VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2023 Martin Rosa

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

SUPERROZLIŠENÍ V OBRAZE PRO ZAJIŠTĚNÍ VYLEPŠENÉHO MONITOROVÁNÍ ZABEZPEČENÝCH PROSTORŮ

SUPER RESOLUTION IN THE IMAGE TO ENSURE IMPROVED MONITORING OF SECURED AREAS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE — Martin Rosa

AUTHOR

VEDOUCÍ PRÁCE — doc. Ing. Radim Burget, Ph.D.

SUPERVISOR

BRNO 2023

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Bakalářská práce

bakalářský studijní program **Informační bezpečnost**

Ústav telekomunikací

Student: Martin Rosa ***ID:*** 230648

Ročník: 3 ***Akademický rok:*** 2022/23

NÁZEV TÉMATU:

Superrozlišení v obraze pro zajištění vylepšeného monitorování zabezpečených prostorů

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou super-rozlišení, kde se zaměřte nikoli jen na využití ke zvýšení rozlišení fotografie, ale na využití v oblasti biometrie. Proveďte podrobnou rešerši stavu vědy a techniky v této oblasti, zmapujte existující datové množiny pro tyto účely a vyberte vhodné metriky pro hodnocení kvality. S použitím vybraných architektur neuronových sítí natrénujte několik modelů, které poté s pomocí vybraných metrik vzájemně srovnejte v tabulce. Na vybraných příkladech ukažte případy, kdy metody selhávají a kdy naopak fungují.

DOPORUČENÁ LITERATURA:

podle pokynů vedoucího práce

Termín zadání: 6.2.2023 ***Termín odevzdání:*** 26.5.2023

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

doc. Ing. Jan Hajný, Ph.D.
předseda rady studijního programu

ABSTRAKT

Cieľom bakalárskej práce bolo porovnať modely super-rozlíšenia s aplikáciou na reštaurovanie fotiek ľudských tvárí. V práci sme spracovali rešerš technológií superrozlíšenia a následne sme natrénovali a porovnali 5 modelov. Zameriavame sa hlavne na oblasť superrozlíšenia, ktorá by mohla byť nápomocná na identifikáciu osôb z bezpečnostných kamier. Použité technológie boli preto vyberané na základe percepčnej kvality a schopnosti identifikácie osoby na výstupnom snímku. Práca ukázala účinnosť porovnaných modelov pomocou objektívnych aj subjektívnych metrík. Výsledky boli porovnané v dotazníku (106 respondentov). Dotazník ukázal účinnosť použitia vlnovej transformácie v superrozlíšení tvarí.

KĽÚČOVÉ SLOVÁ

superrozlíšenie, halucinácia tvárí, neurálne siete, hlboké učenie, bezpečnostné kamery, identifikácia

ABSTRACT

The point of this bachelor thesis was to compare models of super-resolution with the application on resolving human faces. A brief review of the technologies of super-resolution was created and five models were trained and compared. The focus was on the area of super-resolution that could be helpful with identifying people from CCTV cameras. Used technologies were therefore chosen based on their perceptual quality and ability to identify the person in the output image. This thesis has shown the effectivity of the compared models using objective and subjective metrics. The results were compared in a survey (106 respondents). Survey has shown the advantage of using wavelet-transform in the area of the super-resolution of human faces.

KEYWORDS

super-resolution, face hallucination, neural networks, deep learning, CCTV cameras, identification

ROSA, Martin. *Superrozlišení v obraze pro zajištění vylepšeného monitorování zabezpečených prostorů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2023, 51 s. Bakalárska práca. Vedúci práce: doc. Ing. Radim Burget, Ph.D.

Vyhlásenie autora o pôvodnosti diela

Meno a priezvisko autora:	Martin Rosa
VUT ID autora:	230648
Typ práce:	Bakalárska práca
Akademický rok:	2022/23
Téma závěrečnej práce:	Superrozlišení v obraze pro zajištění vylepšeného monitorování zabezpečených prostorů

Vyhlasujem, že svoju záverečnú prácu som vypracoval samostatne pod vedením vedúcej/cého záverečnej práce, s využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej záverečnej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto záverečnej 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 trestnoprávnych dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákonníka Českej republiky č. 40/2009 Sb.

Brno ....

podpis autora*

*Autor podpisuje iba v tlačenej verzii.

POĎAKOVANIE

Rád by som sa poďakoval vedúcemu bakalárskej práce pánovi doc. Ing. Radimu Burgetovi, Ph.D. za odborné vedenie, konzultacie, trpezlivosť a podnetné návrhy k práci.

Obsah

Zoznam obrázkov

Úvod

Pod pojmom superrozlíšenie sa rozumie, zväčšovanie rozlíšenia obrázku, tak aby došlo k zvýšeniu jeho kvality. Super-rozlíšenie je téma zo širokou škálou aplikácie (bezpečnosť, zdravotníctvo, atď.). V našej práci sa budeme zaoberať použitím týchto technológií na bezpečnosť. Konkrétne riešime aplikáciu super-rozlíšenia na vylepšovanie záberov z bezpečnostných CCTV kamier pre účely identifikácie osôb na záberoch z týchto kamier.

Tento problém je komplexný a existuje veľa metód, ako sa ho pokúsiť vyriešiť. Jeho riešenie je však podstatné pre jeho významnú aplikáciu pre bezpečnostné zložky, ktoré pravidelne pracujú s zábermi na základe ktorých sú identifikovaný podozrivý z trestných činov. Zvýšená kvalita týchto záznamov (ktoré zvyknú často trpieť veľmi zlou kvalitou) môže identifikáciu uľahčiť, príp. ju umožniť. V tejto práci porovnávame moderné technológie super-rozlíšenia pri ich aplikácii na rekonštrukciu obrazu ľudských tvárí. Výsledky hodnotíme pomocou objektívnych metrík a subjektívnych metrík, za účelom zistenia aplikovateľnosti týchto technológií na daný problém.

V úvode práce je stručný popis problému super rozlíšenia a jeho aplikácia na tváre. Následne sú popísané technológie zvyšovania kvality obrazu. V ďalších kapitolách sú popísané neurónové siete a ich aplikácia na danú problematiku. Zameriavame sa tu aj na unikátne problémy, ktoré sa objavujú v prípade použitia superrozlíšenia za účelom identifikácie. Na záver uvádzame technológie použité pri experimentácii.

Hlavným prínosom práce je porovnanie 5 technologií superrozlíšenia objektívnymi aj subjektívnymi metrikami. Zároveň sme upravili existujúcu sieť za účelom dosiahnutia lepších výsledkov. Nami upravenej sieti sa podarilo dosiahnuť výsledky porovnateľné s modernými technológiami, v subjektívnom hodnotení skončila na druhom mieste. V subjektívnom hodnotení sme ukázali účinnosť vlnovej transformácie pri superrozlíšení ľudských tvárí.

Dielo končí časťou kde uvádzame výsledky práce. Zaoberáme sa tu spôsobmi ich hodnotenia a procesom trénovania sietí. V závere zhrnieme subjektívne aj objektívne výsledky.

1 Superrozlíšenie

Pod pojmom superrozlíšenie (super-resolution) sa rozumie zvýšenie kvality snímky, prípadne videa. Superrozlíšenie má veľa oblastí (superrozlíšenie tvárí, satelitných snímkov, medicínskych snímkov a pod.) v tejto práci sa zameriavame na superrozlíšenie tvárí. Kľúčovou časťou superrozlíšenia je tzv. prevzorkovanie (upsampling).

1.1 Prevzorkovanie

Superrozlíšenie je založené na zmene snímky s nízkym rozlíšením (LR) na snímku z vysokým rozlíšením (HR). V skutočnosti ide o od základu zle založený problém, keďže z viacerých HR snímok je možné spraviť rovnakú LR snímku. Toto znemožňuje získanie pôvodnej HR snímky, výsledkom superrozlíšenia je len aproximácia HR snímky.[46] Pri získavaní HR snímky musíme zvýšiť rozlíšenie pôvodnej LR snímky. Matematicky možno problém superrozlíšenia zapísať ako:

$$\widehat{HR} = F(LR, \theta)\,, \tag{1.1}$$

kde LR reprezentuje LR snímku, F reprezentuje zvolenú metódu superrozlíšenia a θ reprezentuje parametre funkcie. $\widehat{HR}$ reprezentuje aproximáciu pôvodného HR obrázku. Parametre funkcie závisia od použitej technológie. Pri superrozlíšení založenom na neurálnych sieťach sa získavajú iteratívnym optimalizačným procesom nazývaným učenie. [46]

Metódy superrozlíšenia sú založené na prevzorkovaní. Jedná sa o pridávanie pixelov do obrázku na základe daných pravidiel. V nasledujúcich odsekoch sa zameriame na rôznorodé metódy prevzorkovania.

1.1.1 Interpolačné metódy

Jedná sa o najznámejšie a najjednoduchšie metódy prevzorkovania. Patrí sem interpolácia najbližších susedov (nearest neighbor interpolation), bilineárna interpolácia (bilinear interpolation), bikubická interpolácia (bicubic interpolation) a iné.

Interpolácia najbližších susedov

Jednoduchý algoritmus, založený na princípe najbližších susedov. Pridané pixely si hodnoty preberú od pôvodných pixelov, ktoré sú k nim najbližšie. Jedná sa o veľmi rýchlu a výpočetne nenáročnú metódu. Metóda však produkuje výsledky s veľmi nízkou kvalitou. [46]

Bilineárna interpolácia (BIL)

Štatistická metóda. Doplnené pixely si hodnoty získavajú vypočítaním "priemerných" hodnôt pôvodných pixelov. Jednoduchá, rýchla a nenáročná metóda. Dosahuje lepšej kvality, ako interpolácia najbližších susedov. [46]

Bikubická interpolácia (BIC)

Pracuje na rovnakom princípe ako BIL. Pridáva komplexnosť, keďže, narozdiel od BIL, vypočítava hodnoty pridaným pixelom pomocou polynomickej metódy. "Priemerné" hodnoty sa počítajú na krivke, čo dosahuje lepšie výsledky, ako BIL. Jedná sa o najpopulárnejšiu metódu interpolácie. [46]

1.1.2 Prevzorkovanie založený na učení

Vyššie popísané metódy používajú iba informácie získané zo snímky na zvýšenie rozlíšenia. Momentálne však dosahuje lepších výsledkov prevzorkovanie založené na učení, ktoré používa aj informácie naučené použitou neurálnou sieťou.[6]

Konvolučné neurálne siete

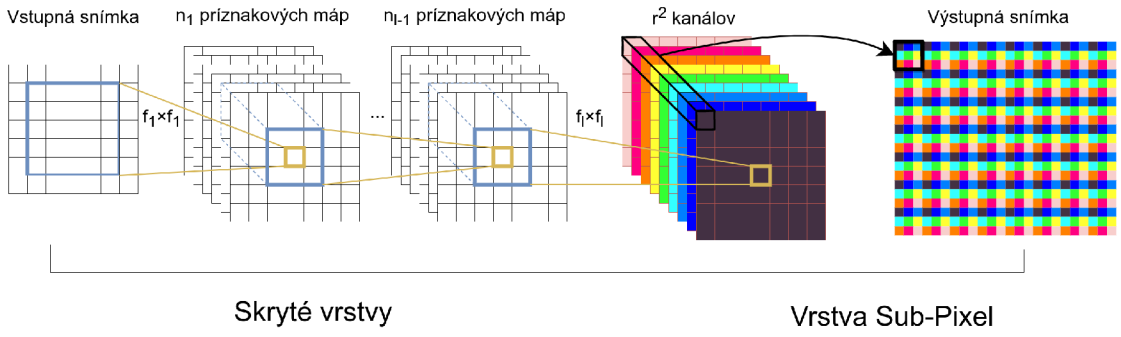
Konvolučné neurálne siete priniesli veľký pokrok v oblasti superrozlíšenia.[10] [6] Táto technológia je založená na tzv. konvolúciách. Konvolúcie sú proces, kde po snímke prechádzame kernelom s danou veľkosťou a posunom (stride). Kernel (niekedy nesprávne nazývaný aj filter) sa pohybuje po snímke, pričom rýchlosť posunu je určená posunom, a sčítava hodnoty, ktoré zachytáva. Výsledkom je tzv. Skonvolovaná vlastnosť (convoluted feature). Jedná sa o snímku so zníženým rozlíšením, ktorá sa následne posiela do ďalších vrstiev siete. Tento proces umožňuje sieti zachytiť podstatné vlastnosti snímky. Čo sa využíva hlavne v problémoch počítačového videnia (computer vision).[10]

Transponovaná konvolúcia

Tento mechanizmus funguje na opačnom princípe, preto je často nesprávne nazývaný dekonvolúcia. Pri transponovanej konvolúcii najprv snímku rozšírime na požadovanú veľkosť, pričom pridané pixely doplníme nulovými hodnotami. Následne do snímky opäť doplníme pixely s nulovými hodnotami, tak aby sme dosiahli dvojnásobok požadovanej veľkosti. Na záver, po snímke necháme prejsť kernel, ktorý snímku skonvoluje na požadovanú veľkosť. Týmto spôsobom dokáže sieť na základe príznakových máp (feature maps) predpovedať, ako by mohla *HR* snímka vyzerať. Táto metóda sa, vďaka svojím uspokojivým výsledkom používa v mnohých technológiách. [46] [10]

Vrstva Sub-pixel (Sub-pixel layer)

Vrstva Sub-pixel vytvára viaceré kanály pomocou viacnásobnej konvolúcie. Konkrétne sa pôvodný obrázok konvoluje, až pokiaľ nevznikne r2 kanálov, kde r je faktor zväčšenia. Pri vstupe s rozmermi $W \times H \times C$ (kde W je šírka snímky, H je výška snímky a C je počet farebných kanálov) bude teda výstupom $W \times H \times C * r^2$. Následne je na výstup aplikovaná tzv. zmiešavacia (shuffle) operácia, kde dôjde k prevodu $W \times H \times C * r^2$ na $rW \times rH \times C$. Týmto spôsobom dôjde ku zvýšeniu rozlíšenia o faktor r. Túto metódu môžeme vidieť bližšie na 1.1. Veľkou výhodou tejto operácie je, že pri konvoluci a následnom upsamplingu nie je nutné pridávať nuly, ani obrázok nejakým spôsobom umelo rozširovať. Vďaka tomuto si dokáže sieť získať viac kontextuálnych informácií, čo umožňuje lepšie generovať detaily. Zároveň, podobne ako pri Transpnovanej konvolucii, je operácia pomerne výpočetne jednoduchá. Prvým príkladom siete, ktorá túto metódu používala je ESPCN, ktorá dosiahla lepšie výsledky ako vtedy dovtedy existujúce technológie a zároveň to dokázala v lepšom výpočtovom čase. [40]



Obr. 1.1: Vrstva Sub-Pixel. Zdroj [40] (prekreslené)

1.2 Single-image vs. multi-image

Superrozlíšenie môžeme rozdeliť na dve kategórie

- Superrozlíšenie z jedného snímku (Single-image super-resolution - **SISR**)
- Superrozlíšenie z viacerých snímkov (Multi-image super-resolution - **MISR**)

Ako z názvu vyplýva, superrozlíšenie z jedného snímku vychádza z použitia jedného snímku ako základ, ktorému je následne zväčšené rozlíšenie. Zatiaľ čo pri superrozlíšení z viacerých snímkov vychádzame z viacerých snímkov, ktoré použijeme ako základ, pričom si z nich berieme informácie, podľa ktorých vygenerujeme *HR* snímku.

Nešťastnou realitou pre MISR je, že väčšina literatúry sa venuje práve SISR. Nejedná sa o dobre preskúmanú oblasť. MISR sa dostáva pozornosti iba v superrozlíšení satelitných snímkov [15] [38]. V poslednej dobe sa však do pozornosti dostáva aj MISR tvárí [37].

Rok 2019 priniesol technológiu DeepSUM. Model sa skladá zo siete ne extrakciu čŕt z snímok. Na základe týchto čŕt potom ďalšia sieť vytvára tzv. registračné filtre, na základe ktorých sa registrujú príznakové mapy. V poslednej časti siete dochádza k splývaniu príznakových máp a generovaniu $\widehat{HR}$ snímky. [33]

Technológia RAMS priniesla využitie reziduálneho príznakového bloku. Tento blok umožňuje modelu zamerať sa na vysokofrekvenčné komponenty snímky. Model dosiahol podobných výsledkov ako DeepSUM. [39]

Jedna z mála technológií superrozlíšenia tvarí bola prestavená v [37]. Autori svoju sieť založili na U-Net sieti (podobne ako ostatné technológie MISR). Následne bol pridaný tzv. GEU blok (gated enhanced unit). Narozdiel od väčšiny technológií, bol tento model ohodnotený nie len pomocou objektívnych metrík ale aj na základe dotazníku, v ktorom respondenti porovnávali výstupy rôznych sietí s pôvodnou HR snímkou a následne hlasovali za najpodobnejší. V danom dotazníku sa ukázala ako najlepšia práve táto technológia. Z čoho môže vyplývať, že sieť dokáže dobre rekonštruovať identifikačné črty.

StatNet je model, ktorý priniesol tzv. štatistický blok. Tento blok dokáže pretvoriť LR snímky na 6 príznakových máp, ktoré sa používajú vdalších častiach modelu. Výhodou štatistického bloku je, že neobsahuje trénovateľné časti, čiže je výpočtovo efektívny. Zároveň je možné do tohto bloku dať hocikoľko vstupných LR snímok. Čo je unikátna vlastnosť v oblasti MISR. [34]

Jednou z najnovších MISR technológií je QA-Net z roku 2022. Tento model používa tzv. mapy kvality (quality maps). Ich úlohou je odfiltrovať časti snímok, ktorých informačná hodnota bola znížená napr. šumom. Model vďaka tomu dokáže diskriminovať voči týmto častiam a získavať väčšinu komponentov z kvalitnejších častí snímok. [23]

Je nutné zmieniť, že QA-Net bol vytvoerný pre účely superrozlíšenia satelitných snímok. V tomto prípade dáva zmysel zamerať sa len na kvalitné časti snímky na získanie lepších informácií, z ktorých je možné generovať lepšie $\widehat{HR}$ snímky. To sa aj empiricky potvrdilo. [23] Je pre nás náročné si predstaviť využitie tohto fenoménu na superrozlíšenie tvárí, a teda aj jeho aplikovateľnosť na túto prácu. Keďže mapy kvality by mohli diskriminovať voči častiam obsahujúcim dôležité identifikačné črty.

2 Aplikácia neurálnych sietí na superrozlíšenie

Neurálne siete sa ukázali ako veľmi účinná metóda riešenie problematiky superrozlíšenia. Dnes sa v podstate stretávame už len z ich aplikáciou na tento problém. V nasledujúcich kapitolách sa budeme venovať prístupom k superrozlíšeniu za pomoci neurálnych sietí.

2.1 Stratové funkcie

Pod pojmom stratová funkcia (loss function), niekedy tiež prekladaná ako chybová funkcia, rozumieme funkciu, ktorá si za vstup berie $\widehat{HR}$ a HR a jej výstupom je odchýlka $\widehat{HR}$ od HR. Jedná sa o nutnú súčasť všetkých technológií super-rozlíšenia založených na neurálnych sieťach. Stratová funkcia vedie neurálnu sieť na základe odchýlky predošlých výsledkov k lepším výsledkom, pričom úlohou siete je minimalizovať túto odchýlku. V nasledujúcej časti sa pozrieme na niektoré stratové funkcie. Bude tu použitá rovnaká notácia ako v 1.1. [46]

Funkcie pixlovej straty (Pixel-wise loss)

Funkcie pixlovej straty vypočítajú odchýlku na každom pixely. Skladajú sa z $\mathcal{L}_1$ ktorá vypočíta priemernú absolútnu chybu (mean absolute error) a $\mathcal{L}_2$, ktorá vypočíta priemernú kvadratickú chybu (mean squared error)

$$\mathcal{L}_1(\widehat{HR}, HR) = \frac{1}{hwc} \sum_{i,j,k} |\widehat{HR}_{i,j,k} - HR_{i,j,k}| \ , \tag{2.1}$$

$$\mathcal{L}_2(\widehat{HR}, HR) = \frac{1}{hwc} \sum_{i,j,k} (\widehat{HR}_{i,j,k} - HR_{i,j,k})^2 \ , \tag{2.2}$$

kde h, w, c sú výška, šírka a počet kanálov vstupných snímok nasledovne. [46]

Nevýhodou $\mathcal{L}_2$ je, že dobre nezachytáva malé rozdiely medzi snímkami. Výsledky sietí používajúce túto funkciu zvyknú byť preto príliš hladké bez zachovania menších detailov. V tomto má $\mathcal{L}_1$ značnú výhodu oproti $\mathcal{L}_2$, keďže malé aj veľké rozdiely zachytáva podobne. [13] Ďalšou výhodou $\mathcal{L}_1$ je, že je inverzne korelované k jednej z najpoužívanejších hodnotiacich metrík – PSNR (peak signal-to-noise ratio). To znamená, že znižovaním $\mathcal{L}_1$ sa bude zároveň zvyšovať PSNR. [46]

Asi najväčšou nevýhodou funkcií Pixel-wise loss je, že nehodnotia oblasti, ako je napr. percepčná kvalita. Tento hendikep spôsobuje, že výsledky neurálnych sietí sú vzorovo príliš monotónne, bez náznaku po detailoch. [46]

Percepčná strata

Funkcie percepčnej straty fungujú na princípe zisťovania, na koľko boli zachované isté črty v $\widehat{HR}$ oproti HR. Získavanie týchto čŕt sa robí pomocou na to prispôsobenej neurálnej siete (napr. VGG [41]). Funkciu percepčnej straty možno zapísať, ako

$$\mathcal{L}_{\text{percept}}(\widehat{HR}, HR, \psi, l) = \|\psi^l(HR) - \psi^l(\widehat{HR})\|_2 \ , \tag{2.3}$$

kde ψ značí zvolenú neurálnu sieť na extrahovanie čŕt. A l značí l-tú vrstvu siete.

Keďže percepčná stratová funkcia núti sieť vytvárať snímky s vyššou percepčnou kvalitou, snímky vyzerajú esteticky krajšie aj keď zvyknú mať nižšie PSNR oproti metódam využívajúcim Pixel-wise loss. [13]

2.2 Konvolučné neurálne siete

Konvolučné neurálne siete (convolutional neural networks – CNN) majú veľa foriem, no v základe sa skladajú z troch hlavných častí – konvolučná vrstva (convolutional layer), združovacia vrstva (pooling layer) a plne prepojená vrstva (fully-connected layer).

Konvolučná vrstva používa kernely na skonvolovanie vstupnej snímky na príznakové mapy. Veľa konvolučných sietí obsahuje viacero konvolučných vrstiev, tzn. tento proces prebieha viackrát. Každý neurón konvolučnej vrstvy je prepojený so špecifickými neurónmi predošlej konvolučnej vrstvy.

Združovacia vrstva slúži na zníženie rozlíšenia príznakových máp. Jej úlohou je dosiahnuť shift-invariance. Združovacie vrstvy sa väčšinou dávajú medzi jednotlivé konvolučné vrstvy.

Úlohou plne prepojenej vrstvy je prepojiť všetky neuróny predošlej vrstvy s momentálnou vrstvou. Táto vrstva nebýva vždy prítomná a môže sa nahradiť 1×1 konvolučnou vrstvou.

Poslednou vrstvou v CNN je výstupná vrstva (output layer). Štruktúra tejto vrstvy záleží na úlohe konkrétnej konvolučnej neurálnej siete. Napr. pri klasifikačných úlohách sa používa funkcia softmax. [10]

Konvolučné neurálne siete prestavujú jeden z prvých prelomov v oblasti superrozlíšenia. Jednou z týchto sietí je SRCNN. Jedná sa o veľmi jednoduchú technológiu. Obsahuje iba 3 vrstvy, no napriek tomu dokáže generovať snímky s vyššou kvalitou, ako dovtedy používané metódy. [6]

V roku 2016 bola vymyslená DRCN. Táto technológia priniesla veľmi hlbokú rekurzívnu vrstvu, čo umožnilo vysoké zvýšenie kvality oproti doposiaľ existujúcim

technológiám. [16] Dong et al. sa pokúsili vylepšiť rýchlosť a znížiť výpočetnú obtiažnosť SRCNN. Ich práca, FSRCNN, vymenila bikúbickú interpoláciu za transponovanú konvolúciu a pridala iný tvar konvolučnej siete. [7] O rok na bola vymyslená ESPCN. Tu bola prvý krát použitá vrstva Sub-pixel, čo umožnilo dosiahnuť lepšie kvantitatívne výsledky. [40] Ďalšou technológiou bola LapSRN. Tu bola vymyslená sieť, ktorá dokáže extrahovať črty z LR snímok, na základe ktorých sú vytvárané tzv. reziduálne snímky. Tieto snímky sa kombinujú s upsampled snímkami na dosiahnutie väčšej kvality. Zároveň LapSRN dosahuje podobného výpočetného času a zložitosti ako FSRCNN. [20] Ying et al. vytvorili DRRN, ktorá používa rekurzívne učenie. Táto technológia dosiahla omnoho lepšie výsledky ako predošlé. [43] Následne vznikla EDSR kde boli použité reziduálne bloky, resp. ResNet čo umožnilo prehĺbenie siete s tým, že je možné sa vyhnúť degradácii výsledku. Táto technológia vyhrala NTIRE2017 výzvu. [25] Yiqun et al. Pouzili metód nelokálnej riedkej pozornosti (Non-local sparse attention) na vytvorenie NLSN, ktorá dosahuje porovnateľných výsledkov. [30]

2.3 Superrozlíšenie založené na GAN sieťach

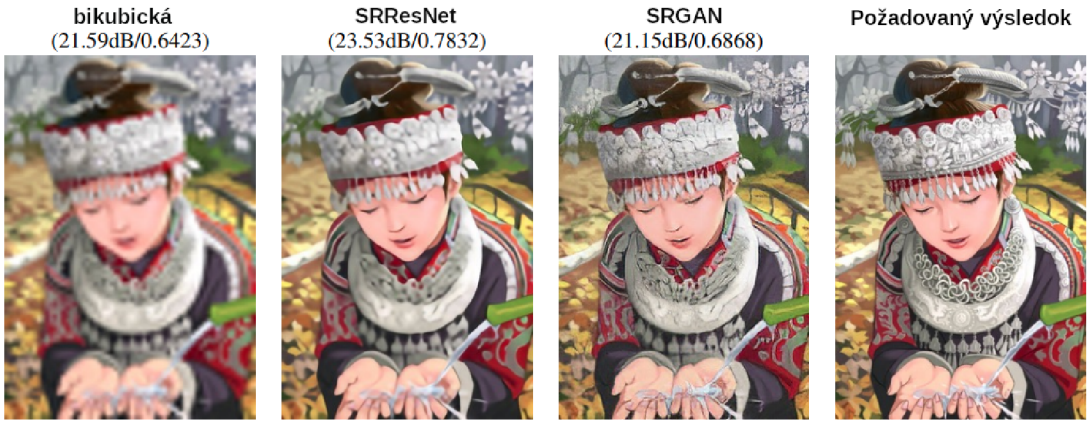
Rok 2014 priniesol nový pokrok v oblasti strojového učenia. Bol tým práve koncept generatívnej aversarialnej siete (generative adversarial network – GAN). Táto technológia zlepšila schopnosť umelej inteligencie generovať „svoje vlastné" výtvory. GAN sú založené na princípoch genetických algoritmov. Konkrétne GAN sa skladá z dvoch neurálnych sietí.

Prvá sa nazýva generátor. Je to prevažne dekonvolučná sieť, ktorej úlohou je generovať chcené dáta na základe naučených dát. Generátor je nutné naučiť hodnotám reálnych dát, na základe ktorých chceme vygenerovať „falošné" dáta, ktoré budú mať čo najbližšie k reálnym.

Diskriminátor, ktorý je druhou neurálnou sieťou v GAN, má za úlohu zistiť, či je výstup z generátora „dostatočne skutočné". Diskriminátor je prevažne konvolučná sieť, ktorá je naučená na vstupoch reálnych dát, a na základe nich zisťuje s akou pravdepodobnosťou je výstup z generátora skutočný alebo umelo vytvorený. Inak povedané, pointou generátoru je oklamať diskriminátor tým, že vygeneruje dáta, ktoré budú tak podobné so skutočnými, že diskriminátor nebude schopný vidieť rozdiel medzi nimi.

Pôvodní autori tohto konceptu prirovnali GAN ku zápasu medzi falšovateľmi peňazí a políciou, ktorá sa ich snaží odhaliť (falšovatelia sú generátor a polícia je diskriminátor). Zároveň, ako postavy v spomínanej analógii sa aj generátor a diskriminátor dokážu zlepšovať vzájomným zápasením. [9]

V oblasti superrozlíšenia zaznamenali GAN veľký úspech. Jednu z prvých implementácií je sieť SRGAN (super-resolution generative adversarial network). Ako generátor tu bola použitá sieť SRResNet. Jedná sa konvolučnú sieť, ktorá dokáže zvyšovať rozlíšenie snímok, zatiaľ čo diskriminátor sa snaží zistiť či je výsledok generátoru skutočná snímka alebo len prevzrokovaná. Vďaka tomuto SRGAN vytvára snímky z vysokou vizuálnou kvalitou. [22] O dva roky bola vytvorená ESRGAN. Táto technológia pridala RRDB bloky a upravila diskriminátor tak aby predpovedal skutočnosť snímok pomocou relatívnej skutočnosti, namiesto absolútnej. Pomocou týchto vylepšení, ESRGAN dokázal prekonať SRGAN v oblasti vizuálnej kvality, čím vyhral PIRM2018-SR výzvu. [45] V tom istom roku bola vytvorená URDGN. Ide sa o jednu z najznámejších GAN v oblasti superrozlíšenia tvárí. WGAN-GP použili tzv. Wassertein GAN aby spravili sieť stabilnejšou. FCGAN používa architektúru diskriminátoru aj generátoru založenú na U-Net sieťach. TDAE je technológia zameraná na superrozlíšenie tvárí s šumom a bez zarovnania. FaceAttr používa atribút tváre (viď 3.2.4) na zlepšenie kvality. Zatiaľčo modernejší Super-FAN používa tepelných máp (viď 3.2.2) ako vopred známu informáciu, ktoré si dokáže aj sám generovať. FSRFCH taktiež používa tepelné mapy, aby bola schopná presnejšie predvídať štruktúru tváre. [26] MLGE používa obrysy tváre vďaka čomu dosahuje výborné výsledky. [19] EIPNet používa obrysy tváre aj zachovanie identity. [17] SiGAN prináša iný spôsob zachovania identity (viď 3.3.2) vďaka čomu dosahuje výbornú percepčnú kvalitu. [11]



Obr. 2.1: Ukážka úspechu GAN (v zátvorkách sú hodnoty PSNR a SSIM). Zdroj [22] (upravené)

Adversariálna strata

Pre správne fungovanie generatívnych adversariálnych sietí je treba špeciálnych, tzv. Adversariálnych stratových funkcii. Adversariálne stratové funkcie sa skladajú z dvoch funkcií – stratová funkcia generátora ($\mathcal{L}_G$) a stratová funkcia diskriminátora ($\mathcal{L}_D$). Pôvodné technológie využívali stratové funkcie založené na krížovej entropií,

$$\mathcal{L}_{\mathrm{G}}(\widehat{HR}) = -log(\mathcal{D}(\widehat{HR})) \tag{2.4}$$

$$\mathcal{L}_{\mathrm{D}}(\widehat{HR}, HR) = -log(\mathcal{D}(HR)) - log(\mathcal{D}(\widehat{HR}))\,, \tag{2.5}$$

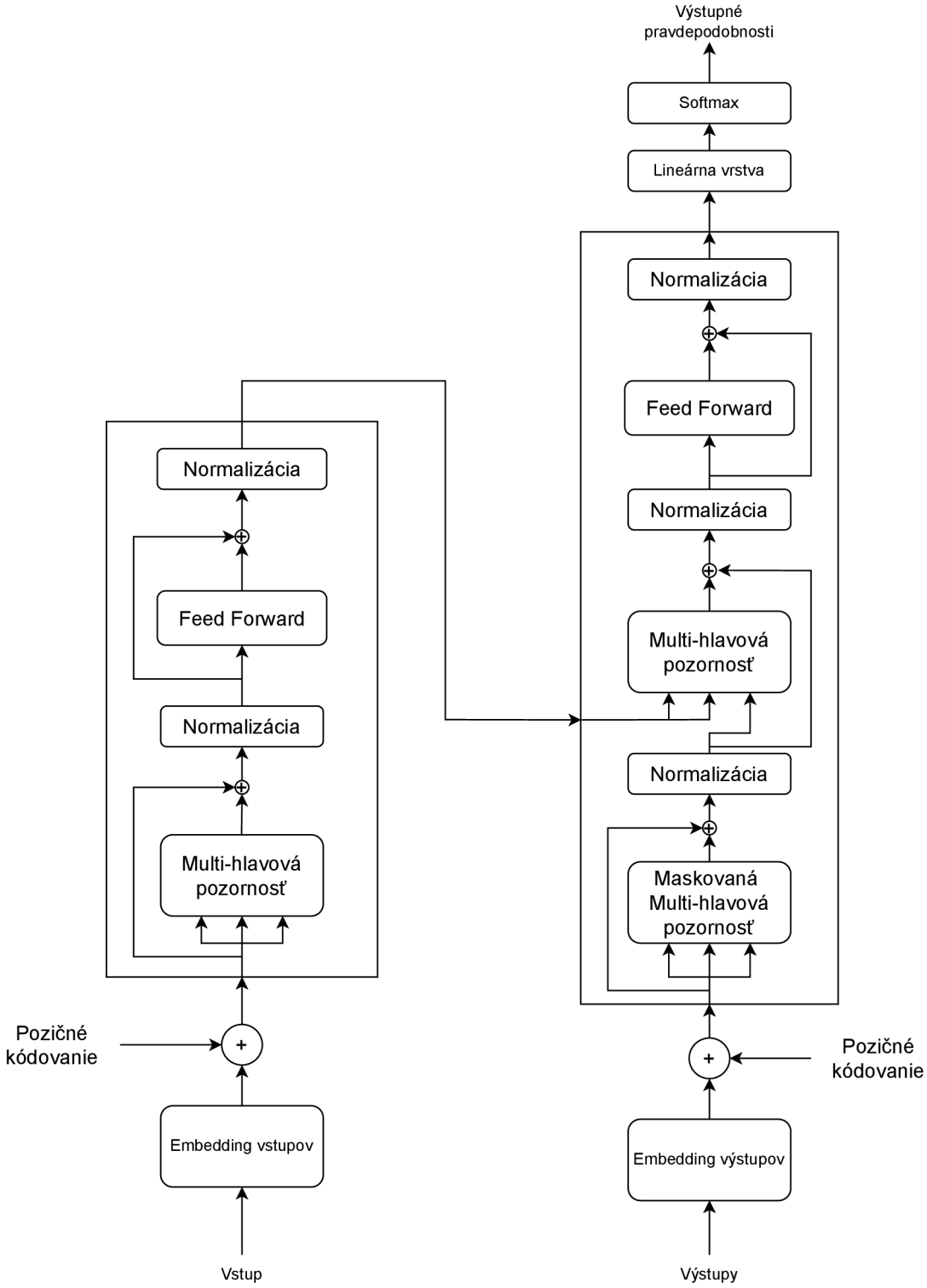
kde $\mathcal{D}$ reprezentuje diskriminátor. HR reprezentuje požadovanú snímku a $\widehat{HR}$ reprezentuje dosiahnutú snímku. [13]

2.4 Architektúra typu Transformer

V roku 2017 bol vytvorený nový druh neurálnych sietí, zameraných na spracovanie jazyka. Transformery boli vytvorené za účelom náhrady komplikovaných rekurentných modelov za jednoduchšiu technológiu. Transformery sú založené na typickej štruktúre autoenkóderu. Ich netypickou časťou je, že takmer výhradne používajú mechanizmy pozornosti. [44]
Transformer priniesol 3 hlavné kontribúcie:

- Pozičné kódovanie
- Pozornosť
- Seba-pozornosť

Pozičné kódovanie je spôsob akým dokážeme zaznamenať informáciu o pozícii slova vo vete. Pozornosťou sa rozumie mechanizmus akým sa model dokáže špeciálne zamerať na sémanticky podstatné dáta (v prípade spracovania jazyka to môžu byť kľúčové slová vo vete, v prípade počítačového videnia podstatné časti snímky). Mechanizmy pozornosti sa používali v neurálnych sieťach už aj pred vynálezom transformeru (hoci samozrejme nie v takej miere ako v transformeri). Unikátom transformeru je tzv. Seba-pozornosť.



Obr. 2.2: Schéma modelu transformer. Zdroj [44] (prekreslené)

Seba-pozornosť

Seba-pozornosť je nový druh pozornostného mechanizmu. Výstupom seba-pozornosti je vážený priemer všetkých vstupov. Seba-pozornosť sa vždy aplikuje na dáta vychádzajúce z jedného predošlého bloku (preto sa nazýva seba-pozornosť). V transformeri je používaná tzv. multi-hlavovná pozornosť (multi-head attention). Multi-hlavová pozornosť je založená na škálovanej skalárne súčinovej pozornosti (scaled dot-product attention). Multi-hlavová pozornosť sa skladá z viacerých blokov, ktoré spolu bežia paralelne.

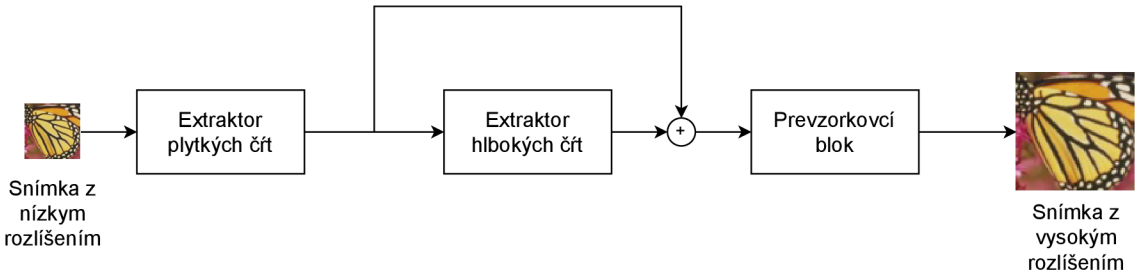
Veľkou výhodou seba-pozornosti je pomerne nízka výpočetná komplexita vzhľadom na počet vrstiev v modely. Ďalšiou výhodou je fakt, že transformery nemajú problém s učením sa vzdialených závislostí (long-range dependecies). Vzdialené závislosti sú závislosti medzi dátami, ktoré prekračujú jednotlivé domény. Napr. ak vo vetách „Vyrastal som vo Nemecku. Hovorím po nemecky.“ ak chceme predpovedať posledné slovo druhej vety potrebujeme na to informáciu z predošlej vety (inej domény). S týmto zvyknú mať rekurentné modely problém. [44] Posledným vylepšením je schopnosť paralerizácie učenia. Paralerizácia je niečo, čoho rekurentné modely neboli schopné, čo malo veľký dopad na škálu možností učenia. Transformerom sa podarilo tento problém vyriešiť, čo umožnilo učiť modely na veľkých dátových sadách s výbornými výsledkami. [44] Vďaka tomuto mohli vzniknúť komprehensívne jazykové modely ako je BERT, GPT-3 alebo GPT-4. BERT bol trénovaný na 3.3 miliardách slov. [5]

2.4.1 Transformery v oblasti superrozlišenia

V roku 2021 bol vytvorený transformer, ktorého úlohou bolo klasifikovať obrázky. Tzv. vidiaci transformer (vision transformer - ViT) používa kodér podobný bežnému transformeru, z ktorého sa informácie následne posielajú ďalej na klasifikáciu. [8] Odvtedy bolo vytvorených veľa modelov pre rôzne oblasti počítačového videnia (klasifikácia, segmetácia, a pod.). Všeobecne sa modely transformerov na superrozlíšenia skladajú z troch častí

- Extraktor plytkých čŕt
- Extraktor hlbokých čŕt
- Prevzorkovací blok

Pričom sa architektúra transformeru zvykne uplatňovať práve v extraktore hlbokých čŕt. [2]



Obr. 2.3: Schéma modelu transformeru s využitím na superrozlíšenie. Zdroj [2] (prekreslené)

Medzi známe modely patrí SwinIR [24], ESRT [28], HAT [4]. Tieto modely sú zamerané na všeobecné superrozlíšenie, t.j. nemusia byť vhodné pre superrozlíšenie tvárí. Transformery zamerané na superrozlíšenie tvárí nám nie sú známe. Výhodou transformerov je, že nezvyknú byť náročné na GPU. Obzvlášť dobrým príkladom tohoto je ESRT, ktorý zaberá len 4GB grafickej pamäte. [28]

3 Superrozlíšenie tvárí

3.1 Unikátne výzvy FSR

Pod pojmom super-rozlíšenie tvárí rozumieme aplikáciu superrozlíšenia na vylepšenie kvality ľudských tvarí. Ľudská tvár je vysoko komplexným objekt s mnohými štruktúrami, vlastnosťami, a pod. Komplikovanosť superrozlíšenia tvarí spôsobila, že sa táto problematika rieši ako individuálny problém. Existujú technológie špeciálne vytvorené pre túto záležitosť (URDGN, WaveletSRNet, SPARNet, Super-FAN, atď.). [46] Ciele technológii super-rozlíšenia tvarí možno rozdeliť na 2 kategórie,

- Vylepšenie kvality za účelom estetickým
- Vylepšenia kvality za účelom identifikácie

V prvom prípade sa jedná iba o vizuálne vylepšenie kvality. Nie je tu podstatné udržanie pôvodnosti snímky alebo identifikačných vizuálnych čŕt. Kľúčová je výsledná kvalita snímky.

Druhý prípad sa, presne naopak, zaoberá zachovaním identifikačných vizuálnych čŕt. Úlohou je zjednodušenie identifikácie osoby zachytenej na LR snímke pomocou super-rozlíšenia. V tomto prípade nie je až tak podstatná estetická kvalita.

Obidva prípady predstavujú rozdielne oblasti výskumu. Táto práca sa zaoberá iba tým druhým.

3.1.1 Kvalitatívne a kvantitatívne meranie kvality

Jedným z hlavných implikácii rozdielov medzi superrozlíšením za účelom identifikácie a superrozlíšením za účelom estetickým je fakt, že vysoké výsledky kvantitatívnych metód merania kvality (PSNR, SSIM, atď.) nemusia zaručiť dostatočné zachovanie vizuálnych identifikačných čŕt, čo dokáže sťažiť identifikáciu. Na potlačenie tohto efektu sa môžu použiť kvalitatívne metódy hodnotenia kvality. Často zvyknú byť v podobe dotazníka, kde sa respondenti rozhodujú, ktorá snímka sa najviac podobá na SR. Bolo dokázané, že rozdiel medzi kvantitatívnymi a kvalitatívnymi výsledkami je často veľký. [37]

3.1.2 Problémy GAN

Napriek tomu, ze GAN dosahujú výborné výsledky v oblasti superrozlíšenia. Nie sú schopné dôveryhodne rekonštruovať obraz. GAN zvyknú dovytvárať neexistujúce detaily, resp. vynechajú iné, podstatné, existujúce detaily. Toto nie je problém pre

superrozlíšenie za účelom estetickým, avšak pre superrozlíšenie za účelom identifikácie to môže byť fatálne. Veľa technológií takéhoto typu je preto pre našu prácu nepoužiteľných. Až extrémnym príkladom tohto fenoménu je technológia PULSE. [32] Tento model je založený na opačnom princípe, ako obvyklé modely super-rozlíšenia. PULSE sa nesnaží naučiť mapovať LR na HR ale naopak, generuje umelo vytvorené HR snímky ľudských tvári a následne im znižuje rozlíšenie, pričom jednotlivé snímky porovnáva s vstupnou LR snímkou. Snímka, ktorá je, po znížení rozlíšenia, najviac podobná vstupnej LR snímke je výstupom tejto technológie. Napriek tomu, že PULSE dokáže generovať $\widehat{HR}$ snímky s vysokou kvalitou, jedná sa o arbitrárne vygenerované snímky, a teda nie je možné aby v nich bola zachovaná pôvodná identita. Zároveň, tieto snímky zvyknú byť veľmi odlišné od požadovaných výsledkov, ako možno vidno na tejto fotke. [37]

3.2 Superrozlíšenia s vopred známymi informáciami

Na uľahčenie práce so snímkami komplexných objektov ako sú ľudské tváre, začali vznikať spôsoby ako zachytávať niektoré charakteristiky a poskytnúť ich technológiám superrozlíšenia. Medzi tieto charakteristiky patria - orientačné body (facial landmarks), tepelné mapy (facial heatmaps), parsovacie mapy (facial parse maps), atributy tváre (facial attributes), informácia o identite (identity information). Tieto charakteristiky sa zvyknú tiež nazývať vopred známe informácie (priori information). [13]

3.2.1 Orientačné body

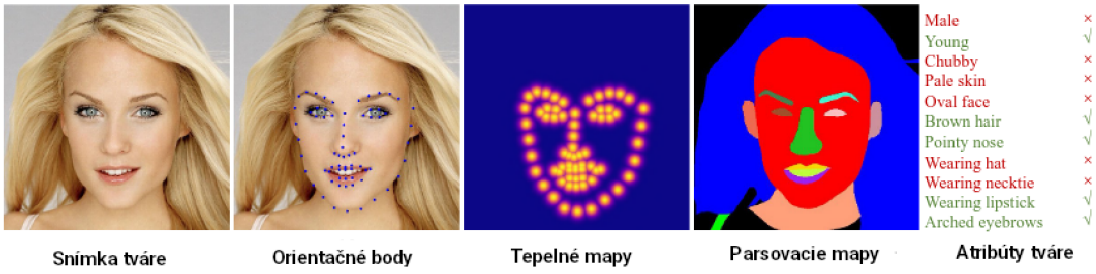
Určujú kľúčové komponenty tváre. Umožňujú technológiám superrozlíšenia lepšie sa orientovať v snímke. Je nutné ich vopred preddefinovať vo fotkách. Existujú databázy ktoré obsahujú veľký počet týchto bodov, napr. Helen [21]

3.2.2 Tepelné mapy

Sú založené na orientačných bodoch. Narozdiel od nich neurčujú priamo kde sa kľúčové komponenty nachádzajú ale určujú pravdepodobnosť, že daný bod bude kľúčový komponent. [13]

3.2.3 Parsovacie mapy

Jedná sa segmentačné mapy, ktorých úlohou je oddeliť od seba sémanticky rozdielne časti tváre, ako sú oči, pery, obočie, koža, atď. [13]



Obr. 3.1: Ukážka vopred známych informácií. Zdroj [13] (preložené)

3.2.4 Atribúty tváre

Atribúty, ako sú pohlavie, má/nemá rúž, má/nemá okuliare, atď. Sa môžu využiť v technológiách superrozlíšenia. Tieto atribúty totižto môžu byť problematické na určenie z LR snímky, a teda sieť nemusí byť schopná ich rekonštrukcie v $\widehat{HR}$ snímke. Vďaka atribútom tváre je možné priamo technológií povedať, čo má v snímke očakávať, čím môžeme zvýšiť presnosť s pôvodnou HR snímkou. Na zjednodušenie sa zvyknú brať v úvahu binárne atribúty. [13]

3.2.5 Informácia o identite

Každá ľudská tvár zodpovedá unikátnej identite. Tento druh informácie je možný zaznamenať a použiť na zlepšenie rekonštrukcie HR snímky. Zrekonštruovaná $\widehat{HR}$ snímka by mala zodpovedať tej istej identite ako pôvodná HR snímka. Informácia o identite sa ukázala, ako veľmi užitočný spôsob zlepšenia výsledkov technológií superrozlíšenia. [13] [47] [11]

3.3 Superrozlíšenie so zachovaním identity

Pod pojmom superrozlíšenie so zachovaním identity rozumieme superrozlíšenie za pomoci znalosti informácie o identite. V posledných rokoch sa tejto oblasti venovalo viac pozornosti, keďže sa ukázalo, že je schopná priniesť výborných výsledkov. [47] [11] Jedná sa zároveň o veľmi zaujímavú oblasť z pohľadu tejto práce, keďže zachovanie identity je kľúčové pre následnú identifikáciu človeka na snímke.

Zachovanie identity môže byť dosiahnuté dvomi spôsobmi, a t.j. zachovanie identity pomocou rozpoznávania tváre a zachovanie identity párovými metódami.

3.3.1 Zachovanie identity pomocou rozpoznávania tváre

Tento systém používa sieť na rozpoznávanie identity tvárí aby bolo možné určiť identitu snímok. Na základe určenia identity HR a $\widehat{HR}$ snímky je následne možné vypočítať rozdiel v identite medzi HR a $\widehat{HR}$. Na základe tohto poznatku môžeme definovať tzv. identitovú stratu (identity loss),

$$\mathcal{L}_{\text{identity}}(\widehat{HR}, HR) = \|FR(\widehat{HR}) - FR(HR)\|_k \,, \tag{3.1}$$

kde FR reprezentuje funkciu siete na rozpoznávanie identity tvárí, k je buď 1 alebo 2, čo záleží od technológie. [13]

Aplikácia tejto straty závisí od individuálnej technológie. Jednou z najjednoduchších je SiCNN (Super-identity convolutional neural network). Tento model používa 2 CNN. Prvá prevzorkuje snímku na vyššie rozlíšenie, zatiaľ čo druhá aplikuje metódu zachovania identity na prevzorkovanú snímku, aby jej znovu zlepšila kvalitu na základe identity. [47]

3.3.2 Zachovanie identity párovými metódami

Najväčším problémom technologii založených na zachovaní identity pomocou rozpoznávania tváre je práve sieť na rozpoznávanie identity tvárí. Tieto siete potrebujú veľmi dobre označené dátové sady. To znamená, že každá snímka má k sebe priradenú presnú identitu. Tvorba takýchto dátových sád je veľmi obtiažna. [13]

Na riešenie tohto problému sa môžu použiť tzv. slabo označené párové dátové sady. Toto označenie hovorí len o tom či snímky z daného páru prislúchajú k rovnakej identite, bez toho aby bolo nutné poznať presnú identitu tvárí v snímkach. [11]

4 Popis experimentu

4.1 Dátové sady

Jednou z kľúčových častí strojového učenia sú dáta. Modely strojového učenia potrebujú dáta na to, aby boli schopné si nastaviť vnútorné parametre, tak aby dokázali pracovať s čo najväčšou presnosťou. Rovnako to funguje pri superrozlíšení tvárí založenom na neurálnych sieťach.

Je empiricky dokázané, že technológie natrénované na snímkach ľudských tvárí dosahujú lepšie výsledky pri superrozlíšení tvárí ako keď sú natrénované na bežných snímkach. [46] Preto budeme výber dátových sád zameriavať výhradne na sady, ktoré obsahujú výhradne ľudské tváre. Dátové sady obsahujú veľa snímok ľudských tvárí. Niektoré obsahujú len *HR* snímky a *LR* si musíme vyrobiť sami, zatiaľčo iné obsahujú páry, resp. skupiny *HR* a *LR* snímok. V nasledujúcej tabuľke môžeme vidieť príklady niekoľkých takýchto dátových sád.

Tab. 4.1: Ukážka niektorých dátových sád.

Meno dátovej sady	počet snímkov
CelebA [27]	202 599
VGGFace [36]	3 310 000
Helen [21]	2 330
MLFDB [37]	14 800
FFHQ [14]	70 000

V našom experimente sme si zvolili dátovú sadu CelebA.

4.1.1 CelebA

CelebA je veľmi známa a osvedčená dátová sada. Obsahuje viac ako 200 tisíc snímok celebrít. Bola vytvorená pokrokom v metóde predpovedania tvárových atribút. Obsahuje výhradne *HR* snímky. CelebA bola použitá vo veľkom počte štúdií v oblasti superrozlíšenia. Je možné ju považovať za jednu z najpopulárnejších dátových sád. [27]

Pri tvorbe *LR* snímok sme použili 4 druhy interpolacií – bikubická, bilineárna, najbližších susedov, lanczos. Na jednotlivé snímky sme použili vždy len jednu interpoláciu, zvolenú náhodným výberom. T.j. každá interpolácia mala 20% šancu, že bude aplikovaná. Tento postup sme zvolili aby sa zamedzilo fenoménu, keď sa sieť naučí pracovať len so snímkami vytvoernými len jedným druhom interpolácie.

Tento fenomén spôsobí, že sieť bude dosahovať horšie výsledky pri iných formách degradácie. [37]

4.1.2 MLFDB

MLFDB je relatívne nová dátová sada, zameraná na superrozlíšenie tvárí s viacerými snímkami. Bola vytvorená z viac ako 300 videí na stránke YouTube a obsahuje viac ako 14 tisíc fotiek. Tieto fotky prislúchajú približne 6 000 rôznym osobám. Dátová sada bola vytvorená pomocou technológii na rozpoznávanie tvárí. Fotky sú rozdelené na osmice. Pričom 1 snímka je *HR* a zvyšných 7 je *LR*. MLFDB otvára dvere do neprebádané oblasti superrozlíšenia tvárí s viacerými snímkami. [37]

4.2 Metódy merania kvality

4.2.1 PSNR

Vrcholový pomer signálu k šumu (Peak signal-to-noise ratio – PSNR) je jedna z najznámejších metrík na porovnávanie kvality snímok. Je založená na priemernej kvadratickej chyba (MSE).

$$\mathrm{MSE}(\widehat{HR}, HR) = \frac{1}{hwc} \sum_{i,j,k} (\widehat{HR}_{i,j,k} - HR_{i,j,k})^2 \ , \tag{4.1}$$

$$\mathrm{PSNR}(\widehat{HR}, HR) = 10\, log_{10} \left(\frac{MAX^2}{MSE} \right) \ , \tag{4.2}$$

kde MAX reprezentuje najvyššiu možnú hodnotu pixela (255 pre 8-bitové snímky), zvyšná notácia je rovnaká ako v 2.2.

Napriek svojej popularite naprieč mnohými oblasťami je veľmi dobre známe, že PSNR nezvykne dobre kolerovať so subjektívnym hodnotením kvality človekom. [13]

4.2.2 SSIM

Meranie indexu štrukturálnej podobnosti (structural similarity index measure – SSIM) sa, na rozdiel od PSNR, zameriava na porovnávanie štrukturálnej podobnosti. SSIM meria tri oblasti fotky - jas, kontrast a štruktúra. SSIM možno matematicky zapísať ako,

$$\mathrm{SSIM}(\widehat{HR}, HR) = L(\widehat{HR}, HR) * C(\widehat{HR}, HR) * S(\widehat{HR}, HR) \ , \tag{4.3}$$

kde $L(\widehat{HR}, HR)$, $C(\widehat{HR}, HR)$, $S(\widehat{HR}, HR)$ sú jas, kontrast a štruktúra snímok nasledovne. Tieto časti možno zapísať ako

$$L(\widehat{HR}, HR) = \frac{2\mu_{\widehat{HR}}\mu_{HR} + c_1}{\mu^2_{\widehat{HR}} + \mu^2_{HR} + c_1}, \tag{4.4}$$

$$C(\widehat{HR}, HR) = \frac{2\sigma_{\widehat{HR}}\sigma_{HR} + c_2}{\sigma^2_{\widehat{HR}} + \sigma^2_{HR} + c_2}, \tag{4.5}$$

$$S(\widehat{HR}, HR) = \frac{\sigma_{\widehat{HR}HR} + \frac{c_2}{2}}{\sigma_{\widehat{HR}}\sigma_{HR} + \frac{c_2}{2}}, \tag{4.6}$$

kde $\mu_{\widehat{HR}}$ je pixel s mediánovou hodnotou $\widehat{HR}$, μ_{HR} je pixel s mediánovou hodnotou HR, $\sigma_{\widehat{HR}}$ je rozptyl $\widehat{HR}$, σ_{HR} je rozptyl HR, $\sigma_{\widehat{HR}HR}$ je kovariancia medzi $\widehat{HR}$ a HR a c_1 a c_2 sú stabilizačné premenné. [46]

4.2.3 LPIPS

Naučená percepčná obrázková "patchová" podobnosť (Learned perceptual image patch similarity – LPIPS) je percepčná hodnotiaca metrika. LPIPS porovnáva rozdiely medzi snímkami pomocou extrahovania čŕt snímok. Na extrakciu sa používajú neurálne siete (napr. VGG, Adam, atď.). Výsledky tejto metriky kolerujú so subjektívnym ľudským hodnotením oveľa viac ako PSNR a SSIM. [48]

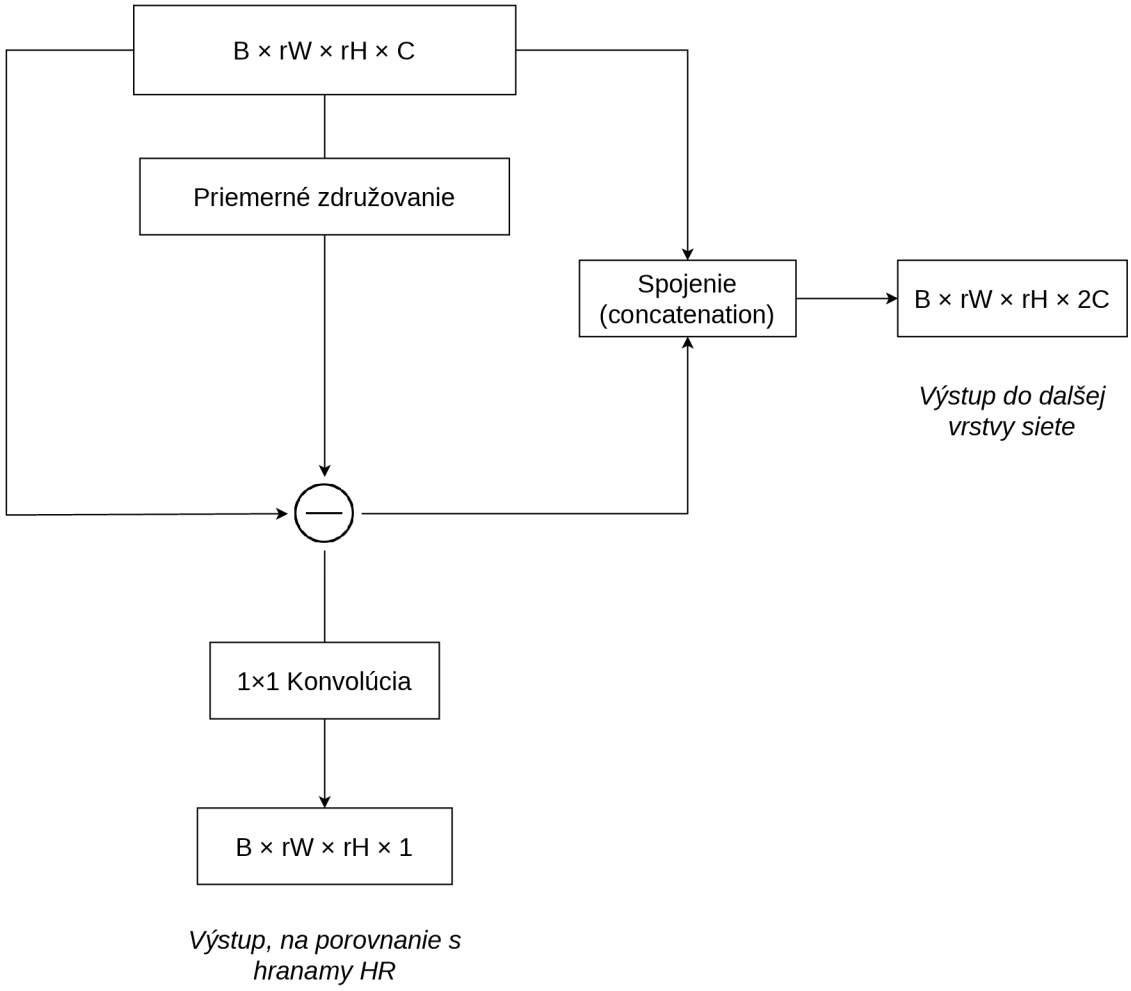
4.3 Zvolene technológie

4.3.1 EIPNet

V roku 2021 bola vytvorená technológia EIPNet. Jedná sa o jednu z najmodernejších technológií v oblasti super-rozlíšenia tvárí. EIPNet využíva dvoch vopred známych informácií, a t.j. Informáciu o identite a snímku obrysov tváre. EIPNet vo svojej architektúre využíva tzv. Hranové bloky (edge blocks) na získavanie hrán snímky počas jej prevzorkovania. V sieti je viacero hranových blokov, tzn. hrany snímky sú získavané pri rôznych rozlíšeniach. Výstup týchto blokov je následne porovnávaný s hranami získanými z HR snímky pomocou tzv. cannyho detektoru hrán. Z toho jasne vyplýva, že sieť dokáže merať rozdiely medzi hranami HR a hranami získanými hranovými blokmi, tzn. tento systém prináša tzv. Hranovú stratovú funkciu, ktorú môžeme zapísať ako,

$$\mathcal{L}_{\mathrm{e}} = \frac{1}{r^2 wh} \sum_{i,j}^{rw,rh} \left(C(HR_{i,j}) - E(I_{i,j}) \right) , \tag{4.7}$$

kde I reprezentuje snímku prevzorkovanú o faktor r (nemusí sa jednať o $\widehat{HR}$, keďže hranových blokov je v sieti niekoľko a všetky okrem posledného majú na vstupe len čiastočne prevzorkovanú snímku), C je cannyho detektor hrán a E je hranový blok. [17]



Obr. 4.1: Nákres hranového bloku. Zdroj [17] (prekreslené)

Zachovanie identity

Okrem Zachovania obrysov sa sieť snaží zachovať aj identity človeka na fotke. Tu sa uplatňuje metóda zachovania identity pomocou rozpoznávania tváre. Konkrétne je tu použitá sieť Inception1, ktorá bola predtrénovaná na dátovej sade VGGFace. [17]

Svetelno-chrominančná strata

EIPNet zároveň počíta tzv. Svetelno-chrominančnú stratu. Autori vychádzajú z idei, že pri reprezentací snímku v YUV farebnom modeli je možné dosiahnuť lepšiu percepčnú kvalitu v superrozlišení, ako pri použití bežného modelu RGB. Technológia si prevedie obrázky z RGB do YUV modelu pomocou konverznej matice a následne vypočíta priemernú kvadratickú chybu v YUV. [17]

Celkovo má stratová funkcia technológie EIPNet tvar,

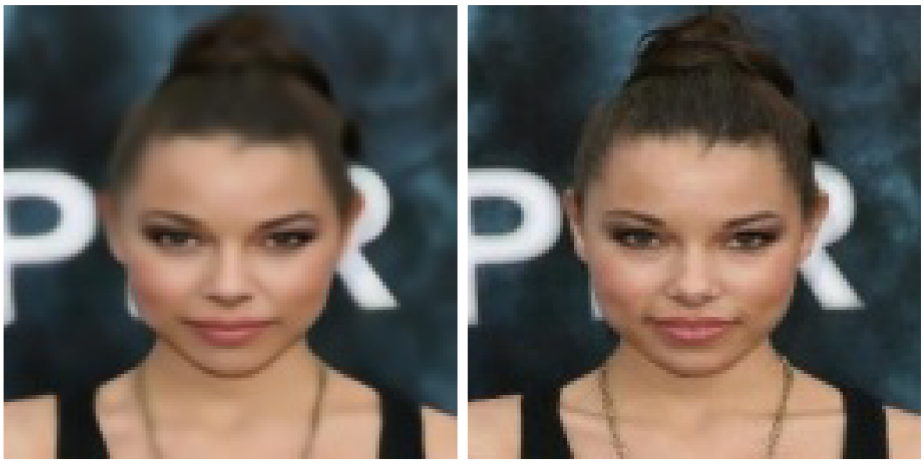
$$\mathcal{L}_{\text{celková}} = \mathcal{L}_2 + \gamma\mathcal{L}_{\text{e}} + \mathcal{L}_{\text{lc}} + \alpha\mathcal{L}_{\text{id}} + \beta\mathcal{L}_{\text{ad}}\,, \tag{4.8}$$

kde $\mathcal{L}_{\text{lc}}$ je svetelno-chrominančná strata, $\mathcal{L}_{\text{id}}$ je strata identity, $\mathcal{L}_{\text{ad}}$ je adversaríalna strata (jedná sa o GAN) a γ, α, β sú váhové parametre. [17]

Trénovanie siete

EIPNet sme natrénovali na CelebA dátovej sade. Sada bola rozdelená na dve množiny – testovaciu a trénovaciu množinu. Pričom trénovacia bola použitá na trénovanie a testovacia na testovanie siete. Testovacia množina obsahovala okolo 60 000 snímok zatiaľčo trénovacia obsahovala okolo 140 000 snímok (t.j. CelebA bola rozdelená približne v pomere 1 : 2).

Na trénovanie sme použili skript `train.py`. Parametre sme zachovali podľa autorov. Počet epoch bol nastavený na 200. Pri 113. epoche sme však trénovanie zastavili, keďže už nedochádzalo k zlepšovaniu. Pred spustením bolo však nutné ešte sieť upraviť, keďže bola naprogramovaná na 8× superrozlíšenie a my pracujeme s 4×. Konkrétne bolo nutné upraviť generátor aby fotku prevzorkoval len 4-krát.



$\widehat{HR}$ HR

Obr. 4.2: Ukažka fotky z modelu EIPNet

4.3.2 WaveletSRNet

Jedná sa o konvolučnú neurálnu sieť, ktorá využíva metód vlnovej transformácie (Wavelet transform). Konkrétne, pred tým než sa sieť pokúsi predpovedať *HR* snímku sa najprv pokúsi vygenerovať tzv. vlnové koeficienty na základe ktorých sa pokúsi zrekonštruovať *HR* snímku. Pomocou vlnovej transformácie sa *HR* snímka pretvorí na vlnové koeficienty, na ktoré sa sieť snaží dostať z *LR* snímky. Výhodou tejto metódy je, že sieť je schopná lepšie zachytiť a zrekonštruovať percepčné črty, čo bežná konvolučná sieť nedokáže. WaveltSRNet sa skladá z troch sietí – vkladacia sieť (embeding net), sieť vlnovej predikcie (wavelet prediction net) a rekonštrukčná sieť (reconstruction net). Vkladacia sieť má za úlohu reprezentovať vstupnú snímku ako prízankové mapy. Tie putujú dalej do siete vlnovej predickie kde budú predikované vlnové koeficienty. V rekoštrukčnej sieti sa vlnové koeficienty prerobia na $\widehat{HR}$. Model používa tzv. vlnovú stratu na porovnávanie odchýlky vygenerovaných vlnových koeficientov s požadovanými. [12]

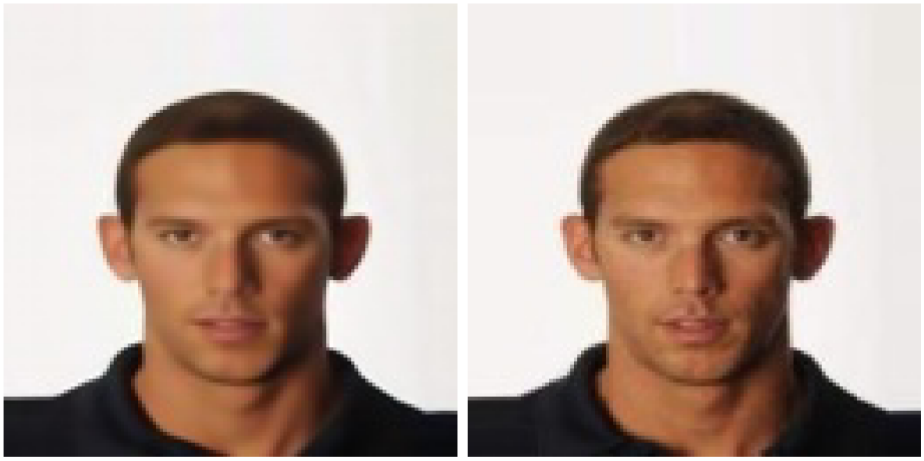
Celkovú stratovú funkciu možeme zapísať ako,

$$\mathcal{L}_{\text{celkova}} = \mathcal{L}_{\text{vlnova}} + \alpha\mathcal{L}_{\text{texturova}} + \beta\mathcal{L}_2\,, \tag{4.9}$$

kde $\mathcal{L}_{\text{texturova}}$ je tzv. texturová strata, ktorej úlohou je zabrániť vysokofrekvenčným vlnovým koeficientom aby konvergovali k nule, čím zabraňuje degradácii kvality snímky a α, β sú váhové parametre. $\mathcal{L}_2$ je taktiež použitá na zachovanie rovnováhy medzi hladkosťou výsledkov a zachovaním podstatných čŕt. [12]

Trénovanie siete

Sieť sme trénovali na dátovej sade CelebA, ktorú sme rozdelili na testovaciu a trénovaciu množinu. Testovacia množina obsahovala približne 35 000 snímok, zatiaľčo trénovacia obsahovala zvyšných približne 167 000 snímok. Sieť sme trénovali po dobu 200 epoch, zvyšné nastavenia sme nechali rovnako ako v pôvodnej štúdií. Technológia je schopná 4×, 8× a 16× superrozlíšenia. Pracovali sme so 4×.



$\widehat{HR}$ HR

Obr. 4.3: Ukažka fotky z modelu WaveletSRNet

4.3.3 NLSN

Nelokálna riedka sieť (Non-local sparse network – NLSN) je model z roku 2021. Je založený na princípoch nelokálnej riedkej pozornosti. Nelokálna pozornosť je špecíalny druh sebapozornosti, ktorá, zjednodušene povedané, počíta podobnostné skóre medzi párom pozícii na obrázku. Touto metódou je možné lepšie zachytávať vzdialené závislosti. NLSN priniesla nelokálnu riedku pozornosť. Táto forma pozornosti sa snaží potláčať obmedzený rozsah medzi porovnávanými pozíciami. [31]

Ako stratovú funkciu sme zvolili

$$\mathcal{L}_{\text{celková}} = 2 * \mathcal{L}_1 + 0.6 * \mathcal{L}_{\text{ssim}} + 0.2 * \mathcal{L}_{\text{vgg}}\,, \tag{4.10}$$

kde $\mathcal{L}_{\text{ssim}}$ je stratová funkcia založená na metrike ssim a $\mathcal{L}_{\text{vgg}}$ je percepčná stratová funkcia založená na sieti VGG.

$\mathcal{L}_{\text{ssim}}$ má tvar

$$\mathcal{L}_{\text{ssim}} = 1 - SSIM(\widehat{HR}, HR) \tag{4.11}$$

kde $SSIM(\widehat{HR}, HR)$ je metrika SSIM z $\widehat{HR}$ a HR snímok.

Úlohou $\mathcal{L}_{\text{ssim}}$ je, okrem udržania štruktúry, aj potláčanie šumu, ktorý spôsobuje $\mathcal{L}_{\text{vgg}}$. Šum sa nám bohužiaľ nepodarilo kompletne odstrániť. Skúšali sme aj tzv. stratu celkového rozptylu (total variation loss) [46], avšak tá príliš zhoršovala kvalitu snímok. Napriek čiastočnému zašumeniu si podľa nášho názoru zachovávajú dobrú mieru vizuálnej kvality.

Trénovanie siete

Sieť sme trénovali na dátovej sade CelebA. Trénovali sme na približne 130 tisíc snímkach, zatiaľ čo testovanie prebehlo na približne 70 tisíc snímkach. Sieť sme trénovali po dobu 200 epoch.

$\widehat{HR}$ HR

Obr. 4.4: Ukažka fotky z modelu NLSN

4.3.4 HAT

Hybridný pozornostný transformer (Hybrit attention transformer – HAT) je technológia z roku 2023. Jedná sa o jeden z najmodernejších modelov v oblasti vidiacich transformerov. Je založený na rovnakej schéme ako 2.3. Pričom na extrakciu plytkých čŕt sa používa bežná konvolúcia. Extraktor hlbokých čŕt sa skladá z tzv. reziduálnych hybridných pozoronostných grúp (Residual Hybrid Attention Groups – RHAG). Tieto bloky obsahujú viacero hybridných pozornostných blokov (Hybrid Attetion Block – HAB). HAB kombinuje kanálovú pozornosť (channel attention) so seba-pozornosťou. Autori vychádzajú z predpokladu, že kombinácia konvolucii (v kanálovej pozornosti) a seba-pozornosti dokážu vylepšiť výsledky modelu. Na záver, sa na prevzorkovanie používa vrstva Sub-pixel. [28]

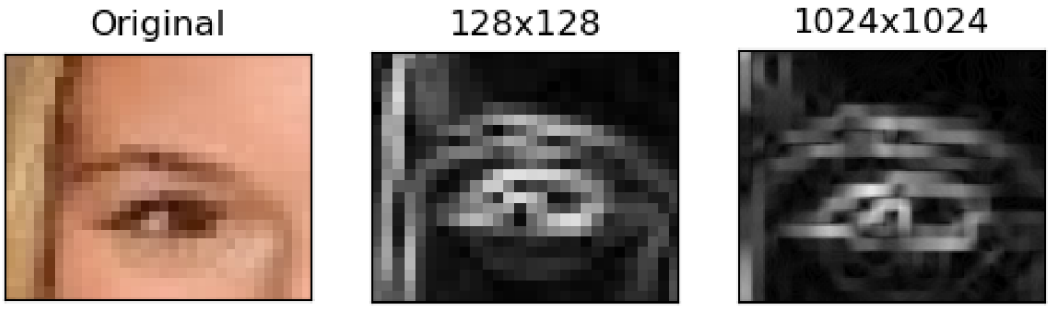
Ako stratovú funkciu sieť používa $\mathcal{L}_1$, táto funkcia nebola dostatočné pre naše účely. Na superrozlíšenie tvárí sme použili nasledujúcu stratovú funkciu

$$\mathcal{L}_{\text{celková}} = 2 * \mathcal{L}_1 + 0.8 * \mathcal{L}_{\text{ssim}} + 0.2 * \mathcal{L}_{\text{vgg}} + 0.5 * \mathcal{L}_{\text{sobel}}\,, \tag{4.12}$$

kde $\mathcal{L}_{\text{ssim}}$ je stratová funkcia založená na metrike ssim, $\mathcal{L}_{\text{vgg}}$ je precepčná strata a $\mathcal{L}_{\text{sobel}}$ je hranová strata.

Hranová strata

Ako hranovú stratu používame Sobelovú detekciu hrán (Sobel edge detection). [42] Pracujeme so snímkami s veľkosťami 128×128 pixelov. Tieto snímky sú natoľko malé, že Sobelov operátor má problémy správne detekovať mnohé črty. Preto pre tým než aplikujeme na snímky Sobelov operátor ich prevzorkujeme na veľkosť 1024×1024 pomocou bikúbickej interpolácie.



Obr. 4.5: Ukážka rozdielu zachytenia obrysov obrázku po aplikácii sobelovho operátora pri rozdielnych veľkostiach

Nevýhodou tohto prístupu je, že práca so snímkami veľkého rozlíšenia je pamäťovo náročná a zároveň spomaľuje trénovanie siete. Zároveň prevzorkovanie snímku spôsobí, že snímky už nebudú identické, keďže bikúbická interpolácia si dopočíta pixely, ktoré tam predtým neboli. V snímkach však nebol veľký rozdiel takže to nepovažujeme za veľký problém. Aplikáciou tejto metódy detekcie hrán sa nám podarilo zvýšiť PSNR fotiek v priemere o pár desatín dB.

Podobne ako v predošlom modely aj tu snímky trpia čiastočným zašumením, ktoré sa nám úplne nepodarilo odstrániť ani pomocou SSIM. Použitie straty celkového rozptylu sa aj v tomto prípade ukázalo ako neúspešné. Našťastie šum ani v tomto prípade nie je veľký a nemyslíme si, že bude mať veľký vplyv na vizuálnu kvalitu snímok z pohľadu identifikácie.

HAT sa ukázal ako model schopný reštaurovať snímky ľudských tvarí s dobrou kvalitou. To svedčí o potencíale architektúr typu transformer v oblasti superrozlíšenia ľudských tvarí. Model zaostáva za ostatnými iba v oblasti zachovania odtieňov tváre.

Trénovanie siete

Model sa (aj bez našej metódy detekcie hrán) trénoval pomaly. Trénovali sme ho len 10 epoch, no to bolo dostačujúce na dosiahnutie dobrých výsledkov. Trénovali

sme na dátovej sade CelebA. Model sme trénovali na približne 130 tisíc obrázkoch, testovanie následne prebehlo na približne 70 tisíc obrázkoch.

$\widehat{HR}$ HR

Obr. 4.6: Ukažka fotky z modelu HAT

4.3.5 SRDD*

Superrozlíšenie pomocou hlbokého slovníku (super-resolution using deep dictionary - SRDD) je pomerne nová metóda. Na rozdiel od bežných metód super rozlíšenia, je založená na tzv. metóde sparse-coding. Sieť sa učí vytvárať tzv. slovník, ktorý sa používa na premenu LR snímky na $\widehat{HR}$. Na rozdiel od iných sparse-coding metód sa v tomto prípade slovník vytvára z náhodného šumu. [29]

Pri práci so sieťou sme pozmenili pôvodnú architektúru za cieľom dosiahnutia lepších výsledkov, preto budeme našu sieť nazývať SRDD*. V nasledujúcich častiach prejdeme niektoré zo zmien.

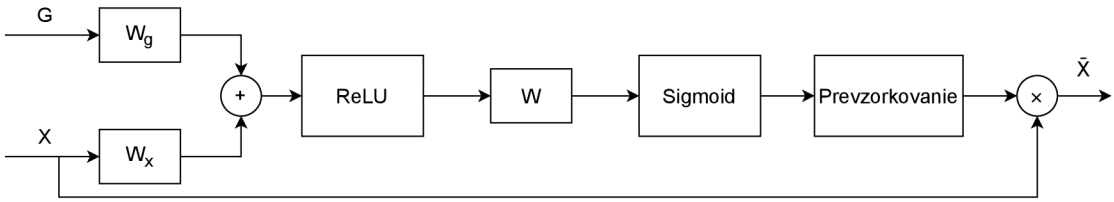
Aktivačná funkcia xUnit

Aktivačná funkcia xUnit je nový druh trénovateľnej aktivačnej funkcie, ktorá dokáže zachytávať viac komplexných čŕt ako bežné aktivačné funkcie. [18] Experimentáciou sa ukázalo, že SRDD* dosahuje najlepšie výsledky ak xUnit použijeme v extraktore hlbokých čŕt a pri rekonštrukcii $\widehat{HR}$ snímky.

Obr. 4.7: Porovnanie aktivacii na sieti SRCNN s pomocou ReLU a s pomocou xUnit. Zdroj [18] (preložené)

AttentionUNet++

SRDD používa upravenú UNet++ ako extraktor hlbokých čŕt. Rozhodli sme sa ju prerobiť na UNet++, ktorá používa pozornosť. Pozornostná UNet++ sa ukázal ako výhodný spôsob segmentácie obrazu hlavne na medicínske účely. [35] Architektúru UNet++ sme ponechali rovnakú ako v pôvodnom SRDD, len sme pridali poznornostné brány na reziduálne prepojenia, rovnako ako v [35].



Obr. 4.8: Schéma pozornostnej brány. Zdroj [35] (prekreslené)

Brána má na vstupe dva signály – X a G. X je vstup z reziduálneho prepojenia, zatiaľ čo G je vstup z uzlu, ktorý sa nachádza „pod“ bránou. Na vstupy sa aplikujú lineárne vrstvy W_x a W_g. Tie sa následne sčítajú a na ich súčet sa aplikuje aktivačná funkcia ReLU. Na výstup z ReLU sa použije ďalšia lineárna vrstva a jej výstup sa ďalej pošle do funkcie Sigmoid. Výstup zo funkcie Sigmoid je následne prevzorkovaný na potrebnú veľkosť a na záver je spojený (concatenated) so signálom X. [35]

Použili sme nasledovnú stratovú funkciu

$$\mathcal{L}_{\text{celková}} = 2 * \mathcal{L}_1 + 0.8 * \mathcal{L}_{\text{ssim}} + 0.5 * \mathcal{L}_{\text{LoG}} \,, \tag{4.13}$$

kde $\mathcal{L}_{\text{LoG}}$ je hranová strata založená na LoG (laplacian of gaussian) operátore. [1] LoG operátor sme použili, keďže bežný Sobelov operátor zachytával zbytočne veľa detailov. Detekciu hrán sme aplikovali rovnako ako v 4.3.4.

Na rozdiel od iných sietí sme v tomto prípade nepoužili VGG stratovú funkciu. VGG opäť vytvárala pomerne veľký šum, čo degradovalo kvalitu fotky. Experimentálne sme zistili, že SRDD* dokáže vytvárať fotky s relatívne dobrou kvalitou ja bez VGG stratovej funkcie. Preto sme sa ju rozhodli odstrániť. Nevýhodou tohto prístupu je, že bez VGG stratovej funkcie sa niektoré fotky javia čiastočne rozmazane (viď 4.9). Zároveň, kvôli nepoužitiu percepčnej straty, dosahujú fotky horší LPIPS. Nevidíme v tom však veľký problém, keďže snímky, napriek tomu, vyzerajú podobne z pohľadu identifikácie.

Trénovanie siete

Sieť sme trénovali na rovnaých parametroch ako predošlé siete. Množinu CelebA sme rozdelili na 130 tisíc trénovacích snímok, zvyšok (okolo 70 tisíc) sme použili na testovanie. Trénovali sme po dobu 200 epoch. Trénovanie trvalo približne 4 dni.

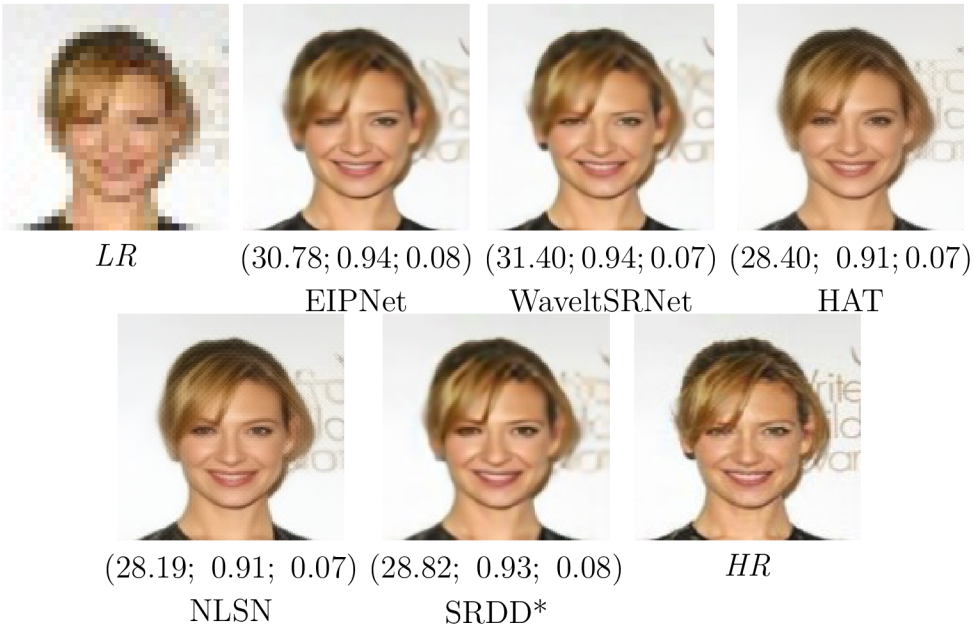
$\widehat{HR}$ HR

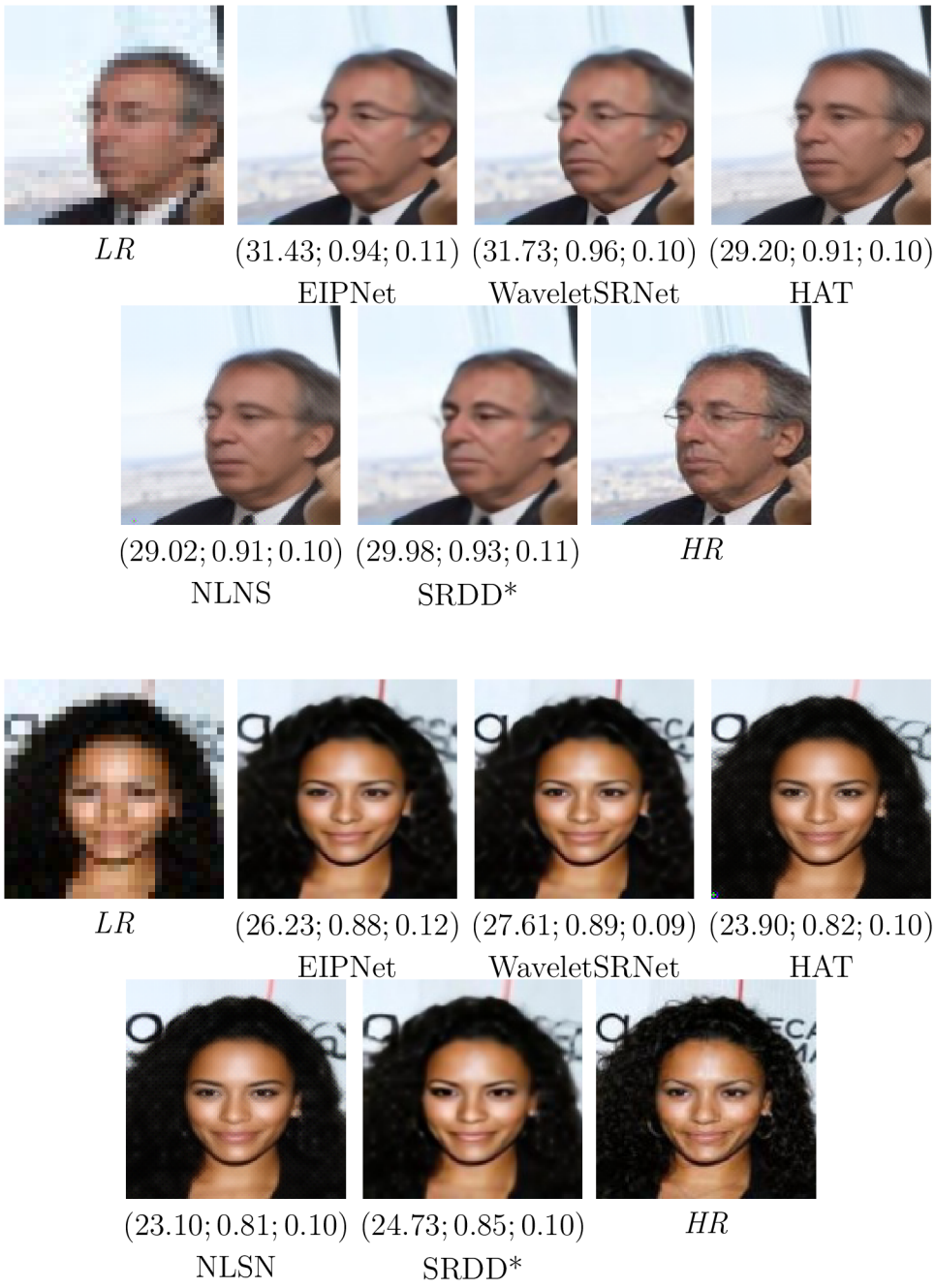
Obr. 4.9: Ukážka výstupu modelu SRDD*

5 Výsledky experimentu

Tab. 5.1: Porovnanie výsledkov na dátovej sade CelebA pri 4× superrozlíšení. Najlepšie hodnoty sú červenou.

CelebA			
Model	**PSNR [dB]**	**SSIM**	**LPIPS**
EIPNet	28.87	0.92	0.119
WaveletSRNet	29.42	0.93	0.109
NLSN	27.34	0.90	0.099
HAT	27.42	0.90	0.097
SRDD*	28.20	0.91	0.113





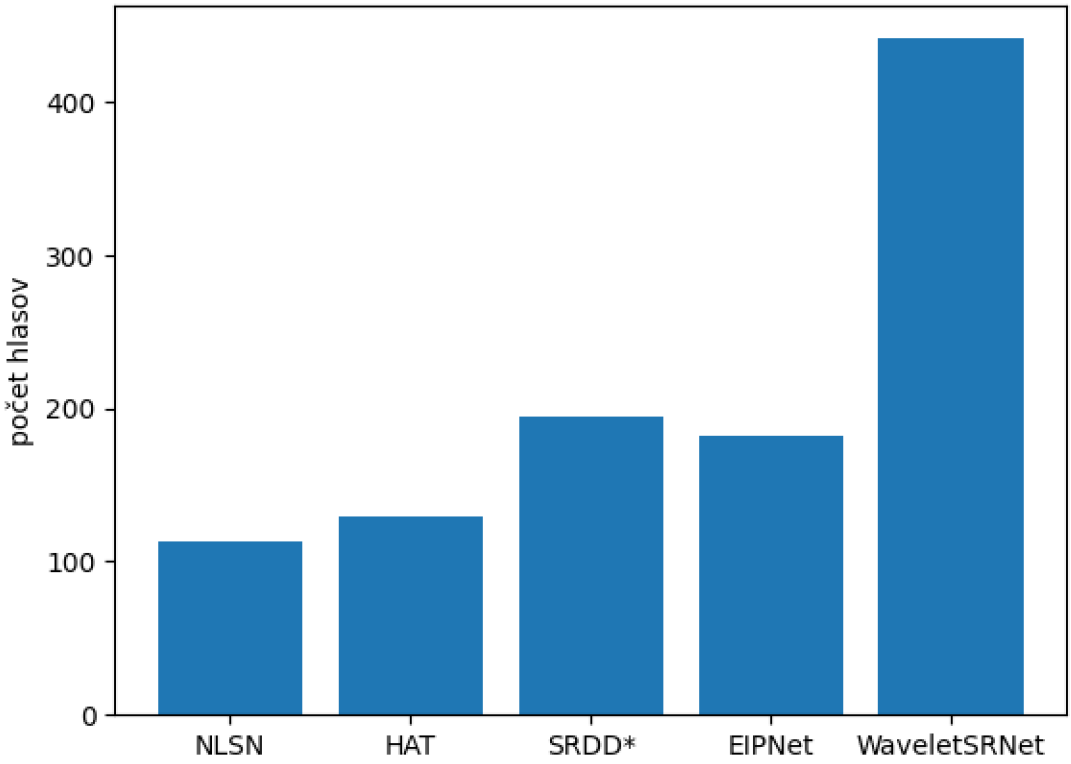
Obr. 5.1: Kvalitatívne porovnanie výsledkov. V zátvorkách sú PSNR [dB], SSIM a LPIPS nasledovne.

Trénovanie a testovanie technológií prebehlo na `NVIDIA Tesla v100s` grafickej karte. Kvantitatívne výsledky sme merali pomocou troch metrík – PSNR, SSIM a LPIPS.

Kvantitatívne výsledky sa ukázali ako uspokojivé. Všetky siete dosiahli výsledkov podobných s momentálnym stavom vedy a techniky superrozlíšenia. NLSN a HAT dosiahli výborných výsledkov v metrike LPIPS. Čo naznačuje výbornú vizuálnu kvalitu. Tieto siete zároveň dosiahli horších výsledkov v metrikách PSNR a SSIM.

Kvalitatívne hodnotenie vyzerá na prvý pohľad ako dostatočné pri väčšine prípadov, pre lepšie hodnotenie sme zostrojili dotazník 5.2. Môžeme predpokladať, že kvalitatívne hodnotenie snímok bude mať väčší význam ako kvantitatívne, vzhľadom na zameranie práce. [37]

5.1 Dotazník



Obr. 5.2: Výsledky dotazníku

Dotazník sme vytvorili náhodným výberom snímok z testovacej časti množiny CelebA. Respondenti v dotazníku porovnávali vizuálnu podobnosť snímok z jednotlivých technológii s pôvodnou $\widehat{HR}$ snímkou a následne zvolili tú, ktorá im prišla najpodobnejšia. Snímok sme zvolili iba 10, keďže sme predpokladali, že veľký počet snímok by mohol respondentov odradiť. Dotazník bol vyplnený 106 respondentmi (t.j. pri 10 snímkoch bolo v dotazníku dokopy 1060 hlasov).

Dotazníku jasne dominovala WaveletSRNet (441 hlasov). Za ňou sa umiestnila naša technoogia SRDD* (194 hlasov). EIPNet mala o 12 hlasov menej (182 hlasov). Na posledných dvoch miestach sa umiestnili HAT (130 hlasov) a NLSN (113 hlasov). Predpokladáme, že HAT a NLSN skončili na posledných miestach, keďže nedokázali dobre zachýtavať odtiene tváre. Zaujímavé je, že na poslených 2 miestach

sa umiestnili technológie s najlepším LPIPS. čo ukazuje, že aj metriky snažiace sa napodobniť ľudské vnímanie majú svoje nedostatky. K tomuto mohol prispieť aj faktor, že snímky z HAT a NLSN boli čiastočne zašumené.

Experiment ukázal účinnosť vlnovej transformácie v oblasti superrozlíšenia. Na prvých miestach sa umiestnili technológie, ktoré boli schopné zachytávať farby a odtiene pokožky. Preto považujeme tento faktor za veľmi podstatný pri tvorbe nových technologií superrozlíšenia tvárí.

Použitie pozornosti v SRDD* prinieslo len o niečo lepšie výsledky ako pôvodný model. Aplikácia pozornosti na superrozlíšenie je preto problematická. Napriek tomu náš model prekonal všetky ostatné, s výjnimkou WaveletSRNet. Predpokladáme, že pre lepšie výsledky by mohli byť použité komplexinejšie pozornostné bloky.

Taktiež je nutné povedať, že práca sa nevenuje tzv. skutočnému superrozlíšeniu (real super-resolution). Pod týmto pojmom sa rozumie superrozlíšenie z *LR* snímok získaných skutočnou degradáciou kvality (napr. kamera s nízkym rozlíšením, šum, nevhodné osvetlenie, zlé počasie, a pod.), na rozdiel od bežne používaných interpolácií. Na túto problematiku bohužial neexistujú dostatočné dátové sady. Toto môže mať negatívny efekt na aplikovateľnosť daných sietí v praxi. [3]

Záver

V rámci našej práce sme sa venovali technológiam superrozlíšenie a následne sme natrénovali a porovnali 5 rôznych neurálnych sietí. Jednou z týchto sietí bola nami upravená verzia. Všetky siete boli natrénované a podrobne sme ich kvantitatívne aj kvalitatívne porovnali. Výsledky kvantitatívnej analýzy ukázali, že všetky siete dosahujú konkurencieschopné výsledky v porovnaní s modernými štandardmi v tejto oblasti. Teda všetky siete sa ukázali ako účinné metódy pre superrozlíšenie.

Objektívne aj subjektívne výsledky ukázali, že siete dosahujú dobré hodnoty percepčnej kvality, čo naznačuje ich potenciál pre identifikáciu. Výsledky sme porovnali pomocou dotazníka, z ktorého vyplynulo, že vlnová transformácia má výnimočný potenciál v oblasti superrozlíšenia tvarí.

Dotazníkové hodnotenie tiež ukázlo, že zachovanie odtieňov a farieb tváre je dôležité pre subjektívne hodnotenie. V tejto oblasti sa najlepšie výsledky dosiahli pomocou siete s názvom WaveletSRNet. Táto siet dokázala prekonať aj technológie, ktoré dosiahli dobré výsledky z hľadiska objektívnych metrík. Nami upravená sieť dosiahla dobrých výsledkov, porovnateľných z modernými technológiami superrozlíšenia.

Je však potrebné zdôrazniť, že naša práca sa nezaoberala reálnym superrozlíšením (real super-resolution). Toto môže mať významný vplyv na aplikovateľnosť týchto technológií v skutočných prípadoch.

Literatúra

[1] Ahmadi, N.; Akbarizadeh, G. *Iris Recognition System based on Canny and LoG Edge Detection Methods.* Journal of Soft Computing and Decision Support Systems, 2, 2015.

[2] Ali, A. M.; Benjdira, B.; Koubaa, A.; El-Shafai, W.; Khan, Z.; Boulila, W. *Vision Transformers in Image Restoration: A Survey.* Sensors, 23(5), 2023. ISSN 1424-8220. doi:10.3390/s23052385. Dostupné z: <https://www.mdpi.com/1424-8220/23/5/2385>.

[3] Chen, H.; He, X.; Qing, L.; Wu, Y.; Ren, C.; Sheriff, R. E.; Zhu, C. *Real-world single image super-resolution: A brief review.* Information Fusion, 79:124–145, 2022. ISSN 1566-2535. doi:https://doi.org/10.1016/j.inffus.2021.09.005. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1566253521001792>.

[4] Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. *Activating More Pixels in Image Super-Resolution Transformer.* 2023.

[5] Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* CoRR, abs/1810.04805, 2018. Dostupné z: <http://arxiv.org/abs/1810.04805>.

[6] Dong, C.; Loy, C. C.; He, K.; Tang, X. *Image Super-Resolution Using Deep Convolutional Networks.* CoRR, abs/1501.00092, 2015. Dostupné z: <http://arxiv.org/abs/1501.00092>.

[7] Dong, C.; Loy, C. C.; Tang, X. *Accelerating the Super-Resolution Convolutional Neural Network.* CoRR, abs/1608.00367, 2016. Dostupné z: <http://arxiv.org/abs/1608.00367> [cit. 26. november 2022].

[8] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; Houlsby, N. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.* CoRR, abs/2010.11929, 2020. Dostupné z: <https://arxiv.org/abs/2010.11929>.

[9] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. *Generative Adversarial Networks*, 2014. doi:10.48550/ARXIV.1406.2661. Dostupné z: <https://arxiv.org/abs/1406.2661> [cit. 26. november 2022].

[10] Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; Chen, T. *Recent advances in convolutional neural networks.* Pattern Recognition, 77:354–377, 2018. ISSN 0031-3203. doi:https://doi.org/10.1016/j.patcog.2017.10.013. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0031320317304120> [cit. 26. november 2022].

[11] Hsu, C.; Lin, C.; Su, W.; Cheung, G. *SiGAN: Siamese Generative Adversarial Network for Identity-Preserving Face Hallucination.* IEEE Transactions on Image Processing, 28(12):6225–6236, 2019. doi:10.1109/TIP.2019.2924554.

[12] Huang, H.; He, R.; Sun, Z.; Tan, T. *Wavelet-SRNet: A Wavelet-Based CNN for Multi-scale Face Super Resolution.* In 2017 IEEE International Conference on Computer Vision (ICCV), s. 1698–1706. 2017. doi:10.1109/ICCV.2017.187.

[13] Jiang, J.; Wang, C.; Liu, X.; Ma, J. *Deep Learning-based Face Super-resolution: A Survey.* CoRR, abs/2101.03749, 2021. Dostupné z: <https://arxiv.org/abs/2101.03749> [cit. 26. november 2022].

[14] Karras, T.; Laine, S.; Aila, T. *A Style-Based Generator Architecture for Generative Adversarial Networks.* CoRR, abs/1812.04948, 2018. Dostupné z: <http://arxiv.org/abs/1812.04948>.

[15] Kawulok, M.; Benecki, P.; Piechaczek, S.; Hrynczenko, K.; Kostrzewa, D.; Nalepa, J. *Deep Learning for Multiple-Image Super-Resolution.* IEEE Geoscience and Remote Sensing Letters, 17(6):1062–1066, 2020. doi:10.1109/LGRS.2019.2940483.

[16] Kim, J.; Lee, J. K.; Lee, K. M. *Deeply-Recursive Convolutional Network for Image Super-Resolution.* CoRR, abs/1511.04491, 2015. Dostupné z: <http://arxiv.org/abs/1511.04491> [cit. 26. november 2022].

[17] Kim, J.; Li, G.; Yun, I.; Jung, C.; Kim, J. *Edge and Identity Preserving Network for Face Super-Resolution.* CoRR, abs/2008.11977, 2020. Dostupné z: <https://arxiv.org/abs/2008.11977>.

[18] Kligvasser, I.; Rott Shaham, T.; Michaeli, T. *xunit: Learning a spatial activation function for efficient image restoration.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, s. 2433–2442. 2018.

[19] Ko, S.; Dai, B.-R. *Multi-Laplacian GAN with Edge Enhancement for Face Super Resolution.* In 2020 25th International Conference on Pattern Recognition (ICPR), s. 3505–3512. 2021. doi:10.1109/ICPR48806.2021.9412950.

[20] Lai, W.; Huang, J.; Ahuja, N.; Yang, M. *Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution.* CoRR, abs/1704.03915, 2017. Dostupné z: <http://arxiv.org/abs/1704.03915> [cit. 26. november 2022].

[21] Le, V.; Brandt, J.; Lin, Z.; Bourdev, L.; Huang, T. S. *Interactive facial feature localization.* In European conference on computer vision, s. 679–692. Springer, 2012.

[22] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Aitken, A. P.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.* CoRR, abs/1609.04802, 2016. Dostupné z: <http://arxiv.org/abs/1609.04802> [cit. 26. november 2022].

[23] Lee, M. *Multi-image Super-resolution via Quality Map Associated Attention Network*, 2022. doi:10.48550/ARXIV.2202.13124. Dostupné z: <https://arxiv.org/abs/2202.13124>.

[24] Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. *Swinir: Image restoration using swin transformer.* In Proceedings of the IEEE/CVF international conference on computer vision, s. 1833–1844. 2021.

[25] Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M. *Enhanced Deep Residual Networks for Single Image Super-Resolution.* CoRR, abs/1707.02921, 2017. Dostupné z: <http://arxiv.org/abs/1707.02921> [cit. 26. november 2022].

[26] Liu, H.; Zheng, X.; Han, J.; Chu, Y.; Tao, T. *Survey on GAN-based face hallucination with its model development.* IET Image Processing, 13(14):2662–2672, 2019. doi:https://doi.org/10.1049/iet-ipr.2018.6545. Dostupné z: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2018.6545>.

[27] Liu, Z.; Luo, P.; Wang, X.; Tang, X. *Deep Learning Face Attributes in the Wild.* CoRR, abs/1411.7766, 2014. Dostupné z: <http://arxiv.org/abs/1411.7766> [cit. 26. november 2022].

[28] Lu, Z.; Liu, H.; Li, J.; Zhang, L. *Efficient transformer for single image super-resolution.* arXiv preprint arXiv:2108.11084, 2021.

[29] Maeda, S. *Image Super-Resolution with Deep Dictionary.* In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX, s. 464–480. Springer, 2022.

[30] Mei, Y.; Fan, Y.; Zhou, Y. *Image Super-Resolution With Non-Local Sparse Attention.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), s. 3517–3526. 2021.

[31] Mei, Y.; Fan, Y.; Zhou, Y. *Image Super-Resolution with Non-Local Sparse Attention.* In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), s. 3516–3525. 2021. doi:10.1109/CVPR46437.2021.00352.

[32] Menon, S.; Damian, A.; Hu, S.; Ravi, N.; Rudin, C. *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models.* CoRR, abs/2003.03808, 2020. Dostupné z: <https://arxiv.org/abs/2003.03808> [cit. 26. november 2022].

[33] Molini, A. B.; Valsesia, D.; Fracastoro, G.; Magli, E. *DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images.* IEEE Transactions on Geoscience and Remote Sensing, 58(5):3644–3656, 2020. doi:10.1109/tgrs.2019.2959248. Dostupné z: <https://doi.org/10.1109%2Ftgrs.2019.2959248>.

[34] Nalepa, J.; Hrynczenko, K.; Kawulok, M. *Multiple-Image Super-Resolution Using Deep Learning and Statistical Features.* In Andrea Torsello; Luca Rossi; Marcello Pelillo; Battista Biggio; Antonio Robles-Kelly, editoøi, Structural, Syntactic, and Statistical Pattern Recognition, s. 261–271. Cham: Springer International Publishing, 2021. ISBN 978-3-030-73973-7.

[35] Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M. C. H.; Heinrich, M. P.; Misawa, K.; Mori, K.; McDonagh, S. G.; Hammerla, N. Y.; Kainz, B.; Glocker, B.; Rueckert, D. *Attention U-Net: Learning Where to Look for the Pancreas.* CoRR, abs/1804.03999, 2018. Dostupné z: <http://arxiv.org/abs/1804.03999>.

[36] Parkhi, O. M.; Vedaldi, A.; Zisserman, A. *Deep Face Recognition.* In British Machine Vision Conference. 2015.

[37] Rajnoha, M.; Mezina, A.; Burget, R. *Multi-Frame Labeled Faces Database: Towards Face Super-Resolution from Realistic Video Sequences.* Applied Sciences, 10(20), 2020. ISSN 2076-3417. doi:10.3390/app10207213. Dostupné z: <https://www.mdpi.com/2076-3417/10/20/7213> [cit. 26. november 2022].

[38] Salvetti, F.; Mazzia, V.; Khaliq, A.; Chiaberge, M. *Multi-Image Super Resolution of Remotely Sensed Images Using Residual Attention Deep*

Neural Networks. Remote Sensing, 12(14), 2020. ISSN 2072-4292. doi: 10.3390/rs12142207. Dostupné z: <https://www.mdpi.com/2072-4292/12/14/2207> [cit. 26. november 2022].

[39] Salvetti, F.; Mazzia, V.; Khaliq, A.; Chiaberge, M. *Multi-Image Super Resolution of Remotely Sensed Images Using Residual Attention Deep Neural Networks*. Remote Sensing, 12(14):2207, 2020. doi:10.3390/rs12142207. Dostupné z: <https://doi.org/10.3390%2Frs12142207>.

[40] Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; Wang, Z. *Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network*. CoRR, abs/1609.05158, 2016. Dostupné z: <http://arxiv.org/abs/1609.05158> [cit. 26. november 2022].

[41] Simonyan, K.; Zisserman, A. *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.

[42] Sobel, I.; Feldman, G. *An Isotropic 3x3 Image Gradient Operator*. 2015. doi:10.13140/RG.2.1.1912.4965.

[43] Tai, Y.; Yang, J.; Liu, X. *Image Super-Resolution via Deep Recursive Residual Network*. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), s. 2790–2798. 2017. doi:10.1109/CVPR.2017.298.

[44] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. *Attention Is All You Need*. CoRR, abs/1706.03762, 2017. Dostupné z: <http://arxiv.org/abs/1706.03762>.

[45] Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C. C.; Qiao, Y.; Tang, X. *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. CoRR, abs/1809.00219, 2018. Dostupné z: <http://arxiv.org/abs/1809.00219> [cit. 26. november 2022].

[46] Wang, Z.; Chen, J.; Hoi, S. C. *Deep learning for image super-resolution: A survey*. IEEE transactions on pattern analysis and machine intelligence, 43(10):3365–3387, 2020.

[47] Zhang, K.; Zhang, Z.; Cheng, C.; Hsu, W. H.; Qiao, Y.; Liu, W.; Zhang, T. *Super-Identity Convolutional Neural Network for Face Hallucination*. CoRR, abs/1811.02328, 2018. Dostupné z: <http://arxiv.org/abs/1811.02328> [cit. 26. november 2022].

[48] Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; Wang, O. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric.* CoRR, abs/1801.03924, 2018. Dostupné z: <http://arxiv.org/abs/1801.03924> [cit. 26. november 2022].

Zoznam symbolov a skratiek

HR	Snímka s vysokým rozlíšením (ground-truth)
LR	Snímka s nízkym rozlíšením
$\widehat{HR}$	Výstup siete
CNN	Konvolučná neurálna sieť (Convolutional Neural Network)
GAN	Generatívna adversaríalna sieť (Generative adversarial network)
ViT	Vidiaci transformer (Vision Transformer)
PSNR	Vrcholový pomer signálu k šumu (Peak signal-to-noise ratio)
SSIM	Meranie indexu štrukturálnej podobnosti (structural similarity index measure)
LPIPS	Naučená percepčná obrázková "patchová" podobnosť (Learned perceptual image patch similarity)
$\mathcal{L}_1$	Stredná absolútna chyba (Mean absolute error)
$\mathcal{L}_2$	Stredná kvadratická chyba (Mean squared error)
ReLU	Lineárny jednotkový usmerňovač (Rectifier linear unit)
xUnit	Aktivačná funkcia xUnit
EIPNet	Sieť zachovávajúca obrys a identitu (Edge and Identity preserving network)
NLSN	Nelokálna riedka sieť (Non-local sparse network)
HAT	Hybridny pozornostný transformer (Hybrid attention transformer)
SRDD	Superrozlíšenia pomocou hlbokého slovníka (super-resolution with deep dictionary)
WaveletSRNet	Vlnová konvolučná neurálna sieť na superrozlíšenie tvarí rôznych škál (A Wavelet-based CNN for Multi-scale Face Super Resolution)