.6 | Základná štruktúra konfiguračného súboru <code>train_ESRGAN.json</code> . . . . .   | 34 |
| 3.7 | Blokový diagram modelu EDSR uvedený v práci [5] . . . . .   | 35 |
| 3.8 | Architektúra siete RDN (vľavo), štruktúra bloku RDB (vpravo) . . . . .  | 37 |
| 4.1 | Dvojnásobne zväčšené obrázky pomocou vybraných metód . . . . .  | 43 |

# Úvod

Pre technickú sekciu národných bezpečnostných zložiek štátu je proces identifikácie osôb každodennou rutinnou prácou. Dôležitá je v tomto prípade nepopierateľnosť rozhodnutia a tým pádom aj presnosť posudzovania faktov. Mimo iného, presnosť identifikovania osoby z obrazových dát je tým vyššia, čím vyššia je kvalita týchto dát (fotografií a videozáznamov). Zvyšovanie kvality je v tomto prípade významným parametrom ovplyvňujúcim presnosť každého konkrétneho rozhodnutia. Práca rozoberá super-rozlíšenie pomocou hlbokých konvolučných a reziduálnych neurónových sietí, čo sú v súčasnosti uznávané techniky pre dosiahnutie vyššej kvality obrazu. Super-rozlíšenie má v tejto súvislosti veľký potenciál s presnosťou identifikácie osôb pre bezpečnostné zložky.

Práca na začiatku približuje všeobecný teoretický základ neurónových sietí. Druhá kapitola bližšie definuje dve hlavné techniky rekonštrukcie obrazu - super-rozlíšenie a odstraňovanie šumu z obrazu. V ďalšej časti sú teoreticky a prakticky vysvetlené použité modely, ich architektúry a implementačné detaily. Táto kapitola rovnako obsahuje aj popis použitých nástrojov a dátovej sady.

Kapitola s výsledkami práce uzatvára toto dielo. Sú v nej vysvetlené metriky hodnotenia a prezentované výsledky testovania natrénovaných modelov super-rozlíšenia. Práca zhodnotila výsledky testovania štyroch modeloch: SRGAN, ESRGAN, ISR, EDSR na dvoch typoch dátových sád.

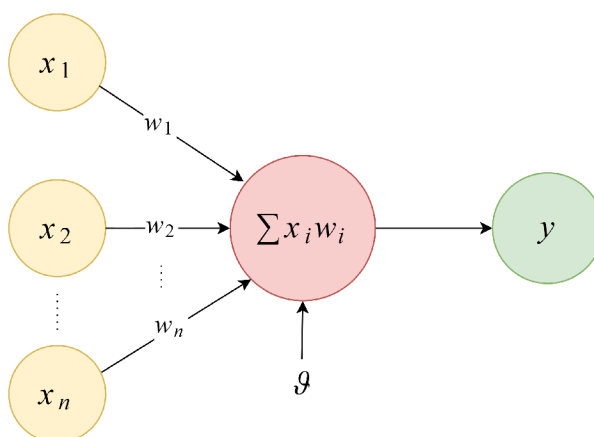
# 1 Neurónové siete

Princípom fungovania neurónových sietí je abstrahovanie základných rysov ľudského mozgu. Tieto základne črty mozgu vychádzajú z jeho neurofyziologických poznatkov. V mozgu je napríklad až  $10^{10}$  navzájom prepojených neurónov, pričom každý z nich môže mať až 5000 spojov s ostatnými neurónmi [1]. Vďaka veľkej zložitosti mozgovej štruktúry súčasná veda stále nedisponuje úplnými poznatkami mozgu a jeho fungovania. Neurónové siete pri výpočte využívajú paralelné spracovanie informácie. Tento fakt ako aj samotná prítomnosť učenia ich odlišuje od konvenčných počítačových algoritmov.

## 1.1 Model neurónu

Neurón je základná funkčná bunka mozgu. Je zistené, že neuróny majú dendrity a axóny. Cez dendrity prichádza do neurónu vzruch, ktorý je axónom cez synapsie posielaný ďalším dendritom vzájomne spojených neurónov.[2]

Najviac citovaný model neurónu, ktorý zjednodušuje biologický neurón, bol popísaný Frankom Rosenblattem už v roku 1958. Autor mu dal označenie „perceptrón“ [3]. Základná funkčnosť neurónu ostáva však nezmenená. V tomto pojatí neurón obsahuje  $n$  vstupov  $x_1, x_2, \dots, x_n$  (s výnimkou neurónov vstupnej vrstvy) a minimálne jeden výstup  $y_1, y_2, \dots, y_n$  (s výnimkou neurónov výstupnej vrstvy). Neurón vypočíta svoj *vnútorný potenciál*  $\xi$  ako vážený súčet  $n$  vstupných hodnôt predaných susednými neurónmi spolu s jeho nastaveným prahom  $\vartheta$ , vzorec 4.1. Vnútorný potenciál je parameter pre prenosovú, alebo tiež aj aktivačnú funkciu  $f(\xi)$ .



Obr. 1.1: Grafická reprezentácia vrstiev perceptrónu

Na obrázku 1.1 je zobrazená jednoduchá jednovrstvová dopredná neurónová sieť, tzv. *perceptrón*. Aktivačná funkcia neurónu sa vypočíta podľa:



$$f(\xi), \text{ kde } \xi = \vartheta + \sum_{i=1}^n x_i w_i. \quad (1.1)$$

Vstupy neurónu obsahujú synaptické *váhy*  $w_1, w_2, \dots, w_n$ . V rozhodnutí, či neurón výsledný signál prepošle na svoj výstup majú váhy kľúčovú úlohu. Dávajú svojim vstupom určitú dôležitosť. Ich zmenou je možné určovať, aké vstupy sú pre neurón dôležité a naopak, ktorý signál nie je vo finále účinný. Váhy jednotlivých vstupov majú interval hodnôt v rozmedzí  $[0; 1]$ . Čím vyššia váha, tým hodnota z tohto vstupu má väčší význam (väzba na daný neurón je väčšia).

Aktivačná funkcia je využívaná kvôli prínosu nelinearity do výpočtu. Druhov aktivačných funkcií existuje viacero a to, aká sa v neuróne nakoniec použije závisí na jeho využití. Perceptrón sa rozdeľuje na diskrétny a spojitý. Produkovaný výstup môže byť buď skokový (0, resp.  $-1$  a  $1$ ) alebo spojitý. Medzi najrozšírenejšie aktivačné funkcie patrí skoková funkcia, hyperbolický tangens, ReLU (Rectified Linear Unit) a sigmoid. Z čoho v praxi je najpoužívanejšia práve ReLU [3].

## 1.2 Vrstvy neurónovej siete

Existujú 3 základné vrstvy doprednej neurónovej siete.

*Vstupná vrstva* obsahuje vstupné parametre. Každý parameter predstavuje nejakú abstrakciu, resp. má nejakú vlastnosť. Týchto parametrov môže byť niekoľko málo, alebo (v praxi častejšie) kludne aj niekoľko miliónov [5]. Čím viac parametrov neurónová sieť má, tým viac je zložitejšia a tým pádom môže predstavovať dôveryhodnejší model. Toto tvrdenie je však nutné brať s určitou nadsádzkou, nakoľko počet vstupných parametrov nie je zďaleka jediná známka kvality neurónovej siete. Vstupné parametre sú nezávislé a vstupné neuróny nedisponujú aktivačnou funkciou. Hodnoty vstupných parametrov sú načítavané zo súboru údajov zvaného dátová sada, alebo dátová množina. Vstupná vrstva je ďalej spojená so skrytou vrstvou.

*Skrytá vrstva* môže byť často zložená z viacero ďalších vrstiev, ku ktorým pravidla nie je možné pristupovať priamo. Pri hlbokom učení môžu mať neurónové siete celkovo až niekoľko desiatok vrstiev [8]. Čím viac skrytých vrstiev sieť má, tým môže viac kombinovať signály medzi neurónmi, má väčšiu flexibilitu a dokáže sa zamerať na viac špecifické detaily.

*Výstupná vrstva* obsahuje výstupné neuróny, ktoré predstavujú výstup celej neurónovej siete.

## 1.3 Učenie neurónovej siete

Učenie, alebo tréovanie neurónovej siete je jej základný proces (algoritmus). Mnoho konvenčných počítačových algoritmov nemení svoje naprogramované chovanie. Zatiaľ čo neurónové siete si svoje vnútorné parametre menia každým prechodom vrstiev na základe doposiaľ naučených poznatkov. Učenie v tejto problematike spočíva v opakovanej zmene váh na vstupoch konkrétnych neurónov v neurónovej sieti. Existujú štyri základné spôsoby učenia:

1. Učenie s učiteľom – typ učenia, pri ktorom je vopred známy výstup pre konkrétny vstup. V tejto práci rozvediem iba tento prístup.
2. Učenie bez učiteľa – typ učenia, pri ktorom nie je vopred známy výstup.
3. Spätnoväzobné učenie – obsahujú slučky, využívajú ich rekurentné neurónové siete.
4. Kombinované učenie - kombinácia tretieho s jedným z prvých dvoch spomenutých spôsobov

### 1.3.1 Dataset - dátová množina

Dátová množina (angl. **dataset**) je normalizovaná množina údajov určitého typu. To, akého typu sú dáta v nej obsiahnuté závisí na použití neurónovej siete. Je rozdelená na množinu *trénovaciu*, *testovaciu* a *validačnú* množinu.

Trénovacia množina je použitá pri učení neurónovej siete tzv. s učiteľom. Validačná množina sa používa pre kontrolu pretrénovania neurónovej siete. Testovacia množina je použitá na určenie presnosti finálneho natrénovaného modelu.

### 1.3.2 Backpropragation - učenie s algoritmom spätného šírenia

Pri učení s učiteľom je využitý algoritmus spätného šírenia (angl. **backpropagation**). Na vstupné hodnoty je použitá aktivačná funkcia, ktorej výsledok je označený ako skutočný výstup  $\hat{y}$ . Skutočný výstup  $\hat{y}$  sa porovná s očakávaným výstupom  $y$  (rola učiteľa). Porovnávanie prebieha prostredníctvom *chybovej funkcie*  $E$  (angl. *loss function*). Stredná kvadratická chyba MSE (angl. Mean Squarred Error) je najrozšírenejšia chybová funkcia široko používaná v lineárnej regresii. Stredná kvadratická chyba je vyjadrená ako: [9]

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (1.2)$$

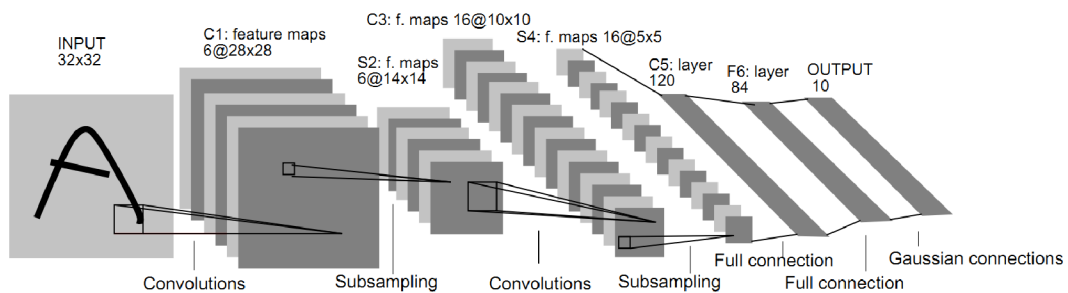
Cielom učenia je na trénovacej množine túto chybu minimalizovať. Algoritmus spätného učenia preto používa spätnú adaptáciu váh. Rozšíreným postupom je znižovanie chybového **gradientu** (angl. *gradient descent*), ktorý určuje smer a strmnosť

sklonu chybovej funkcie. Na základe gradientu sú vo fáze učenia aktualizované váhy v správnom smere voči lokálnemu minimu chybovej funkcie.

Veľkosť kroku pri úprave váh určuje *koeficient učenia*  $\alpha$  (angl. **learning rate**). Nadobúda nezáporných hodnôt z intervalu  $[0; 1]$ , kde 0 znamená, že k úprave váh nedôjde. Čím väčší koeficient je, tým je miera učenia rýchlejšia, avšak za cenu nižšej presnosti trénovaného modelu.

## 1.4 Konvolučné neurónové siete

*Konvolučná neurónová sieť* (CNN, angl. *Convolutional neural network*) je typ dopredných viacvrstvových neurónových sietí, ktoré používajú operáciu diskretnej konvolúcie aspoň v jednej jej vrstve. Pracujú s učením spätnou propagáciou a v širokej miere sú zastúpené pri spracovávaní obrazu. Boli navrhnuté pre rozpoznávanie obrazových charakteristík priamo z pixelov obrázka s minimálnym predspracovaním. Ukážka konvolučnej neurónovej siete je na obrázku 1.2



Obr. 1.2: Zobrazenie konvolučnej neurónovej siete LeNet5, zdroj [10]

Konvolučná neurónová sieť obsahuje tri typy vrstiev: *konvolučnú*, *združovaciu* a *plne prepojenú vrstvu*. Výstup konvolučnej vrstvy je vstupom pre združovaciu vrstvu, ktorá zmenší veľkosť príznakových máp. Tieto dve vrstvy sa striedajú až ich výstup je nakoniec pripojený do plne prepojenej vrstvy.

### 1.4.1 Konvolučná vrstva

Konvolučná vrstva je základnou časťou CNN. Zameriava sa na učenie vzorov z obrazových dát na jej vstupe. Vstupný obrázok je privedený na prvú konvolučnú vrstvu. Táto vrstva je zložená z konvolučných príznakových máp, ktoré medzi sebou zdieľajú svoje parametre. Prvá vrstva deteguje relatívne jednoduché príznaky, akými sú napríklad hrany, či okraje. V každom ďalšom prechode konvolučnými vrstvami sú príznaky stále viac abstraktné.

| Hodnoty jasů<br>původného obrázku |   |                 |                 |                 |
|-----------------------------------|---|-----------------|-----------------|-----------------|
| 1                                 | 1 | 1               | 0               | 0               |
| 0                                 | 1 | 1               | 1               | 0               |
| 0                                 | 0 | 1 <sub>x1</sub> | 1 <sub>x0</sub> | 1 <sub>x1</sub> |
| 0                                 | 0 | 1 <sub>x0</sub> | 1 <sub>x1</sub> | 0 <sub>x0</sub> |
| 0                                 | 1 | 1 <sub>x1</sub> | 0 <sub>x0</sub> | 0 <sub>x1</sub> |

| Konvolučné<br>jadro |   |   |
|---------------------|---|---|
| 4                   | 3 | 4 |
| 2                   | 4 | 3 |
| 2                   | 3 | 4 |

Obr. 1.3: Násobenie matíc ako príklad 2D diskřetnej konvolúcie

CNN obsahuje *konvolučné jadrá* (angl. *convolutional kernels*). Jadro je matica o malej veľkosti, napr.  $3 \times 3$ , ktorá sa pixel za pixelom posúva po celom vstupnom obrázku. Neuróny tejto vrstvy počítajú v každom kroku 2D diskřetnú konvolúciu medzi jadrom a časťou obrázku, nad ktorým sa jadro práve nachádza. Výsledok operácie konvolúcie je ukladaný do tzv. *príznakových máp* (angl. *feature maps*). Ak jadro presiahne okraj obrázku, daná plocha sa vyplní nulami [11].

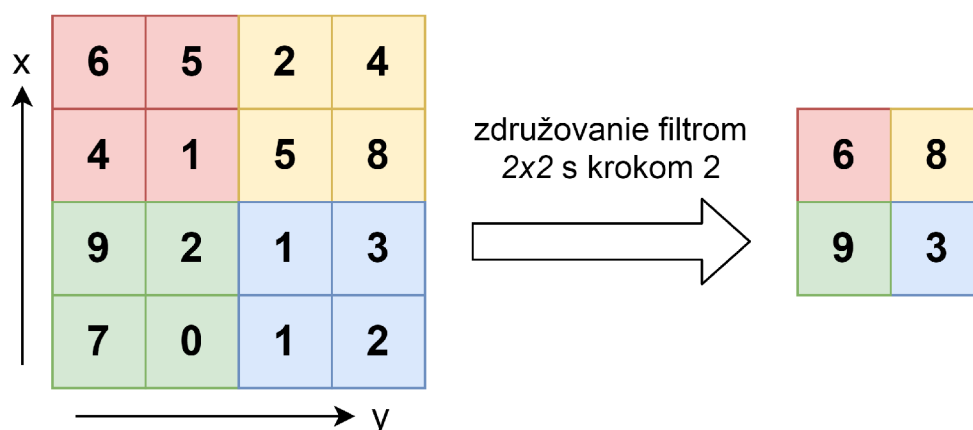
$$\mathbf{f} * \mathbf{k} = f(x, y) * k(x, y) = \sum_{(m,n) \in O} k(x - m, y - n) f(m, n), \quad (1.3)$$

kde funkcia  $f(x, y)$  vracia hodnoty pixelov vstupného obrázku  $\mathbf{f}$  na súradniciach  $x$  a  $y$ ,  $k(x, y)$  je funkcia konvolučného jadra  $\mathbf{k}$ , ktorej výstupom je hodnota na daných súradniciach.  $O$  je okolie o veľkosti  $m \times n$  a túto veľkosť má aj konvolučné jadro.

### 1.4.2 Zduřovacia vrstva

Táto vrstva zmenšuje veľkosti príznakových máp konvolučnej vrstvy a dáva tak väčšiu variabilitu vstupného obrázku. Príznaková mapa je po častiach postupne prekryvaná menšou štvorcovou maticou tzv. *filtr*om o veľkosti, napr.  $2 \times 2$ , pričom vynecháva tú časť, ktorú už prekryla.

V každom kroku je výsledkom maximálna (angl. *Max pooling*) alebo priemerná hodnota (angl. *Average pooling*) danej časti príznakovej mapy (viď obr. 1.4).



Obr. 1.4: Ukážka algoritmu Max pooling zobrazuje dvojnásobné zmenšenie vstupného obrazu reprezentovaného maticou o veľkosti 4x4

### 1.4.3 Plne prepojená vrstva

Dáta na výstupe konvolučnej vrstvy majú veľkú mieru abstrakcie. Na výstup posledných konvolučných a združovacích vrstiev je dodaná plne prepojená vrstva 1.2. V tejto vrstve sú neuróny plne prepojené s predošlými aj nasledujúcimi neurónmi a jej výstup je ďalej spracovávaný ako v bežných umelých neurónových sieťach. Táto vrstva predstavuje lineárnu operáciu pre spracovanie naučených príznakov poslednej konvolučnej vrstvy v ich nelineárnej podobe.

## 1.5 Reziduálne neurónové siete

*Hlboké reziduálne siete* sú v oblasti zvyšovania kvality obrazu čoraz populárnejšie [5, 13, 14, 15, 8, 18] už od svojho uvedenia v práci *ResNet* [8] v roku 2015. Dôkazy z minulých rokov [21] potvrdzujú výraznú dôležitosť počtu vrstiev modelu neurónovej siete. Ako sa ukázalo väčší počet vrstiev vedie k väčšiemu obohateniu nájdených vzorov pri spracovaní obrazu, tým pádom výsledný obraz môže byť kvalitnejší [22]. Samotné pridávanie vrstiev je však dosť problematické. S rastúcim počtom vrstiev neurónovej siete exponenciálne rastie aj počítaný chybový gradient, kde môže dôjsť až k tzv. *explodovaniu gradientu* (angl. *exploding gradient*), či k tzv. *vymiznutiu gradientu* (angl. *vanishing gradient*) [23]. Preto je optimalizácia veľmi hlbokých neurónových sietí náročný proces.

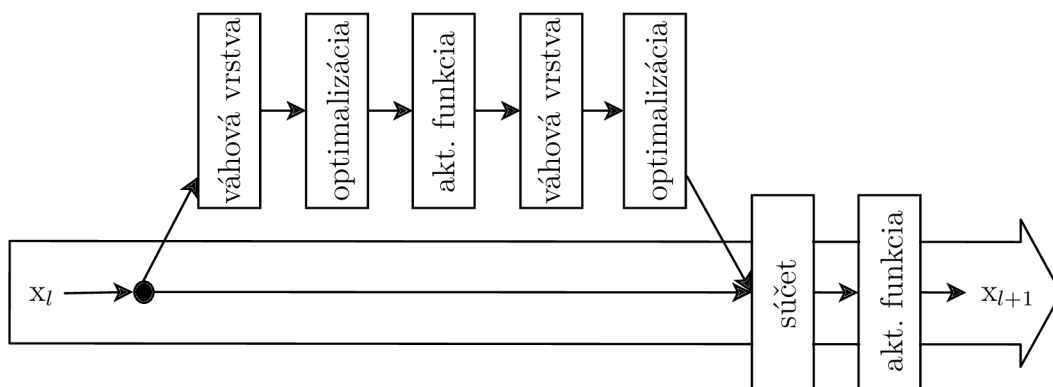
Riešenie môže poskytnúť reziduálne učenie, ktoré definuje tzv. *reziduálne bloky*, ktoré znížia počet vrstiev nutných na prechod neurónovou sieťou pri zachovaní dostatočnej presnosti modelu a odstránil tak problém spojený s hľadaním gradientu.

## 1.5.1 Reziduálny blok

Každý reziduálny blok je vyjadrený ako:

$$\begin{aligned}y_l &= h(x_l) + F(x_l, W_l), \\x_{l+1} &= f(y_l),\end{aligned}\tag{1.4}$$

kde  $x_l$  je vstup a  $x_{l+1}$  výstup  $l$ -tého reziduálneho bloku,  $F$  je reziduálna funkcia. V práci *ResNet* je  $h(x_l) = x_l$  *identické zobrazenie* (angl. *identity mapping*) a  $f$  je aktivačná funkcia ReLU. Hlavnou myšlienkou reziduálnych sietí je učenie reziduálnej funkcie  $F$  s ohľadom na  $h(x_l)$  za použitia identického zobrazenia  $h(x_l) = x_l$ .



Obr. 1.5: Reziduálny blok uvedený v práci ResNet [8]

Na obrázku 1.5 je reziduálny blok, ktorý pozostáva z dvoch vetiev. Prvá je tvorená váhovými, resp. konvolučnými vrstvami s jej optimalizáciami (angl. *batch normalization*) a aktivačnou funkciou ReLU. Druhá vetva je identita  $x$  reziduálneho bloku  $l$ . Výhodou reziduálneho učenia je tzv. *preskakovanie blokov* (angl. *skipping connections*), aby výstup  $x_{l+1}$  bolo možné aproximovať identitou  $x$  bez nutného prechodu konvolučnými vrstvami. Avšak, súčte oboch vetiev na konci reziduálneho bloku zjednodušuje tok naučených informácií [8]. Ako odpoveď na tento problém vznikol koncept husto prepojených blokov (angl. *Dense blocks*).

## 1.6 Sieť husto prepojených blokov

Práca *DenseNet* z roku 2016 predložila myšlienku husto prepojených blokov, ktorá bola odpoveďou na problém s využiteľnosťou informácií medzi vrstvami v hlbokých konvolučných neurónových sietach.

Na potrebu riešiť tento problém, ktorý vznikol s pribudajúcim počtom vrstiev v modeli, poukázali práce v rovnakej [8] a [7]. V *DenseNet* nie je nutné na začiatku tréningu potrebné inicializovať veľké množstvo parametrov ako v spomínaných

sieťach. V týchto sieťach je cesta od vstupnej vrstvy k vrstve výstupnej veľmi dlhá a môže sa stať, že niektoré podstatné informácie sa počas tejto cesty vytratia, nebudú ďalej zohľadňované a to sa vo finále prejaví zníženou efektivitou modelu. DenseNet priniesol priamy prístup ku gradientom z chybovej funkcie a k originálnemu vstupnému obrázku. Všetky predávané informácie z každej vrstvy sú predstavené prostredníctvom príznakových máp.

Štruktúra siete pozostáva z husto prepojených blokov, pričom medzi nimi sa nachádzajú tzv. *prechodové vrstvy* (angl. transition layers). Tieto vrstvy predstavujú filtre a starajú sa o zmenšenie rozmeru využitia optimalizácie. Model DenseNet reťazí príznakové mapy a každou pridanou vrstvou vzrastá dimenzia tohto kanálu. V modeli sa množstvo získaných informácií z  $l$  vrstiev reguluje pomocou tzv. *tempa rastu* (angl. growth rate):

$$k_l = k_0 + k * (l - 1), \quad (1.5)$$

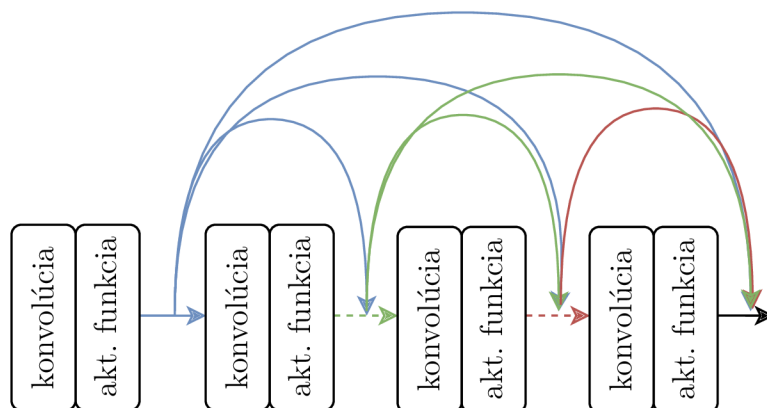
kde hyperparameter  $k$  je tempo rastu. [14]

### 1.6.1 Husto prepojený blok

Hlavne zmena oproti minulým prácam nastala v priamom prepojení výstupu každej vrstvy na vstupy všetkých nasledujúcich vrstiev neurónovej siete (viď 1.6).  $l$ -tá vrstva na vstupe obdrží príznakové mapy zo všetkých predchádzajúcich vrstiev  $x_0, x_1, \dots, x_{l-1}$ , pričom platí:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (1.6)$$

kde  $[x_0, x_1, \dots, x_{l-1}]$  odkazuje na konkatenáciu (zreťazenie)  $H(l)$  príznakových máp, ktoré sú výstupom z vrstiev  $0, \dots, l - 1$ . [14]



Obr. 1.6: Husto prepojený (Dense) blok uvedený v práci DenseNet [14]

## 2 Techniky rekonštrukcie obrazu

Technika rekonštrukcie obrazu je kvalitatívna metóda, ktorá sa na základe nedokonalého (poškodeného) obrazu snaží hľadať jeho dokonalú podobu [12]. Typ a miera poškodenia obrazu môže mať rôzne podoby a príkladom môže byť rozmazaný snímok, nízke rozlíšenie či šum v obraze.

V tejto práci sú rozobraté dve techniky rekonštrukcie obrazu za použitia hlbokých konvolučných neurónových sietí. Podkapitola *Super-rozlíšenie* vymedzuje pojem a problematiku zvyšovania rozlíšenia obrazu v chronologickom prehľade kvalitatívnych metód, ktoré významne prispeli do tejto oblasti. Druhá podkapitola *Odstraňovanie šumu z obrazu* stručne približuje túto tému v rozobratých prácach, ktoré s ňou súvisia.

Zo zadania vyplýva nutnosť orientácie práce na zvyšovanie kvality snímkov ľudí pre bezpečnostné zložky. Preto som zvolil techniku super-rozlíšenia obrazu ako základ tejto práce s dôrazom pre testovanie modelov na obrazových dátach tváří ľudí. V ďalších kapitolách tejto práce sa na odstraňovanie šumu z obrazu preto už nedovoľávam.

### 2.1 Super-rozlíšenie

*Super-rozlíšenie* (SR) je technika zlepšovania kvality obrazových dát získavaním vysokého rozlíšenia na výstupe zo vstupných obrázkov s pôvodne nízkym rozlíšením (viď obr. 2.1) [28]. Obyčajne sú modely trénované metódou učenia s učiteľom, kde je k obrázku s nízkym rozlíšením priradený jeho variant vo vysokom rozlíšení.



Obr. 2.1: Zvyšovanie kvality obrazových záznamov, model *FSRGAN* [18]

Pre zoznámenie so super-rozlíšením sú v tejto práci rozobraté niektoré existujúce modely počínajúc od najdôležitejších po najnovšie. Použité modely sa týkajú super-rozlíšenia jednotlivých statických obrázkov (SISR, angl. *Single Image Super-Resolution*).



### 2.1.1 Hlavné prínosy v oblasti super-rozlíšenia

Zásadným skokom v implementovaní super-rozlíšenia prostredníctvom hlbokých konvolučných neurónových sietí (DCNN, angl. *Deep Convolutional Neural Network*) bola práca *ImageNet Classification with Deep Convolutional Neural Networks* [10] z roku 2012. Autori v nej predstavili model siete, ktorý vychádzal z modelu *SRCNN* [30]. Ako dátovú sadu použili databázu ImageNet [31], ktorú ďalej rozšírili o ďalšie transformácie pôvodných obrázkov (tzv. *augmentácia dát*). Dostali tak 60 miliónov parametrov v 22 tisíc triedach a model dosahoval výborné kvalitatívne výsledky. V modeli zefektívnilo operáciu konvolúcie a pridali vlastnosť náhodného zahadzovania neurónov v plne prepojených vrstvách (angl. *dropout*). Aj keď práca bola primárne určená na klasifikáciu objektov v obraze, hlboko zasiahla všeobecné použitie konvolučných neurónových sietí pri spracovávaní obrazu. [10]

Rok 2015 priniesol spomínané použitie reziduálneho bloku ako odpoveď na poukázané problémy spojené s volením vhodného gradientu pri zmene váh neurónov [8]. So vzrastajúcim počtom vrstiev v sieťach bolo čoraz viac obtiažné pracovať s príznakovými mapami, a preto vznikol koncept husto prepojených blokov [14].

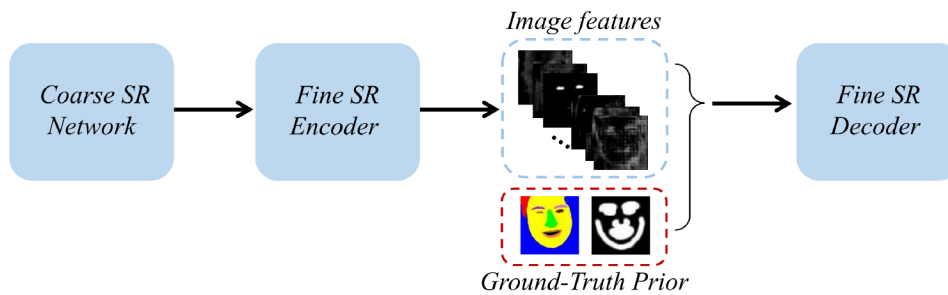
V septembri roku 2016 na základe celkom odlišnej architektúry práce *Generative Adversarial Network* [17] bola predstavená nová architektúra neurónových sietí v oblasti super-rozlíšenia (obr. 3.2), ktorú autori nazvali *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network* (SRGAN) [16] a zožala veľký úspech. Jej výhoda spočívala v použití novej chybovej funkcie, ktorá viac zohľadňovala percepčné vnímanie obrazu podobne ako človek ako aj v návrhu novej architektúry, ktorá bola schopná rekonštruovať realistické textúry v obraze. Ďalšie informácie ohľadom tohto modelu obsahuje nasledujúca podkapitola.

#### FSRNet a FSRGAN - modely zvyšovania kvality obrazov tváří

Z modelu reziduálnych sietí vychádzajú autori práce *FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors* z roku 2017. V nej predstavili model FSRNet, resp. FSRGAN, ktoré sa sústreďia výlučne na tváre ľudí.

Model FSRNet využíva geometrických vlastností obrazu tváre a berie do úvahy jej markanty. Má dve hlavné časti. Prvá časť zvyšuje kvalitu obrazu s nízkym rozlíšením tzv. "*nahrubo*" (angl. Coarse SR Network). Tento obraz je vstupom do druhej časti "*jemnej*" SR siete (angl. Fine SR Network), pozostávajúcu z dvoch vetiev (obr. 2.2).

Prvá vetva je tzv. *SR enkodér* (angl. SR Encoder). SR enkodér hľadá obrazové vzory ako je to u iných reziduálnych modeloch neurónových sietí. Jeho architektúru tvorí 12 reziduálnych blokov. Druhá vetva je *sieť predbežného odhadu*. (angl. prior



Obr. 2.2: Ukážka architektúry modelu *FSRNet* [18]

estimation network). Táto vetva hľadá tvárové prvky (angl. facial landmark heat-maps), akými sú napríklad oči, obočie či nos. Oba výstupy z týchto dvoch vetiev sú vstupom pre záverečnú časť SR siete, SR dekodér (angl. SR Decoder). Na jeho výstupe je získaný zväčšený super-rozlíšený snímok. [18]

Z modelu SRGAN a práce o podmienených GAN sieťach *Image-to-Image Translation with Conditional Adversarial Networks* [19] vychádza model FSRGAN predkladaný v rovnakej práci ako FSRNet ([18]).

Predložené metódy FSRNet a FSRGAN sa sústreďujú na zväčšovanie kvality obrazových snímkov ľudí, čo je aj hlavný zámerný cieľ tejto práce. Avšak o oba modely od svojho vzniku bol malý záujem širokej odbornej verejnosti (nízky počet implementácií na verejných platformách, práca nebola neaktualizovaná od svojho začiatku, apod.). Na základe toho je práca orientovaná na novšie prístupy a architektúry iných modelov v rámci oblasti super-rozlíšenia.

### 2.1.2 NTIRE - výzva nových techník v oblasti super-rozlíšenia

*NTIRE 2017 Challenge on Single Image Super-Resolution* je výzva v oblasti rekonštrukcie obrazu konkrétne pre oblasť super-rozlíšenia a je určená pre širokú odbornú verejnosť (<http://www.vision.ee.ethz.ch/ntire17/>). Od 2017 je každoročne otváraná a vďaka nej sa do oblasti spracovania obrazu prispelo viacerými úspešnými patentmi a trendmi. Súťaží sa v tímoch, pričom organizátori určujú ciele, o ktoré sa zúčastnení budú snažiť. Výsledky sú hodnotené objektívnymi metrikami, ktoré berie do úvahy aj autor tejto práce.

Trénovanie a testovanie prebieha na vlastnom datasete poskytovanom organizátormi výzvy, ktorá je po udalosti zverejnená prostredníctvom <https://data.vision.ee.ethz.ch/cv1/DIV2K/>. Dataset má názov DIV2K (HR obrázky majú rozlíšenie 2K) a vo výzve NTIRE 2018 sa oproti minulému roku zmenila tréningová a validačná časť obrázkov s nízkym rozlíšením.

Vo výzve NTIRE 2017 sa na prvej priečke umiestnil model *Enhanced Deep Residual Networks for Single Image Super-Resolution* (EDSR) [5]. EDSR zväčšil a zoptimalizoval klasickú formálnu reziduálnu neurónovú sieť. Taktiež priniesol koncept tréovania viacúrovňového zväčšovania obrazu v rámci jedného modelu *Multi-scale Deep Super-Resolution system* (MDSR). NTIRE 2018 vyhral model vychádzajúci a stále úzko súvisiaci s modelom EDSR, ktorý nesie názov *Wide Activation for Efficient and Accurate Image Super-Resolution* (WDSR).

V súčasnosti (rok 2019) medzi modelmi v oblasti rekonštrukcie obrazu sa na najvyšších priečkach drží model *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks* [4]. Táto metóda bola predstavená v rámci inej súťaže v oblasti super-rozlíšenia a síce v PIRM2018-SR Challenge (<https://www.pirm2018.org>). ESRGAN zlepšil tri kľúčové komponenty referenčného modelu SRGAN: architektúru siete, protikladnú chybu a chybu vnímania. Týmto počínom sa z vysokofrekvenčných detailov zväčšeného obrazu znížila prítomnosť artefaktov, čo bola najväčšia slabina modelu SRGAN.

## 2.2 Odstraňovanie šumu z obrazu

Predmetom techník odstraňovania šumu z obrazu je zníženie intenzity šumu v digitálnom obraze pri zachovaní čo najväčšej miery detailov (pôvodnosť obsahu). V tejto podkapitole bude načrtnutý dôvod prečo sa šum v digitálnom obraze nachádza a metódy jeho odstraňovania z obrazu.

### 2.2.1 Digitálny šum v obraze

Šum je výsledkom chýb v procese získavania obrazu, ktorého hodnoty pixelov neodrážajú skutočnú hodnotu intenzity snímanej scény. Digitálny snímok získaný optickým senzorom vždy obsahuje nejaký druh šumu. Existuje niekoľko druhov šumu: *impulzný* (Salt-and-papper), *aditívny* (Gaussovský) či *multiplikatívny* šum. Spolu so zhoršujúcimi sa podmienkami snímaného prostredia (napr. nedostatočné osvetlenie) sa zvyšuje výskyt šumu v obraze.

Digitálny obraz je kódovaný ako matica šedotónových a farebných hodnôt. Každá dvojica  $(i, u(i))$ , kde  $u(i)$  predstavuje hodnotu v bode  $i$ , teda jeden pixel  $[px]$ . Jedným z hlavných obmedzení kvality obrazových dát je zašumenie (angl. noise). Nasledujúci vzorec zobrazuje získanú hodnotu pixelu  $v(i)$ :

$$v(i) = u(i) + n(i), \quad (2.1)$$

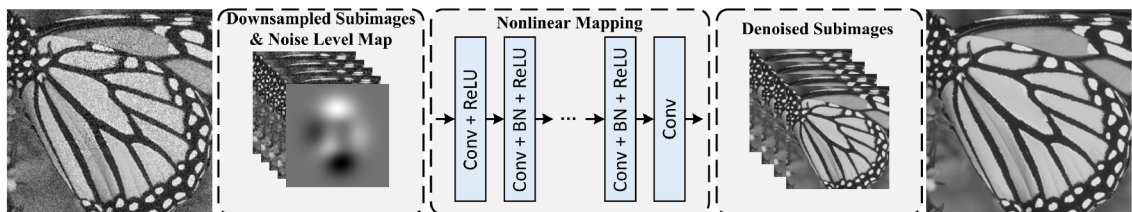
kde  $n(i)$  je šum (v praxi biely aditívny Gaussovský šum so smerodatnou odchýlkou  $\sigma$ ), ktorý sa sčíta so skutočnou hodnotou pixelu  $u(i)$ . [20]

Existujú rôzne rozšírené techniky na odstránenie šumu z obrazu. Základné metódy potlačujú ostré rozdiely medzi susednými pixelmi poškodeného snímku. Jedná sa buď o lokálne vyhladzovacie metódy s použitím ich špeciálnych filtrov (mediánový, Gaussov, anizotropický, filter okolia, apod.), či filtráciu vo frekvenčnej oblasti obrazu tzv. prahovaním, prípadne nelokálne. Zo spomenutých základných metód vychádzajú metódy postavené na hlbokom učení.

## 2.2.2 Odstraňovanie šumu pomocou hlbokého učenia

Posledné roky dosvedčujú veľký potenciál hlbokých konvolučných neurónových sietí v rámci rekonštrukcie obrazových dát. V roku 2016 bol predstavená metóda *Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising* [32]. Táto metóda je založená na reziduálnom učení s optimalizáciou konvolučných vrstiev. Naproti predošlým projektom, tento model má väčší počet vrstiev a dokáže pracovať s vopred neznámou mierou zašumenia Gaussovským šumom (angl. blind Gaussian denoising).

V nasledujúcom roku (2017) bola predstavená práca *FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising* [33]. Dôležitým prínosom tejto architektúry bolo zavedenie tzv. máp stupňov zašumenia, ktoré boli predávané na vstup siete (obr. 2.3). Model dosahoval dobré výsledky pri efektívnom odstraňovaní širokého spektra šumu.



Obr. 2.3: Architektúra modelu *FFDNet* uvedená v práci [33]

Predmetom techník rekonštrukcie obrazu je zotavenie originálneho obrazu  $x$  z poškodeného obrazu  $x_0$ . Takýto problém je možné zapísať ako optimalizovanú úlohu:

$$\min_x E(x; x_0) + R(x), \quad (2.2)$$

kde  $E(x; x_0)$  značí obrazové dáta a  $R(x)$  je regularizér, ktorý zastupuje informácie známe o vlastnostiach týchto obrazových dát.

Obrazové dáta  $E(x; x_0)$  je v súčasnosti jednoduché navrhnuť v širokom spektre problémov, ale navrhnuť schému vlastností týchto dát je omnoho zložitejší proces.

Túto výzvu prijali autori práce *Deep Image Prior* [34]. Vytvorili model siete, ktorý sa sústreďuje na aproximáciu vlastností zašumených obrázkov v nízkej kvalite

na vstupe. Získané vlastosti si tak vystačia v procese rekonštrukcie obrazu samé. Sieť teda nepotrebuje originálne vyčistené obrázky vo vysokom rozlíšení. Model je náhodne inicializovaný a môže byť použitý aj v rámci super-rozlíšenia.

## 3 Modely hlbokého učenia pre zvyšovanie kvality obrazu technikou super-rozlíšenia

Práca obsahuje 4 modely, z ktorých každý priniesol oblasti super-rozlíšenia určité výrazné zlepšenie. Táto podkapitola obsahuje bližší opis ich architektúr a ich implementačné detaily. Opis začína použitým prostredím a nástrojmi, ďalej opisuje vybrané datasety a na konci každého modelu sú opísané jeho implementačné detaily.

### 3.1 Použité nástroje

Implementovanie modelov neurónových sietí je komplexný a zdĺhavý proces, ktorý si žiada dostatok výpočtového výkonu. Jedná sa prevažne o výpočtovo zložité paralelné prepočítavanie matíc. To zaisťuje špecializovaný hardvér - grafické a tenzorové karty. Pre softvérovú časť v dnešnej dobe existuje mnoho nástrojov a rozhraní. Medzi najpoužívanejšie programovacie jazyky (nielen v oblasti super-rozlíšenia) patrí jazyk Python a jeho knižnice PyTorch, Theano a *Tensorflow*, resp. jeho nadstavba *Keras*, ktoré sú využité v rámci tejto práce.

V začiatkoch, práca prebiehala na notebooku s operačným systémom Windows 8.1. Hardvér notebooku obsahoval externú grafickú kartu NVIDIA K1100M Quadro, ktorá obsahuje 384 samostatných, tzv. stream procesorov (CUDA jadier). Nejedná sa o enormný výpočtový výkon, avšak pre potreby tejto práce je dostačujúci. Použité implementačné nástroje boli jazyk Python vo verziách 2.7, 3.5 a 3.6 v prostredí *Anaconda* (4.5.11), základné knižnice *tensorflow*, *tensorflow-gpu* (obe verzie 1.11), *tensorlayer* a *keras* (2.2.4) boli použité pre jednoduchú tvorbu modelov neurónových sietí a *tensorboard* pre grafické spracovanie výsledkov.

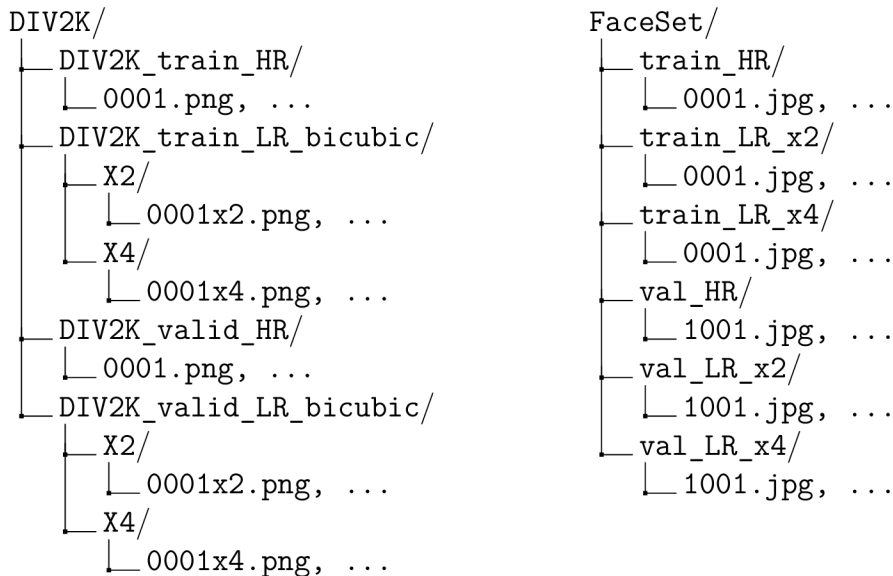
V ďalšej fáze bakalárskej práce bol projekt presunutý do prostredia Google Colaboratory, resp. *Google Colab* ([colab.research.google.com](https://colab.research.google.com)). Toto prostredie poskytuje vzdialený prístup k výpočtom na grafickej karte NVIDIA modelu *Tesla K80* (12GB RAM, 2496 stream procesorov). Prostredie *Google Colab* má zavedené a nastavené prostredie (interpretér jazyka Python, rozhrania TensorFlow a PyTorch, apod.) pre prácu s neurónovými sieťami a pre vývojára je tak možné sa fakticky hneď sústrediť na vlastný projekt.

### 3.2 Vybrané datasety

Táto práca je zameraná na zvyšovanie kvality obrazových záznamov z kamerových zariadení. Využíva modely hlbokých konvolučných neurónových sietí pomocou uče-

nia s učiteľom. To znamená, že datasety sú tvorené párami obrázkov, pričom prvý z nich plní rolu referenčného obrázka k zväčšovanému (SR) a je vo vysokom rozlíšení (HR) a druhý obrázok je v nízkom rozlíšení (LR).

Po konzultáciách s vedúcim práce autor zvolil tri databáze obrázkov (datasety) pre tréovanie a testovanie. Prvé dve datasety sú všeobecné, teda obsahujú rozmanité obrázky prírody, ľudí a vecí v rôznom kontexte. Tretí dataset je vytvorený autorom a obsahuje fotky ľudských tvárí. Adresárová štruktúra vybraných datasetov je na obrázku 3.1.



Obr. 3.1: Štruktúra použitých datasetov

Spomínané všeobecné datasety boli prevzaté z výziev NTIRE z rokoch 2017, resp. 2018 a nesú názov *DIV2K*. Tento dataset bol vytvorený špecificky pre spomínanú výzvu. Tréovaciu množinu tvorí 800 párov obrázkov (0001.png, ..., 0800.png), validačnú časť tvorí 100 párov obrázkov (0801.png, ..., 0900.png) a testovaciu časť taktiež 100 párov obrázkov (0901.png, ..., 1000.png). Každý pár obrázkov tvorí dvojica LR a HR obrázkov s rovnakým obsahom. Rok 2018 priniesol viac LR obrázkov v niekoľkých rozlíšeníach. Všetky obrázky sú vo formáte *PNG*. Obrázky v nízkom rozlíšení boli vytvorené bikubickou transformáciou príslušného obrázku s vysokým rozlíšením a formát ich názvu je *'0123xY.png'*, kde *Y* je úroveň rozlíšenia. Tréovanie na týchto datasetoch prebehlo v dvoch úrovniach rozlíšenia: 2x a 4x.

Vlastný dataset tvárí, ktorý autor nazval *FaceSet* vznikol špecificky pre účely tejto práce. Obsahuje 1250 obrázkov vo formate *JPG* rozdelených do tréovacej a testovacej časti pomerom 1000/250. Štruktúra je podobná adresarovej štruktúre datasetu *DIV2K*.

## 3.3 Super-Resolution using Generative Adversial Network (SRGAN)

Model vychádza z práce [16] publikovanej v roku 2016. V tom čase bolo nutné vyriešiť problém lepšej rekonštrukcie vyšších frekvencií (detailov) vo zväčšovanom obraze pri použití relatívne veľkého faktoru pre zväčšenie rozlíšenia (4x a viac).

Metóda priniesla novú chybovú funkciu - tzv. *chybu vnímania* - perceptual loss (vzorec 4.1), ktorá pozostáva z chyby obsahu (angl. content loss) a protikladnej chyby (angl. adversial loss). SRGAN tak produkuje reálnejšie a detailnejšie textúry ako doterajšie modely a vykazuje dobré výsledky na obvyklých testovacích sadách.

Na testovanie modelu bol použitý zdrojový kód zdieľaný na platforme *GitHub*.

### 3.3.1 Chybové funkcie modelu SRGAN

Nová chybová funkcia nahradila doterajšie metódy založené na optimalizácii MSE. Chyba vnímania je definovaná ako:

$$l_{SR} = l_{SR}^{VGG} + 10^{-3}l_{Gen}^{SR}, \quad (3.1)$$

kde  $l_{SR}^{VGG}$  je chyba obsahu použitá z chybovej funkcie VGG a  $l_{Gen}^{SR}$  je protikladná chyba.

#### Chyba obsahu

Chybová funkcia obsahu  $l_{SR}^{VGG}$  je založená na aktivačných vrstvách ReLU predtrénovanej siete VGG19 popísanej v práci [21]. Zapisuje sa ako:

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2, \quad (3.2)$$

pričom  $\phi_{i,j}$  je príznaková mapa  $j$ -tej konvolúcie pred  $i$ -tou vzorkovacou vrstvou v rámci siete VGG19 a  $W_{i,j}, H_{i,j}$  sú jej dimenzie.

#### Protikladná chyba

Pre lepšie porovnanie výstupu siete generátora je v rámci chybovej funkcie vnímania pridaná protikladná chyba. Táto chyba je kľúčová v procese rozhodovania diskriminátora, ktorý z predložených obrázkov je pôvodný a ktorý je rekonštruovaný generátorom. Vyjadrená je ako:



$$l_{Gen}^{SR} = \sum_{n=1}^N (-\log D_{\theta_D}(G_{\theta_G}(I_n^{LR}))), \quad (3.3)$$

kde  $D_{\theta_D}(G_{\theta_G}(I_n^{LR}))$  je pravdepodobnosť, že predložený obrázok rekonštruovaný obrázok je originálny ( $I_n^{HR}$ ).

### 3.3.2 Architektúra siete

Základom modelu SRGAN sú GAN siete, ktoré boli predstavené v roku 2014 v práci *Generative Adversarial Network* [17]. SRGAN obsahuje dve siete: generátor  $G$  a diskriminátor  $D$  (obr. 3.2), ktoré sú v praxi implementované ako samostatné neurónové siete.

Všeobecná súčinnosť týchto dvoch sietí spočíva v tom, že diskriminátor  $D$  odhaduje pravdepodobnosť  $\theta_D$ , že dáta na jeho vstupe sú reálne alebo nie. Generátor  $G$  sa snaží zvyšovať pravdepodobnosť  $\theta_G$ , že generované dáta sú reálne. Pri trénovaní sú diskriminátoru predložené dve dávky dát o rovnakej veľkosti (batch size): Jedna dávka je vybratá z reálnych (originálnych) dát a druhá z rekonštruovaných (falošných) dát. Pričom diskriminátor na začiatku nevie, ktorá dávka je z akých dát. V každej iterácii siete stanovujú svoje pravdepodobnosti a chybu ďalej propagujú späť na vstupy oboch sietí. Model sa takýmto spôsobom dokáže trénovať (zlepšovať pravdepodobnosť odhadu).

V praxi, generátor  $G_{\theta_G}$  generuje obrázky a predkladá ich diskriminátoru  $D_{\theta_D}$ , ktorý háda, či daný obrázok je vygenerovaný alebo originál. Vyššia chyba diskriminátoru dokazuje, že generátor generuje kvalitnejšie obrázky.

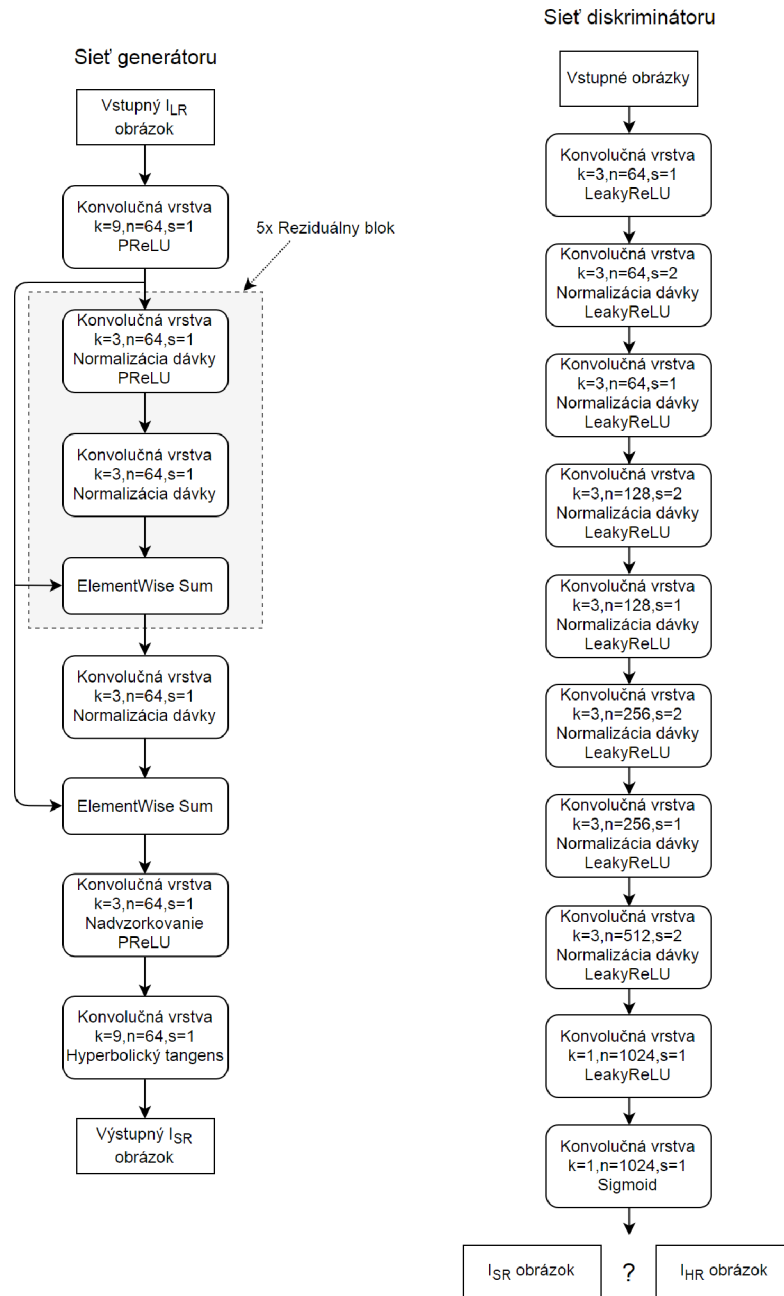
Autori predstavili generátor modelu ako doprednú konvolučnú neurónovú sieť  $G_{\theta_G}$ . Parametre siete generátora  $G_{\theta_G}$  sú  $\theta_G = \{W_{1:L}; b_1 : L\}$ , čo sú váhy a prahy  $L$ -tej vrstvy siete. Parametre  $\theta_G$  sú získané pomocou chybovej funkcie  $l^{SR}$ . Funkcia siete je vyjadrená ako:

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^N l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}), \quad (3.4)$$

kde  $I_n^{LR}$  a  $I_n^{HR}$  je dvojica trénovacích obrázkov  $n = 1, \dots, N$  s nízkym a vysokým rozlíšením.

Na obrázku 3.2 je zobrazená architektúra modelu aj s jej hlavnými parametrami.  $k$  je veľkosť konvolučného jadra,  $n$  počet príznakových máp,  $s$  je krok. Naľavo je generatívna sieť, ktorej vstupom je obrázok s nízkym rozlíšením  $I^{LR}$  a jej výstupom je rekonštruovaný obrázok  $I^{SR}$  v jeho super-rozlíšení. Ten je ďalej v druhej časti

predkladaný sieti diskriminátoru na vstup. Výstup diskriminátoru je potom pravdepodobnosť  $D_{\theta_D}(G_{\theta_G}(I^{LR}))$  či daný obrázok je rekonštruovaný  $I^{SR}$  alebo originál  $I^{HR}$ .



Obr. 3.2: Generátor a diskriminátor modelu SRGAN

## 3.4 Enhanced Super-Resolution using Generative Adversarial Network (ESRGAN)

Model vyhral výzvu *PIRM2018-SR Challenge*. Vychádza z modelu *SRGAN* a významne ho vylepšuje. Hlavnou zmenou v architektúre oproti referenčnému modelu je zaradenie reziduálnych husto prepojených blokov do vnútra nadradeného reziduálneho bloku. Vznikol tak nový viac úrovňový základný blok, tzv. *Reziduálny blok reziduálnych husto prepojených blokov* (angl. Residual-in-Residual Dense Block, RRDB). Autori projektu vďaka tomu dosiahli ľahšie trénovateľnú sieť pri ešte vyššej kapacite akú má SRGAN.

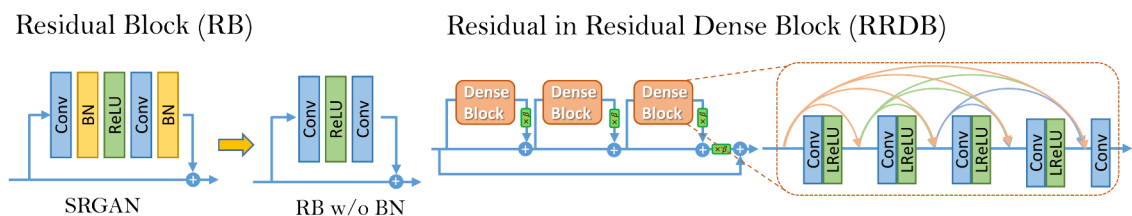
### 3.4.1 Architektúra siete

Architektúru referenčného modelu autori modelu ESRGAN na prvý pohľad nezmenili. Dôležité zmeny nastali skôr v jeho častiach. Implementovaním bloku RRDB, odstránením normalizačných blokov, vylepšením siete diskriminátora a hlavnej chybovej funkcie významne vylepšili pôvodný model SRGAN [16].

#### Vylepšenia oproti referenčnému modelu

Pôvodný základný reziduálny blok v sieti generátora autori nahradili blokom RRDB, čo je hlavný prínos k zlepšeniu jeho vlastností [4]. Tento blok využíva reziduálne učenie v dvoch vrstvách namiesto jednej (obr. 3.3). Spojenie s husto prepojenými blokmi prinieslo modelu možnosť spracovania väčšieho množstva informácie.

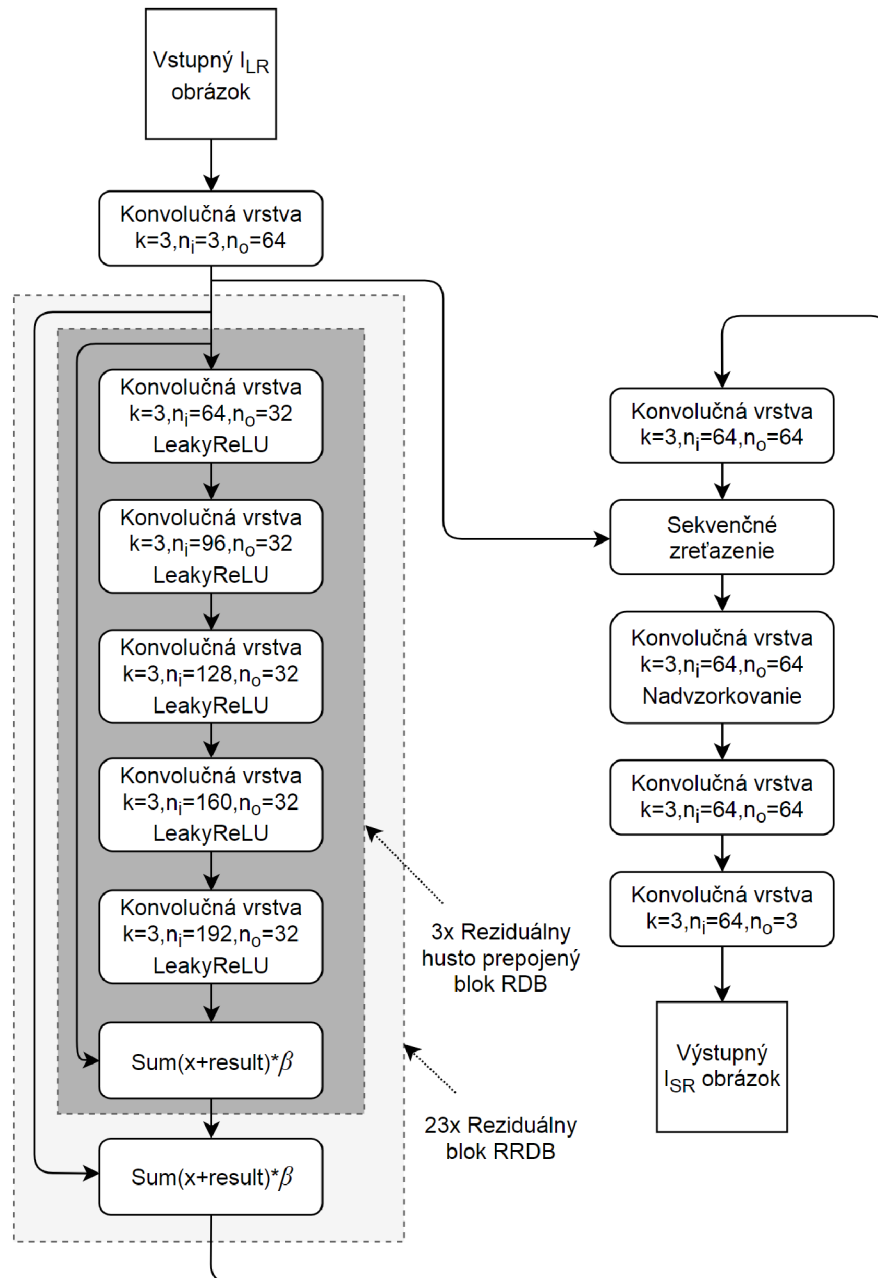
V modeli bola odstránená dávková normalizácia na výstupe každej konvolučnej vrstvy. Na skutočnosť, že to len zlepšilo vlastnosti modelu bolo poukázané už skôr v práci [5]. V nej autori potvrdili zníženie výpočtovej zložitosti pri zlepšení efektivity siete vo viacerých oblastiach rekonštrukcie obrazu vrátane modelov pre super-rozlíšenie.



Obr. 3.3: Reziduálny blok reziduálnych husto prepojených blokov RRDB použitých v modeli ESRGAN, publikovaný v práci [4]

Do modelu bol pridaný *parameter reziduálneho škálovania*  $\beta$  (angl. residual scaling parameter). Parameter  $\beta$  (z práce [29]) je konštanta, ktorá nadobúda hodnoty v intervale  $[0; 1]$  a slúži na znižovanie reziduálnych blokov pre nestabilitu pri použití veľkého počtu vrstiev. Počet parametrov potrebných pre inicializáciu modelu bol obmedzený kvôli zväčšeniu hĺbky siete pri zachovaní, resp. zlepšení učenia modelu.

Základná architektúra modelu ESRGAN je zobrazená na obrázku 3.4.



Obr. 3.4: Architektúra siete generátora modelu ESRGAN

## Relativistický diskriminátor

Zmena nastala aj v štruktúre diskriminátora založeného na práci *Relativistic GAN* [24]. Klasický diskriminátor  $D$  sa snaží predpovedať či obraz  $x$  na jeho vstupe je reálny  $x_r$  alebo falošný  $x_f$ . Naproti tomu, relativistický (priemerný) diskriminátor  $D_{Ra}$  (angl. Relativistic average Discriminator) sa snaží predikovať pravdepodobnosť reálny obraz  $x_r$  je relatívne viac realistickejší ako falošný  $x_f$ .

Štandardný diskriminátor v modeli SRGAN je vyjadrený ako funkcia  $D(x) = \sigma(C(x))$ , kde  $x$  je vstupný obraz (reálny alebo falošný) a  $\sigma$  prekladá skóre do pravdepodobnosti v intervale  $[0; 1]$ . V relativistických GAN sieťach diskriminátor  $D$  odhaduje pravdepodobnosť, že reálny obraz  $x_r$  sú realistickejšie ako náhodne vzorkovaný falošný obraz  $x_f$ :

$$D(\tilde{x}) = \sigma(C(x_r) - C(x_f)), \quad (3.5)$$

kde  $\tilde{x} = (x_r, x_f)$ . Pre potreby modelu pristupovať k spracovanému obrazu viac všeobecne a pre zníženie náhodnosti pri vzorkovaní párov obrázkov, autori predstavili koncept *relativistického priemerného diskriminátora*  $D_{Ra}$ .

$$\bar{D}(x) = \begin{cases} \sigma(C(x) - E_{x_f \sim Q} C(x_f)), & \text{ak } x \text{ je reálne} \\ \sigma(C(x) - E_{x_r \sim P} C(x_r)), & \text{ak } x \text{ je falošné.} \end{cases} \quad (3.6)$$

To znamená, že vždy, keď diskriminátor  $\bar{D}(x)$  dostane reálny obraz, vyhodnotí, ako viac je tento obraz realistickejší než priemerný falošný obraz. Analogicky to platí aj vice versa. Falošný obraz diskriminátor  $\bar{D}(x)$  porovnáva s priemerom všetkých reálnych obrazov v dávke. Táto formulácia relativistického diskriminátora umožňuje porovnať všetky možné kombinácie skutočných a falošných údajov v aktuálnej dávke.

Chyba diskriminátora je vyjadrená ako:

$$L_D^{Ra} = -E_{x_r}[\log(D_{Ra}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}(x_f, x_r))]. \quad (3.7)$$

Protikladná chyba generátora je formulovaná nasledovne:

$$L_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_f, x_r))], \quad (3.8)$$

kde platí  $x_f = G(x_i)$  a vstupným obrázkom sa rozumie  $x_i$ . Oproti základnému modelu SRGAN sú v protikladnej chybe obsiahnuté  $x_r$  a  $x_f$  a generátor tak berie do úvahy gradienty ako z generovaných obrazových dát, tak aj z dát reálnych. [4]

### 3.4.2 Popis implementácie

Použitý bol repozitár (<https://github.com/xinntao/ESRGAN>) uvedený v zadaní tejto práce, ktorý bol zverejnený na platforme GitHub. Jedná sa o pôvodný originálny projekt vytvorený pomocou rozhrania PyTorch autormi modelu. Tento repozitár však poskytuje iba možnosť testovania pomocou predtrénovaného modelu. Testovacia časť je ďalej rozvinutá v nasledujúcej kapitole, avšak autor projektu vytvoril jeden všeobecný repozitár (<https://github.com/xinntao/BasicSR>), ktorý obsahuje zdrojové kódy pre *GAN* orientované modely (SRGAN, ESRGAN, SFTGAN) a *PSNR* orientované modely (SRCNN, SRResNet apod.). Trénovací proces modelu ESRGAN teda modelu vychádza z jeho použitia. Adresárová štruktúra zobrazujúca časť repozitára použitého v tejto práci sa nachádza na obrázku 3.5.

```
BasicSR/ .....repozitár pre GAN a PSNR orientované modely.
├── codes/
│   ├── options/ .....adresár s konfiguračnými súbormi.
│   ├── scripts/ .....adresár so skriptami na predspracovanie.
│   ├── train.py .....súbor pre proces tréovania.
│   └── test.py .....súbor pre proces testovania.
├── experiments/ ..... adresár uložených aktuálnych modelov.
│   └── pretrained_models/ ..adresár predtrénovaných modelov.
├── results/ .....adresár s výsledkami.
└── tb_logger/ .....adresár s výsledkami pre tensorboard.
```

Obr. 3.5: Časť repozitára <https://github.com/xinntao/BasicSR>

Pre proces tréovania je nutné predspracovať obrázky v tréningovej sade, pripraviť predtrénovaný model a spustiť príkaz `python train.py` s parametrom `'-opt BasicSR/codes/options/train/train_ESRGAN.json'`, kde súbor `train_ESRGAN.json` definuje konfiguráciu tréovacieho procesu uloženého vo formáte JSON. Štruktúra parametrov tréovacieho procesu je na obrázku 3.6.

Použitý bol predtrénovaný model `RRDB_ESRGAN.pth` poskytnutý autormi projektu. Typ základného bloku ostal pôvodný (RRDB). Zvolené bolo štvornásobné zväčšenie, augmentácia obrázkov (otočenie a rotácia) bola zapnutá, veľkosť dávky je 16 a výrez z obrázka je o veľkosti 128x128 pixelov. Počiatočný koeficient učenia pre generátor a diskriminátor je 0.0001. Parameter  $\beta$  je 0.9. Počet iterácii je 30000 (pri zvolenej veľkosti dávky je potrebné urobiť 119 epoch).

Použitý dataset DIV2K bol upravený pre potreby projektu. Z pôvodného počtu 800 obrázkov v HR (2K) rozlíšení bolo vybratých 100 obrázkov pre tréovanie a 50 pre validáciu. Nakoľko sú tieto obrázky veľmi veľké, bolo nutné použiť skript `extract_subimgs_single.py`, ktorý každý z obrázkov rozdelil na niekoľko menších orezaných častí. Pre rýchlejšie vstupno-výstupné spracovanie bol využitý skript

```

train_ESRGAN.json
├── "name": .....název tréningového procesu.
├── "model": .....typ siete modelu.
├── "scale": .....faktor zväčšenia rozlíšenia.
├── "datasets": .....definícia datasetu.
├── "path": .....cesta k predtrénovanému modelu.
├── "network_G": .....inicializácia generátora.
│   ├── "nf": 64, "nb": 23, "gc": 32, apod
├── "network_D": .....inicializácia diskriminátora.
│   ├── "nf": 64, "norm_type": "batch", apod
├── "train": .....inicializácia procesu učenia.
│   ├── "lr_G": 0.0001, "beta1_G": 0.9, "niter": 30000, apod

```

Obr. 3.6: Základná štruktúra konfiguračného súboru `train_ESRGAN.json`

`create_lmdb.py`, ktorý pomocou knižnice LMDB celý dataset vložil do databázovej štruktúry typu kľúč-hodnota. Vzniklo tak 4040 tréningových a 1992 validačných obrázkov.

## 3.5 Enhanced Deep Super-Resolution (EDSR)

EDSR vyhral výzvu NTIRE 2017 Super-Resolution Challenge. Model vychádza z modelu reziduálnej neurónovej siete *SRResNet* [16] z roku 2016. Dôležitou zmenou voči pôvodnému modelu prinieslo odstránenie blokov optimalizácie konvolučných vrstiev (angl. *batch normalization layers*) v jej každom reziduálnom bloku. A taktiež odstránenie aktivačnej funkcie ReLU z konca reziduálnych blokov. Je to tak preto, lebo optimalizácia nájdených vzorov je komplexný proces náročný na systémové zdroje. Optimalizácia znižuje flexibilitu siete a zvyšuje nároky na výpočtový výkon. Ušetreným priestorom autori zvýšili výkon modelu zväčšením počtu jeho vstupných parametrov.

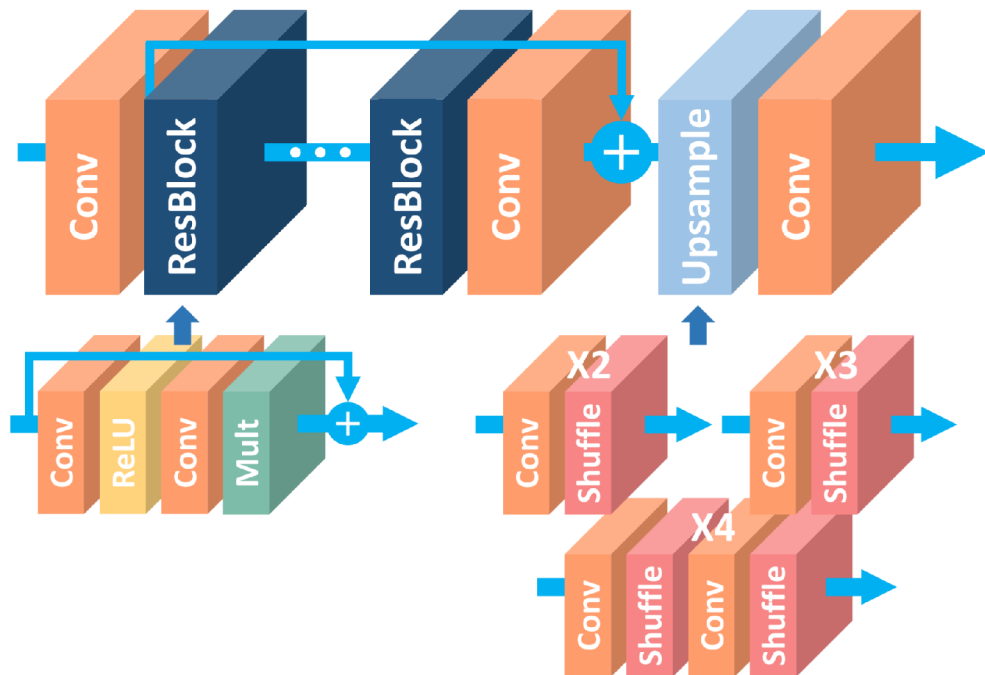
Spolu s predstaveným modelom EDSR autori uvideli aj model *Multi-scale Deep Super-Resolution system* MDSR. Tento model umožňuje v jednom procese natrénovať model pre viacero faktorov rozlíšenia a dokazuje porovnateľné vlastnosti ako jeho varianta EDSR pre jeden faktor zväčšenia.

### 3.5.1 Architektúra siete

Cieľom autorov predkladanej práce bolo zvýšiť výkon modelu. Je dokázané, že výkon v tejto súvislosti je možné významne zvýšiť dvoma základnými spôsobmi: zvýšiť počet parametrov alebo navrstviť do modelu ďalšie filtre a zväčšiť tak jeho hĺbku. Druhý spomínaný prípad je cesta, ktorou sa začali uberať. Všimli si jav popísaný v

[29], ktorý vraví o nestabilite siete, ktorá nastáva pri zväčšení príznakových máp nad určitú úroveň. Vyriešili to parametrom reziduálneho škálovania  $\beta$ , ktorý znižuje mieru informácií prechádzajúcimi sieťou. Jeho miesto je v každom reziduálnom bloku (blok *Mult* na obr. 3.7).

Architektúra modelu EDSR je zobrazená na obrázku 3.7.



Obr. 3.7: Blokový diagram modelu EDSR uvedený v práci [5]

### 3.5.2 Popis implementácie

V tejto práci bola použitá implementácia modelu zdieľaného na platforme *GitHub*. Implementácia sa drží praktickej časti, ktorá je popísaná v oficiálnej práci modelu [5].

Model pozostáva z 32 reziduálnych blokov, jednej vstupnej a jednej výstupnej samostatnej konvolučnej vrstvy a nadzorkovacej (angl. upsampling) vrstvy, pričom všetky tieto vrstvy obsahujú 256 príznakových máp. Hodnota parametra  $\beta$  je 0.1 a veľkosť dávky je 10. Faktor zväčšenia bol nastavený na dvojnásobok. Model v procese tréningu pracuje s výrezom stredu obrázka (angl. patch size) o veľkosti 300x300 pixelov. Pre optimalizáciu váh bol použitý Adam Optimizer [27].

Projekt sa skladá z piatich súborov: *model.py*, v ktorom je implementovaný model EDSR, *data.py*, v ktorom sú definované manipulačné metódy, *test.py*, kde je definovaná sekvencia príkazov pre testovanie, *train.py* je súbor pre tréningový proces a *utils.py*, čo je súbor s pomocnými funkciami, ktorý ďalej využíva súbor *model.py*.



Pre spustenie bol vytvorený skript (notebook) *EDSR.ipynb* spustiteľný v prostredí Google Colaboratory. Pri tréňovaní je nutné určiť cestu k datasetu. Povinný argument pri testovaní je názov obrázka, ktorý sa musí nachádzať v koreňovom adresári projektu. Ostatné parametre nie sú povinné a dajú sa zmeniť samostatne cez argument, alebo priamo v zdrojovom kóde súborov *train.py* a *test.py*.

## 3.6 Residual Dense Network for Image SR (RDN)

V práci RDN z roku 2018 [14] bol predstavený koncept reziduálneho husto prepojeného bloku (angl. Residual dense block, RDB) (obr. 3.8) pre vylepšenie extrakcie príznakov z obrázka. RDB spája výhody navzájom prepojených jednotlivých konvolučných vrstiev pre spoľahlivejšie end-to-end mapvanie pôvodného obrazu  $I_{LR}$  do obrazu s vysokým rozlíšením  $I_{HR}$  a výhody reziduálneho učenia v rámci jedného reziduálneho bloku. Princíp husto prepojených konvolučných vrstiev spočíva v prepojení všetkých výstupov predošlých vrstiev na všetky vstupy nasledujúcich vrstiev.

### 3.6.1 Architektúra siete

Reziduálna husto prepojená sieť RDN (obr. 3.8) pozostáva zo štyroch častí. Dve konvolučné vrstvy na začiatku tvoria prvú časť, tzv. plytkú sieť pre extrakciu príznakov (angl. Shallow Feature Extraction Net, SFENet). Prvá konvolučná vrstva extrahuje znaky  $F_{-1}$  zo vstupného obrazu  $I_{LR}$ , ktoré sú vstupom pre ďalšiu vrstvu extrakcie prvkov  $F_0$  využívajúcu preň operáciu konvolúcie. Výstup z časti SFENet je vstupom pre druhú a hlavnú časť siete RDN, a síce bloky RDB.

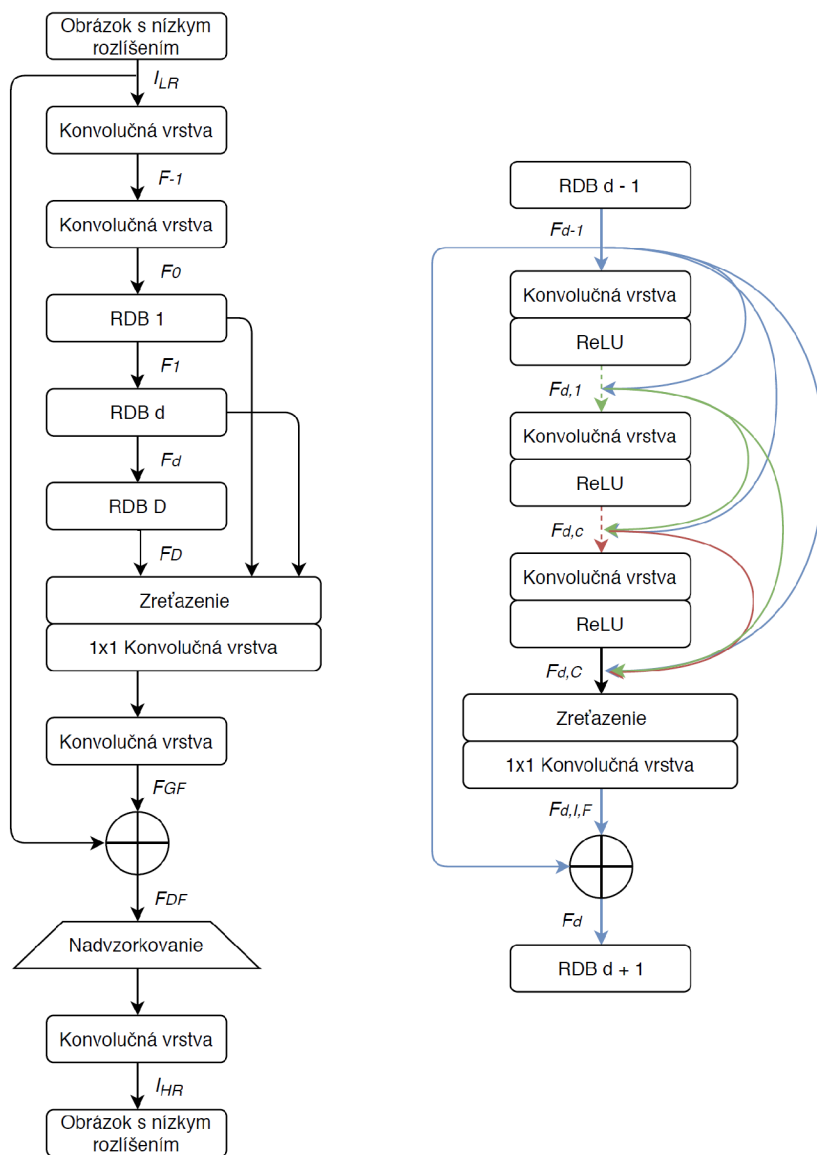
Existuje  $D$  za sebou nasledujúcich reziduálnych hustoprepojených blokov RDB, kde výstup  $F_d$  je výstupom  $d$ -tého RDB bloku a je vypočítaný ako:

$$F_d = H_{RDB,d}(F_{d-1}) = H_{RDB,d}(H_{RDB,d-1}(\cdots(H_{RDB,1}(F_0))\cdots)), \quad (3.9)$$

kde  $H_{RDB,d}$  označuje operáciu  $d$ -tého RDB bloku.  $H_{RDB,d}$  môže byť zložená funkcia pozostávajúca z konvolúcie a aktivačnej funkcie ReLU [25]. Nakoľko v rámci jednotlivých RDB blokov je plne využitá každá konvolučná vrstva, každý výstup  $F_d$  predstavuje tzv. lokálny znak. Na konci tejto časti sú znaky hierarchicky usporiadané a predané tretej časti nazývanej DFF (z angl. dense feature fusion). Táto časť zlučuje lokálne znaky  $[F_1; \cdots; F_D]$  do globálneho znaku  $F_{GF}$  nasledujúcim spôsobom:

$$F_{DF} = H_{DF}(F_{-1}; F_0; F_1; \cdots; F_D), \quad (3.10)$$

kde  $F_{DF}$  je výstup príznakovej mapy pozostávajúcej z lokálnych a globálnych príznakov vypočítaný využitím funkcie  $H_{DF}$ .



Obr. 3.8: Architektúra siete RDN (vľavo), štruktúra bloku RDB (vpravo)

Poslednou časťou je UPNet (angl. Up-sampling net), ktorá má na starosti nadvzorkovanie pôvodného priestoru s nízkym rozlíšením do priestoru so super-rozlíšením. Autori sa v návrhu časti UPNet inšpirovali obdobnou časťou z prác EDSR [5] a ESPCN [6]. Super-rozlíšený obrázok je potom definovaný nasledovne:

$$I_{SR} = H_{RDN}(I_{LR}), \quad (3.11)$$

kde  $H_{RDN}$  označuje funkciu predstavenej siete Residual Dense Network.

## 3.6.2 Popis implementácie

V tejto časti bude priblížená implementácia modelu z použitého repozitára <https://github.com/idealo/image-super-resolution> zverejneného na platforme GitHub. Zvolený projekt je z konca roku 2018, nesie názov Image Super-Resolution (ISR) a je napísaný v rozhraní Keras. Projekt je zaujímavý tým, že spája modely ESRGAN a Residual Dense Network a je užívateľsky nenáročný. Vývojár tak môže experimentovať a modifikovať základné bloky architektúry. Zdrojové kódy je možné stiahnuť a nainštalovať ako moduly prostredníctvom inštalátora Python balíkov *pip*. Pri tréovaní sa vychádzalo z doporučených nastavení siete modelu podľa autorov projektu. Ďalej bude popísaný tréovací proces v jednotlivých logických blokoch. Predpokladá sa nainštalovaný projekt v systéme.

Na začiatku je nutné nainportovať potrebné moduly:

```
from ISR.models import RRDN          # základný blok RRDN z ESRGAN
from ISR.models import Discriminator # Diskriminátor pre GAN
from ISR.models import Cut_VGG19     # Predtrénovaný model VGG19
from ISR.train import Trainer        # Tréovací modul
```

Model VGG19 bol natrénovaný na datasete ImageNet a je ďalej použitý ako príznakový extraktor pre chybovú funkciu vnímania (perceptual loss). Veľkosť obrázkov pri učení je určená parametrom *patch\_size*. Faktor rozlíšenia je v tomto prípade dvojnásobný.

```
lr_train_patch_size = 40    # Veľkosť výrezu LR obrázka
layers_to_extract = [5, 9] # Stanovenie výstupných vrstiev pre f_ext
scale = 2                  # Faktor rozlíšenia
hr_train_patch_size = lr_train_patch_size * scale
```

Nasleduje inicializácia siete generátora (RRDN), kde  $C$  je počet konvolučných vrstiev v každom RDB,  $D$  je počet blokov RDB,  $T$  je počet blokov RRDB,  $G$ ,  $G_0$  je počet príznakových máp (výstupných filtrov) v rámci bloku RDB, resp. za ním,  $f\_ext$  je príznakový extraktor umiestnený na konci siete.

```
rrdn = RRDN(arch_params={'C':4, 'D':3, 'G':64, 'G0':64,
                        'T':10, 'x':scale}, patch_size=lr_train_patch_size)
f_ext = Cut_VGG19(patch_size=hr_train_patch_size,
                  layers_to_extract=layers_to_extract)
discr = Discriminator(patch_size=hr_train_patch_size,
                     kernel_size=3)
```

Nastavenie váh pre jednotlivé komponenty chybovej funkcie *perceptual loss*:

```

loss_weights = {
    'generator': 0.0,
    'feature_extractor': 0.0833,
    'discriminator': 0.01,
}

```

V tejto chvíli sú všetky potrebné časti pripravené na inicializáciu hlavného objektu *Trainer*. Ten tieto časti skombinuje a začne proces tréovania s nastavenými parametrami. Inicializácia dodatočných parametrov procesu tréovania je nasledovná:

```

trainer = Trainer(
    generator=rrdn,
    discriminator=discr,
    feature_extractor=f_ext,
    lr_train_dir='FaceSet/train_HR/',
    hr_train_dir='FaceSet/train_LR_x2/',
    lr_valid_dir='FaceSet/valid_HR/',
    hr_valid_dir='FaceSet/valid_LR_x2/',
    loss_weights=loss_weights,
    dataname='FaceSet',
    logs_dir='./logs',
    weights_dir='./weights',
    weights_generator=None,
    weights_discriminator=None,
    n_validation=40,
    lr_decay_frequency=30,
    lr_decay_factor=0.5,
)

```

Nakoniec je nutné už len spustiť metódu *train* pripraveného tréovacieho modulu a nastaviť mu počet epoch, iterácií v rámci jednej epochy a veľkosť dávky:

```

trainer.train(
    epochs=100,
    steps_per_epoch=20,
    batch_size=16,
)

```

Model je uložený vo formáte HDF5, ktorý používa rozhranie Keras. Logovací súbor sa po procesoch tréovania a testovania nachádza v adresári *logs/*.

## 4 Výsledky práce

Výstupom tejto bakalárskej práce je porovnanie vybraných modelov neurónových sietí v rámci oblasti super-rozlíšenia. Neurónové siete boli natrénované na tréningových sadách popísaných v časti 3.2. Testovacia sada pozostáva zo 100 párov obrázkov (pár tvorí snímok s vysokým rozlíšením a jeho zmenšený variant pomocou bikubickej interpolácie). Použité obrázky sú výlučne tváre ľudí a to práve preto, že zvyšovanie rozlíšenia tváří patrí pod oblasť záujmu bezpečnostných zložiek, na čom si výstup tejto práce primárne zakladá. Tréningová sada bola pri tréningu rozdelená na tréningovú a validačnú sadu v pomere 80/20. Testovacia sada nevynecháva podmienku jej nezávislého obsahu od tréningovej sady.

Porovnanie modelov neurónových sietí je rozdelené do dvoch kategórií, objektívny a subjektívny pohľad na hodnotené modely neurónových sietí.

### 4.1 Metriky hodnotenia

Obsiahnuté objektívne metriky v tejto práci sú stredná kvadratická chyba MSE a s ňou úzko súvisiaci špičkový odstup signálu od šumu PSNR a structural similarity index SSIM. Tieto metriky sú v rámci merania kvality stratovej rekonštrukcie obrazu najrozšírenejšie.

#### 4.1.1 Stredná kvadratická chyba (MSE)

Stredná kvadratická chyba MSE je definovaná nasledujúcim vzťahom:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (I(i) - \hat{I}(i))^2, \quad (4.1)$$

v ktorom je vyjadrený originálny obrázok  $I$  a rekonštruovaný obrázok  $\hat{I}$ , oba o šírke  $N$  pixelov. MSE vyjadruje kvantitatívne skóre, ktoré popisuje stupeň skreslenia medzi dvoma signálmi a má jednotku decibel [dB]. Čím je hodnota MSE bližšia k nule, tým je väčšia vzájomná podobnosť dvoch signálov, čo znamená lepšiu kvalitu rekonštruovaného obrázka  $\hat{I}$ . [37]

#### 4.1.2 Špičkový odstup signálu od šumu (PSNR)

Špičkový odstup signálu od šumu PSNR (Peak signal-to-noise ratio) vyjadruje pomer medzi maximálnou možnou energiou signálu a energiou šumu. Definovaný je ako:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right), \quad (4.2)$$

a to pomocou maximálnej možnej honoty pixela  $L$  a strednej kvadratickej chyby MSE. PSNR má jednotku decibel [dB], býva vyjadrený logaritmickou stupnicou a jeho typický rozsah je od 20 do 40, kde vyššia hodnota udáva vyššiu kvalitu obrazu, resp. vyššiu vzájomnú podobnosť dvoch tenzorov medzi sebou. Nesie informáciu medzi zmenami hodnôt pixelov na rovnakej pozícii zadaného obrázka. Toto porovnanie kvality je však iba matematicky odvodené a nerešpektuje psychovizuálne vnímanie človeka. Ľudské vnímanie obrazu je komplexnejšie a PSNR ani MSE tak často nekoreluje so subjektívnym hodnotením. [38]

Túto metriku som na testovanie použil z dôvodu miery použitia v značnom množstve rešpektovaných prác a štúdií, ktoré sa venujú porovnávaniu kvality obrazových dát.

### 4.1.3 Structural similarity index (SSIM)

SSIM je metrika, ktorá je oproti UQI (angl. Universal quality index), jej predchodcovi (počítaný ako kombinácia straty korelácie, skreslenia jasu a kontrastu) [39], založená na výpočte troch charakteristík vizuálneho vnímania obrazu. Sú nimi *svietivosť* (angl. luminance), *kontrast* (angl. contrast) a *štruktúra* (angl. structure). Celkový index, ktorý je vypočítaný podľa nasledujúceho vzťahu je ich kombináciou:

$$\text{SSIM}(x, y) = [l(x, y)^\alpha] \cdot [c(x, y)^\beta] \cdot [s(x, y)^\gamma], \quad (4.3)$$

kde  $l(x, y)$  je svietivosť,  $c(x, y)$  je kontrast,  $s(x, y)$  je štruktúra v rámci obrázkov  $x, y$  a ktoré sú počítané ako:

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x, y) &= \frac{2\rho_x\rho_y + C_2}{\rho_x^2 + \rho_y^2 + C_2} \\ s(x, y) &= \frac{\rho_{xy} + C_3}{\rho_x\rho_y + C_3}, \end{aligned} \quad (4.4)$$

kde  $\mu_x, \mu_y, \rho_x, \rho_y, \rho_{xy}$  sú vážené priemery, vážené variancie, resp. kovariancia signálov  $x, y$  a  $C$  sú konštanty. SSIM nadobúda hodnoty v intervale  $[-1; 1]$ . Ak  $\alpha = \beta = \gamma = 1$  (predvolené hodnoty exponentov) a  $C_3 = C_2/2$  (predvolená hodnota  $C_3$ ) je možné SSIM zjednodušiť na: [40]

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\rho_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\rho_x^2 + \rho_y^2 + C_2)}. \quad (4.5)$$

#### 4.1.4 Subjektívne metriky

Ľudský zrak je komplexná biologická a psychologická záležitosť, ktorá stále ešte v súčasnosti nie je špecialistami úplne popísaná. V oblasti spracovania obrazu, videa či počítačového videnia sa používa jej zjednodušený model - ľudský zrakový systém *HVS* (angl. human visual system).

Subjektívne hodnotenie je proces porovnania obrázkov, pri ktorom človek hodnotí kvalitu predložených fotografií iba na základe videného.

## 4.2 Hodnotenie vybraných modelov

Vybrané modely prešli fázou testovania na všeobecnej a vlastnej testovacej sade obrázkov. V nej boli aplikované objektívne a subjektívne kvalitatívne metriky pre meranie kvality obrazu.

Hodnoty z testovania vybraných modelov natrénovaných na všeobecnom datasete *DIV2K* zobrazuje tabuľka 4.1.

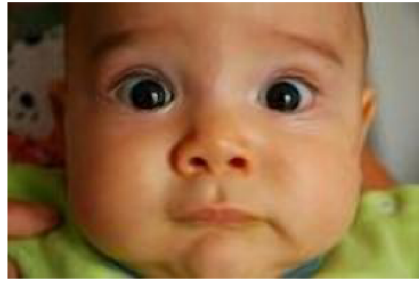
| Model     | MSE     | PSNR   | SSIM  |
|-----------|---------|--------|-------|
| Bikubická | 26.792  | 29.347 | 0.914 |
| SRGAN     | 112.872 | 27.351 | 0.822 |
| ESRGAN    | 32.771  | 30.165 | 0.859 |
| ISR       | 30.027  | 32.113 | 0.921 |
| EDSR      | 31.981  | 30.132 | 0.923 |

Tab. 4.1: Porovnanie výsledkov na testovacej sade *DIV2K*

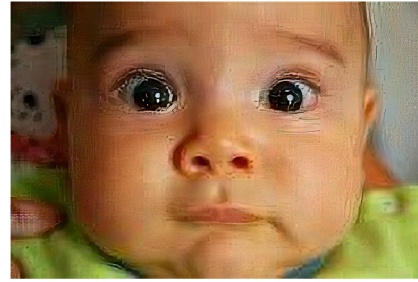
| Model  | MSE     | PSNR   | SSIM  |
|--------|---------|--------|-------|
| SRGAN  | 147.332 | 23.291 | 0.805 |
| ESRGAN | 37.429  | 27.561 | 0.900 |
| ISR    | 31.097  | 30.457 | 0.908 |
| EDSR   | 29.204  | 30.256 | 0.926 |

Tab. 4.2: Porovnanie výsledkov na vlastnej testovacej sade *FaceSet*

Ako je z tabuliek 4.1 a 4.2 vidieť, v testovaní vybraných metód v oblasti superrozlíšenia najlepšie obstál model hlbokaj reziduálnej siete EDSR a to v oboch variantách dátových sád. Druhý v poradí sa podľa objektívneho hodnotenia umiestnil model ISR a za ním ESRGAN. Metriky SSIM a PSNR sú medzi modelmi iba o málo horšie, aj keď v prípade vlastnej testovacej sady hodnoty MSE a PSNR dosiahol model ISR dokonca o niečo vyššie než u EDSR. Ako posledný skončil v oboch



a) Bikubická 29.347dB/0.914



b) SRGAN 23.291dB/0.805



c) ESRGAN 27.561dB/0.900



d) ISR 30.457dB/0.908



e) EDSR 30.256dB/0.926



f) Originál PSNR/SSIM

Obr. 4.1: Dvojnásobne zväčšené obrázky pomocou vybraných metód

hodnoteniach model SRGAN, ktorý sa ukázal ako celkom nevhodný, hlavne v obrázkoch s početnými vysokofrekvenčnými detailmi a dokazuje dokonca horšie výsledky než rekonštrukcia pomocou bikubickej interpolácie.

Subjektívne hodnotenie ukázalo, že modely EDSR, resp. ISR spomedzi testovaných modelov vedia rekonštruovať obrázky tváří ľudí najhodnovernejšie. Tretie miesto obsadil model ESRGAN, ktorý skončil pred posledným modelom SRGAN.



## 5 Záver

Cielom tejto práce bolo zoznámenie s problematikou super-rozlíšenia za pomoci umelých neurónových sietí. V prvej kapitole som priblížil základnú štruktúru neurónových sietí, od jeho základného prvku - neurónu. Pokračoval som uvedením teórie strojového učenia pomocou algoritmu spätnej propagácie, nakoľko je to kľúčový princíp v tejto oblasti. Ďalej bol v práci vysvetlený proces učenia pomocou konvolučných sietí ako aj jej jednotlivých základných prvkov. Teória o reziduálnych a husto prepojených sieťach s popisom ich stavebných blokov uzavrela túto kapitolu.

Nasledujúca, druhá kapitola pojednáva o super-rozlíšení a jej hlavných prínosoch. V nej som priblížil významné práce spojené s aktuálnymi výzvami v oblasti zvyšovania kvality obrazových dát za pomoci hlbokých konvolučných neurónových sietí.

Ďalej, v tretej kapitole, som spomenul použité nástroje a vybrané datasety. Jedným z datasetov bol DIV2K, čo je oficiálny dataset poskytnutý z výziev NTIRE 2017 a 2018. Ako druhý bol použitý aj vlastný dataset autora tejto práce (FaceSet). V tretej kapitole je rozvinutá architektúra zvolených modelov a proces testovania každého z nich. Použité modely sú SRGAN, ESRGAN, EDSR a ISR, ktoré patria v súčasnosti k jedným z najlepšie hodnotených. Príloha tejto práce obsahuje zdrojové kódy použitých modelov a niekoľko dátových sád.

V poslednej kapitole sú zhrnuté výsledky, ktoré si táto práca dala za úlohu. Obsahuje popis objektívnych kvalitatívnych metrík a ich následnú aplikáciu na testované modely. Modely z oblasti odstraňovania šumu som v práci nezamýšľal a sústredil sa iba na oblasť super-rozlíšenia. Modely EDSR a ISR, ktorý spája vlastnosti metód ESRGAN a RDN obstáli spomedzi testovaných modelov najlepšie ako v objektívnom, tak aj subjektívnom hodnotení.

## Literatúra

- [1] HERCULANO-HOUZEL, Suzana a Roberto LENT. Isotropic Fractionator: A Simple, Rapid Method for the Quantification of Total Cell and Neuron Numbers in the Brain. *Journal of Neuroscience* [online]. 2005, s. 2518-2521 [cit. 2018-12-14]. DOI: 10.1523/JNEUROSCI.4526-04.2005. Dostupné z URL: <<http://www.jneurosci.org/content/25/10/2518>>
- [2] BULLOCK, Theodore H., Michael V. L. BENNET, Daniel JOHNSTON a kol. The Neuron Doctrine, Redux. *Science*. 2005, s. 791-793. DOI: 10.1126/science.1114394.
- [3] HAOHAN, Wang, Raj BHIKSHA a Eric P. XING. On the Origin of Deep Learning. *CoRR - Computing Research Repository* [online]. 2017 [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1702.07800>>
- [4] WANG, Xintao, Ke YU, Shixiang WU a kol. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2019-05-21]. Dostupné z URL: <<http://arxiv.org/abs/1809.00219>>
- [5] LIM, Bee, Sanghyun SON, Heewon KIM, a kol. Enhanced deep residual networks for single image super-resolution *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* [online]. 2017, s. 1132-1140 [cit. 2018-12-13]. DOI: 10.1109/CVPRW.2017.151. ISBN: 978-1-5386-0733-6. Dostupné z URL: <<https://ieeexplore.ieee.org/document/8014885>>
- [6] SHI, Wenzhe, Jose CABALLERO, Ferenc HUSZÁR a kol. Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. 2016, s. 1874-1883 [cit. 2019-05-16]. DOI: 10.1109/CVPR.2016.207. ISBN 978-1-4673-8851-1. Dostupné z URL: <<http://ieeexplore.ieee.org/document/7780576>>
- [7] SRIVASTAVA, Rupesh Kumar, Klaus GREFF a Jurgen SCHMIDHUBER. Training Very Deep Networks *The Swiss AI Lab IDSIA / USI / SUPSI* [online]. 2015, [cit. 2019-05-17]. Dostupné z URL: <<https://arxiv.org/abs/1507.06228>>
- [8] KAIMING, He, Zhang XIANGYU, Ren SHAOQING a kol. Deep Residual Learning for Image Recognition *2016 IEEE Conference on Computer Vision and*

- Pattern Recognition (CVPR)* [online]. 2016, s. 770-778 [cit. 2018-12-14]. DOI: 10.1109/CVPR.2016.90. Dostupné z URL: <<https://ieeexplore.ieee.org/document/7780459/citations#citations>>
- [9] HRINČÁR, Matej. *Konvoluční neuronové sítě a jejich využití při detekci objektů* [online]. Praha, 2013. Diplomová práce (Mgr.). Univerzita Karlova, Matematicko-fyzikální fakulta, Katedra teoretické informatiky a matematické logiky, 2013-01-22.
- [10] KRIZHEVSKY, Alex, Ilya SUTSKEVER a Geoffrey E. HINTON. Image-net classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* [online]. 2012, [cit. 2018-12-14]. Dostupné z URL: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.299.205>>
- [11] ALBAWI, Saad, Tareq Abed MOHAMMED a Saad Al-Zawi. Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)* [online]. 2017, s. 1-6 [cit. 2018-12-14]. DOI: 10.1109/ICEngTechnol.2017.8308186. Dostupné z URL: <<https://ieeexplore.ieee.org/document/8308186>>
- [12] LIU, Po-Yu a Edmund Y. LAM. Image Reconstruction Using Deep Learning. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2019-05-24]. Dostupné z URL: <<https://arxiv.org/abs/1809.10410>>
- [13] YU, Jiahui, Yuchen FAN, Jianchao YANG, a kol. Wide Activation for Efficient and Accurate Image Super-Resolution. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1808.08718>>
- [14] ZHANG, Yulun, Yapeng TIAN, Yu KONG a kol. Residual Dense Network for Image Super-Resolution. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2018-12-14]. DOI: 10.1109/CVPR.2017.298. Dostupné z URL: <<https://arxiv.org/abs/1802.08797>>
- [15] TAI, Ying, Jian YANG a Xiaoming LIU. Image Super-Resolution via Deep Recursive Residual Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. 2017, [cit. 2018-12-14]. Dostupné z URL: <<https://ieeexplore.ieee.org/document/8099781>>
- [16] LEDIG, Christian, Lucas THEIS, Ferenc HUSZÁR a kol. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE*

- Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. 2016, s. 105-114 [cit. 2018-05-06]. Dostupné z URL: <<https://arxiv.org/abs/1609.04802>>
- [17] GOODFELLOW, Ian J., Jean POUGET-ABADIE, Mehdi MIRZA a kol. Generative Adversarial Networks. *Department of Computer Science and Operations Research, University of Montreal* [online]. 2014, s. 105-114 [cit. 2019-05-06]. Dostupné z URL: <<https://arxiv.org/abs/1406.2661>>
- [18] CHEN, Yu, Ying TAI, Xiaoming LIU a kol. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. *CoRR - Computing Research Repository* [online]. 2017, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1711.10703>>
- [19] ISOLA, Phillip, Jun-Yan ZHU, Tinghui ZHOU a kol. Image-to-Image Translation with Conditional Adversarial Networks. *CoRR - Computing Research Repository* [online]. 2016, [cit. 2019-05-25]. Dostupné z URL: <<https://arxiv.org/abs/1611.07004>>
- [20] ČIŠECKÝ, Roman. *Metody pro odstranění šumu z digitálních obrazů* [online]. Brno, 2012. [cit. 2019-05-25]. Dostupné z URL: <<http://hdl.handle.net/11012/9097>>. Diplomová práce. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací. Vedoucí práce Petr Číka.
- [21] SIMONYAN, Karen a Andrew ZISSERMAN. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR - Computing Research Repository* [online]. 2014, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1409.1556>>
- [22] KAIMING, He, Zhang XIANGYU, Ren SHAOQING a kol. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification *The IEEE International Conference on Computer Vision (ICCV)* [online]. 2015, s. 1026-1034 [cit. 2018-12-14]. Dostupné z URL: <[https://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/html/He\\_Delving\\_Deep\\_into\\_ICCV\\_2015\\_paper.html](https://www.cv-foundation.org/openaccess/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_paper.html)>
- [23] GLOROT, Xavier a Yoshua BENGIO. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* [online]. 2010, s. 249-256 [cit. 2018-12-14]. Dostupné z URL: <<http://proceedings.mlr.press/v9/lorot10a.html>>

- [24] JOLICOEUR-MARTINEAU, Alexia. The relativistic discriminator: a key element missing from standard GAN. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2019-05-21]. Dostupné z URL: <<https://arxiv.org/abs/1807.00734>>
- [25] GLOT, Xavier, Antoine BORDES a Yoshua BENGIO. Deep sparse rectifier neural networks. *International Conference on Artificial Intelligence and Statistics* [online]. 2011, [cit. 2019-05-16]. Dostupné z URL: <<http://proceedings.mlr.press/v15/glot11a/glot11a.pdf>>
- [26] LEHTINEN, Jaakko, Jacob MUNKBERG, Jon HASSELGREN a kol. Noise2Noise: Learning Image Restoration without Clean Data *CoRR - Computing Research Repository* [online]. 2018, [cit. 2019-05-24]. Dostupné z URL: <<https://arxiv.org/abs/1803.04189.pdf>>
- [27] P. KINGMA, Diederik a Jimmy BA. Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations* [online]. 2015, [cit. 2019-05-21]. Dostupné z URL: <<https://arxiv.org/abs/1412.6980>>
- [28] YANG, Wenming, Xuechen ZHANG, Yapeng TIAN a kol. Deep Learning for Single Image Super-Resolution: A Brief Review. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1808.03344>>
- [29] SZEGEDY, Christian, Sergey IOFFE, Vincent VANHOUCKE a kol. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *CoRR - Computing Research Repository* [online]. 2016, [cit. 2019-05-21]. Dostupné z URL: <<https://arxiv.org/abs/1602.07261>>
- [30] DONG, Chao, Chen Change LOY, Kaiming HE a kol. Image Super-Resolution Using Deep Convolutional Networks *CoRR - Computing Research Repository* [online]. 2015, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1501.00092>>
- [31] ImageNet [online]. [cit. 2018-12-14]. Dostupné z URL: <<http://www.image-net.org/>>
- [32] ZHANG, Kai, Wangmeng ZUO, Yunjin CHEN a kol. Beyond a Gaussian Denoiser: Residual Learning of Deep {CNN} for Image Denoising. *CoRR - Computing Research Repository* [online]. 2016, [cit. 2018-05-25]. Dostupné z URL: <<https://arxiv.org/abs/1608.03981>>

- [33] ZHANG, Kai, Wangmeng ZUO a Lei ZHANG. FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising. *CoRR - Computing Research Repository* [online]. 2017, [cit. 2018-05-25]. Dostupné z URL: <<https://arxiv.org/abs/1710.04026>>
- [34] ULYANOV, Dmitry, Andrea VEDALDI a Victor S. LEMPITSKY. Deep Image Prior. *CoRR - Computing Research Repository* [online]. 2017, [cit. 2018-05-25]. Dostupné z URL: <<https://arxiv.org/abs/1711.10925>>
- [35] MAHPOD, Shahar, Rig DAS, Emanuele MAIORANA a kol. Facial Landmark Point Localization using Coarse-to-Fine Deep Recurrent Neural Network. *CoRR - Computing Research Repository* [online]. 2018, [cit. 2018-12-14]. Dostupné z URL: <<https://arxiv.org/abs/1805.01760>>
- [36] TensorFlow [online]. [cit. 2018-12-14]. Dostupné z URL: <[https://www.tensorflow.org/api\\_guides/python/nm](https://www.tensorflow.org/api_guides/python/nm)>
- [37] ŠKURLA, Adam. *Srovnání metod měřících kvalitu obrazů*. Brno, 2018. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. David Bařina, Ph.D.
- [38] WANG, Zhihao, Jian CHEN, Steven C.H. HOI a kol. Deep Learning for Image Super-resolution: A Survey. *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online] 2019, [cit. 2019-04-25]. Dostupné z URL: <<https://arxiv.org/abs/1902.06068>>
- [39] WANG, Zhou a Alan BOVIK. A Universal Image Quality Index. *Signal Processing Letters IEEE* [online]. 2002, s. 81-84. DOI: 10.1109/97.995823. Dostupné z URL: <[https://www.researchgate.net/publication/3342733\\_A\\_Universal\\_Image\\_Quality\\_Index](https://www.researchgate.net/publication/3342733_A_Universal_Image_Quality_Index)>
- [40] WANG, Zhihao, Eero P. SIMONCELLI a Alan C. BOVIK. Multiscale structural similarity for image quality assessment. *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers* [online]. 2004, s. 1398–1402. DOI: 10.1109/ACSSC.2003.1292216. ISBN 978-0-7803-8104-9. Dostupné z URL: <<http://ieeexplore.ieee.org/document/1292216>>

# Zoznam symbolov, veličín a skratiek

|                 |  |
|-----------------|--|
| <b>SR</b>       | Super-rozlíšenie – Super-Resolution                                  |
| <b>LR</b>       | Nízke rozlíšenie – Low resolution                                    |
| <b>HR</b>       | Vysoké rozlíšenie – High resolution                                  |
| <b>MSE</b>      | Stredná kvadratická chyba – Mean Squared Error                       |
| <b>MAE</b>      | Stredná absolútna chyba – Mean Absolute Error                        |
| <b>PSNR</b>     | Špičkový odstup signálu od šumu – Peak signal-to-noise ratio         |
| <b>SSIM</b>     | Structural Similarity Index  |
| <b>CNN</b>      | Konvolučná neurónová sieť – Convolutional Neural Network             |
| <b>ReLU</b>     | Rectifier Linear Unit  |
| <b>SISR</b>     | Single Image Super-Resolution  |
| <b>DCNN</b>     | Hlboká konvolučná neurónová sieť – Deep Convolutional Neural Network |
| <b>CUDA</b>     | Compute Unified Device Architecture                                  |
| <b>RAM</b>      | Random Access Memory   |
| <b>JPEG</b>     | Joint Photographic Expert Group                                      |
| <b>RDB</b>      | Residual dense block   |
| <b>RRDB</b>     | Residual-in-Residual dense block                                     |
| <b>DFF</b>      | Dense Feature Fusion   |
| <b>UPNet</b>    | Up-sampling net  |
| <b>HVS</b>      | Human Visual System  |
| <b>HDF</b>      | Hierarchical Data Format   |
| <b>SRCNN</b>    | Super-Resolution Convolutional Neural Network                        |
| <b>ResNet</b>   | Residual network   |
| <b>DenseNet</b> | Dense network  |
| <b>GAN</b>      | Generative Adversarial Network                                       |
| <b>SRGAN</b>    | Super-Resolution using a Generative Adversarial Network              |
| <b>SRResNet</b> | Super-Resolution using a Residual Network                            |
| <b>VDSR</b>     | Very Deep Convolutional Networks                                     |
| <b>FSRNet</b>   | Face Super-Resolution Network  |
| <b>FSRGAN</b>   | Face Super-Resolution using a Generative Adversarial Network         |
| <b>FFDNet</b>   | Toward a Fast and Flexible Solution for CNN based Image Denoising    |
| <b>DIV2K</b>    | DIVERse 2K   |
| <b>EDSR</b>     | Enhanced Deep Residual Networks for Single Image Super-Resolution    |
| <b>MDSR</b>     | Multi-scale Deep Super-Resolution system                             |
| <b>WDSR</b>     | Wide Activation for Efficient and Accurate Image Super-Resolution    |
| <b>ESRGAN</b>   | Enhanced Super-Resolution using a Generative Adversarial Network     |

|               |  |
|---------------|--|
| <b>VGG</b>    | Visual Geometry Group                                    |
| <b>SFTGAN</b> | Spatial Feature Transform Generative Adversarial Network |
| <b>RDN</b>    | Residual Dense Network                                   |
| <b>RRDN</b>   | Residual-in-Residual Dense Network                       |
| <b>ESPCN</b>  | Efficient Sub-Pixel Convolutional Neural Network         |
| <b>ISR</b>    | Image Super Resolution                                   |
| <b>NTIRE</b>  | New Trends in Image Restoration and Enhancement          |
| <b>PIRM</b>   | Perceptual Image Restoration and Manipulation            |