

Zvýšenie rozlíšenia a rekonštrukcia obrazu pomocou generatívnych neurónových sietí

Analýza a návrh riešenia

Bc. Roman Rapco

1Im, 2024 – 2025

1. Úvod

V digitálnom svete, kde kvalita obrazu hrá kľúčovú rolu vo vede, medicíne, bezpečnosti alebo aj vo video-hernom priemysle, sa zvyšovanie rozlíšenia a rekonštrukcia obrazu stávajú čoraz dôležitejšími úlohami. Či už ide o obnovu poškodených historických fotografií, zlepšenie záberov z bezpečnostných kamier alebo zlepšenie kvality snímok (napr. MRI alebo CT), ktoré môže pomôcť pri diagnostike, či pri satelitných snímkach pre detailnejšie analýzy terénu. Schopnosť transformovať nekvalitný alebo rozmazaný obraz na detailnejší a ostrejší je vysoko žiadané.

Tradičné prístupy k super-rezolúcii (SR) sa opierajú o interpoláciu a deterministické modely, ktoré však často nedokážu zachytiť jemné textúry ani realistické detaily. V posledných rokoch sa preto do popredia dostávajú generatívne neurónové siete (GAN - Generative Adversarial Networks), ktoré vďaka svojej schopnosti generovať detailné obrazy prinášajú revolučné výsledky v oblasti generatívneho spracovania obrazu.

Tento dokument sa zameriava hlavne na analýzu problému SR a jeho implementácie, ktoré sa využívajú na zvýšenie rozlíšenia a rekonštrukciu obrazových dát. Cieľom je nielen dosiahnuť vyššiu kvalitu výstupných obrazov z pohľadu tradičných metrick ako PSNR a SSIM, ale aj dosiahnuť vizuálne realistické výsledky, ktoré korešpondujú s ľudským vnímaním detailov a štruktúr.

2. Čo je Image Super-Resolution

Zvyšovanie rozlíšenia obrazu (image super-resolution, SR) je proces zlepšenia kvality obrázku a zvýšenie rozlíšenia z jedného alebo viacerých nízkorozlíšných obrazov (Low resolution - LR) na vysokorozlíšený obraz (High resolution - HR). Táto technika zvyšuje detaily, ostrosť a presnosť obrazu čo je veľmi dôležité v rôznych oblastiach. Ide o inverzný problém, pretože informácie stratené pri redukcii rozlíšenia nie sú priamo dostupné, a preto sa využívajú rôzne matematické modely, štatistické

metódy alebo algoritmy strojového učenia na ich predikciu. Ako už bolo spomenuté v úvode dokumentu, tento proces je obzvlášť dôležitý v aplikáciách, ako je medicínske zobrazovanie, kde zlepšenie kvality snímok (napr. MRI alebo CT) môže pomôcť pri diagnostike, alebo v satelitných snímkach pre detailnejšie analýzy terénu. [1]

Image Super-Resolution vieme klasifikovať viacerými spôsobmi ako môžeme vidieť na obrázku č. 1. Teraz si ich podrobnejšie popíšeme.

Typ vstupného obrazu:

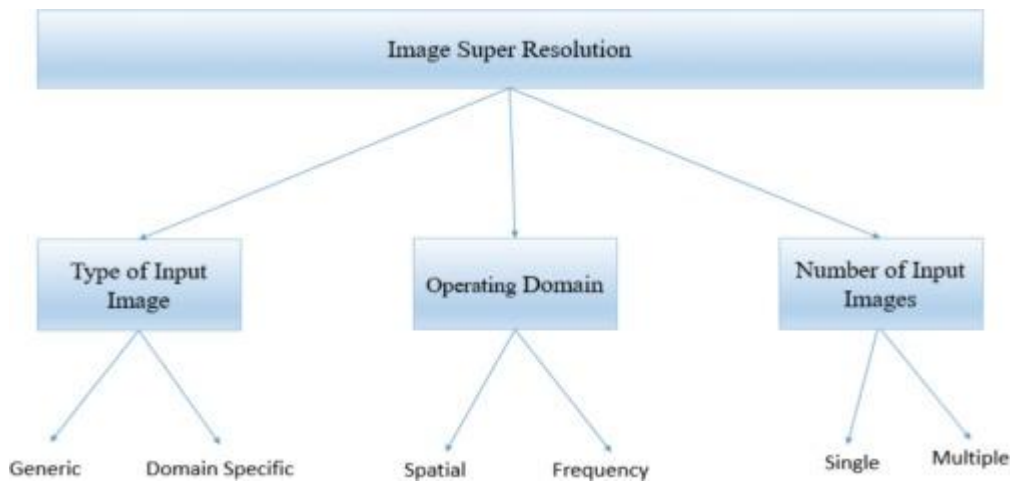
- **Generic:** univerzálne metódy, ktoré fungujú na akomkoľvek type obrazu bez špecifických úprav.
- **Domain specific:** Špecializované metódy prispôbené pre konkrétne oblasti, napr. medicínske alebo satelitné snímky.

Operačná doména:

- **Spatial:** zameriava sa na priestorovú štruktúru obrazu (napr. pixely a ich usporiadanie).
- **Frequency:** pracuje v frekvenčnom priestore (napr. pomocou Fourierovej transformácie) na analýzu a zlepšenie detailov.

Počet vstupných obrazov:

- **Single:** zvyšovanie rozlíšenia z jedného obrazu (Single Image Super-Resolution, SISR).
- **Multiple:** použitie viacerých obrazov (napr. z videa) na získanie lepších výsledkov (Multi-Image Super-Resolution, MISR).



Obr. 1 Klasifikácia Image Super-Resolution [2]

2.1 Single Image Super-Resolution

Zvyšovanie rozlíšenia jedného obrazu (Single Image Super-Resolution, SISR) je špecifický prípad zvyšovania rozlíšenia kde vstupom je iba jeden obrázok s nižším rozlíšením, z ktorého sa snažíme vygenerovať obraz s vyšším rozlíšením. Na rozdiel od multi-image super-resolution (MISR), ktoré využíva viacero snímok (napr. z videa) na získanie dodatočných informácií, SISR pracuje iba s jedným obrázkom čo zvyšuje jeho zložitosť. SISR sa často rieši pomocou hlbokého učenia, napríklad konvolučných neurónových sietí (CNN) alebo generatívnych adversariálnych sietí (GAN), kde sa model naučí predpovedať chýbajúce detaily na základe rôznych tréningových dátových súb, ako napríklad DIV2K. Single Image Super-Resolution (SISR) je dôležitý, pretože je jednou z najpoužívanejších techník a zohráva kľúčovú úlohu v oblasti neurónových sietí a strojového učenia, kde umožňuje zlepšenie kvality obrazov pomocou pokročilých modelov. [3]

2.2 Problematika

Image Super-Resolution je zložitá problematika kvôli podstatnej nejednoznačnosti problému. Z jedného LR obrazu môže existovať viacero možných HR výstupov, pretože stratené detaily (napr. textúry alebo jemné línie) nie sú

jednoznačne určené. Tento problém zhoršuje fakt, že redukcia rozlíšenia zahŕňa stratu informácií, a algoritmus sa snaží nájsť pravdepodobné detaily na základe tréningových dát alebo preddefinovaných modelov. To vedie k výzvam ako zachovať realističnosť a konzistenciu s pôvodným obsahom.

Tu môžeme vidieť vzorec pre degradáciu $I_x = D(I_y) + \sigma$. Degradácia zahŕňa neznáme faktory, ako rozmazanie, downsampling (zmenšenie rozlíšenia), kompresia a šum, ktoré sťažujú presnú inverznú rekonštrukciu. Neurónová sieť musí preto nájsť **inverznú funkciu** degradácie iba na základe párov HR a LR dát, čo zvyšuje komplexnosť úlohy. [1]

Cieľom je teda generovať realistický a vysokokvalitný HR obraz z LR vstupu, ktorý presne zachováva pôvodné detaily, textúry a štruktúru, pričom minimalizuje chyby a artefakty.

3 Prehľad doterajších prístupov

Táto kapitola poskytuje prehľad doterajších metód zvyšovania rozlíšenia obrazu (super-resolution, SR) a metód rekonštrukcie obrazu, ktoré sa vyvinuli od klasických prístupov po moderné techniky založené na hlbokom učení a generatívnych sieťach.

3.1 Klasické prístupy

Klasické tradičné metódy zvyšovania rozlíšenia sa zameriavajú na jednoduché matematické techniky, ktoré nevyžadujú tréning na dátach. Medzi najznámejšie patria napríklad:

- **Nearest neighbour interpolation (Najbližší sused):**

Táto metóda je najjednoduchšia forma interpolácie. Pri zväčšovaní obrazu sa hodnota nového pixelu nastaví na hodnotu najbližšieho existujúceho pixelu z pôvodného obrazu. Výsledkom je rýchly proces, no obraz často vyzerá veľmi pixelizovane, najmä pri vyšších faktoroch zväčšenia, kvôli absencii plynulých prechodov.

- **Bilinear interpolation (Bilineárna interpolácia):**

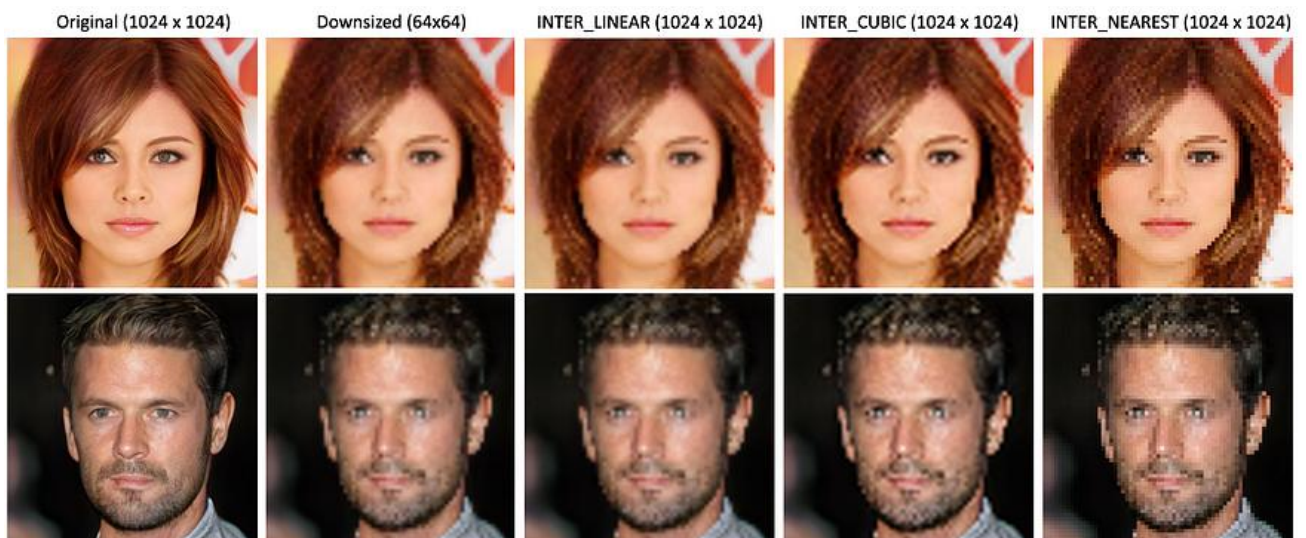
Bilineárna interpolácia používa metódu lineárnej interpolácie. Vypočítava hodnotu nového pixelu ako vážený priemer štyroch najbližších susedných pixelov, pričom váhy závisia od ich vzdialenosti. Táto metóda dosahuje plynulejšie výsledky ako nearest neighbour, no výsledky pre hrany (ostré prechody) nie sú ideálne.

- **Bicubic interpolation (Bikubická interpolácia):**

Bikubická interpolácia rozširuje bilineárnu interpoláciu o použitie 16 najbližších pixelov (4x4 okolie). Vďaka tomu produkuje kvalitnejšie výsledky s menej výrazným rozmazaním v porovnaní s bilineárnou interpoláciou.

- **Lanczos interpolation (Lanczosova interpolácia):**

Lanczosova interpolácia je pokročilejšia technika, ktorá používa sinc funkciu (matematickú funkciu založenú na sinusoide) s orezaním na výpočet hodnôt nových pixelov. Zohľadňuje 8x8 okolie pixelov, čo vedie k ostrým hranám a zachovaniu detailov lepšie ako bicubic interpolácia. Je však výpočtovo náročnejšia a môže zaviesť artefakty v oblastiach s vysokým kontrastom.



Obr. 2 Porovnanie tradičných metód interpolácie [5]

Tieto techniky sú výpočtovo nenáročné, no ich obmedzenia ako napríklad strata detailov a artefakty viedli k vývoju pokročilejších riešení. [5,8]

3.2 Neurónové siete

S nedávnym pokrokom v oblasti umelej inteligencie sa aktívne skúmajú modely super-resolution, ktoré sú založené na hlbokom učení. Existuje viacero open-source modelov, ktoré dosiahli pôsobivé výsledky.

Super Resolution Convolution Neural Networks (**SRCNN**) je jednou z prvých metód hlbokého učenia, ktorá prekonala skoršie tradičné metódy super-resolution. Použitie architektúry CNN umožnilo modelu umelej inteligencie naučiť sa mapovanie medzi obrazmi s nízkym a vysokým rozlíšením. SRCNN má tri konvolučné vrstvy s ktorými berie ako vstup jediný obraz, ktorý je predtým zväčšený pomocou bikubickej interpolácie a následne vylepšuje výsledný obraz pomocou funkcie straty **Mean Square Error** (MSE). Od vzniku SRCNN (2014) vyvinulo viacero výskumníkov ďalšie modely super-resolution založené na hlbokom učení, z ktorých niektoré riešia nedostatky SRCNN. Medzi tieto patria FSRCNN, VDSR, ESDR, PixelCNN a SRResNet. [5]

VDSR (Very Deep Super-Resolution) a EDSR (Enhanced Deep Super-Resolution) predstavili hlbšie siete s reziduálnymi spojmi.

VDSR zaviedlo hlbšiu architektúru s 20 konvolučnými vrstvami, čo bolo výrazným pokrokom oproti SRCNN s tromi vrstvami. Použitie reziduálnych spojení umožnilo sieti učiť sa reziduálne mapovanie, zlepšujúc presnosť pri korekcii rozdielov medzi bikubicky interpolovaným LR a cieľovým HR obrazom. Stabilizáciu tréningu zabezpečoval gradient clipping, a model bol trénovaný na dátach ako ImageNet. Napriek tomu VDSR produkuje rozmazané výstupy, najmä pri vysokých faktoroch zväčšenia (4x, 8x), kvôli závislosti na MSE strate. [6]

EDSR rozšírilo hlboké siete pridaním reziduálnych blokov a odstránením batch normalization z VDSR, čím zlepšilo tok informácií a stabilitu tréningu, najmä pri sieťach s až 32 vrstvami. Optimalizované pre dataset DIV2K, dosiahlo lepšie výsledky v PSNR a SSIM a znížilo výpočtovú záťaž, no stále trpí rozmazaním pri veľkých zväčšeniach. [7]

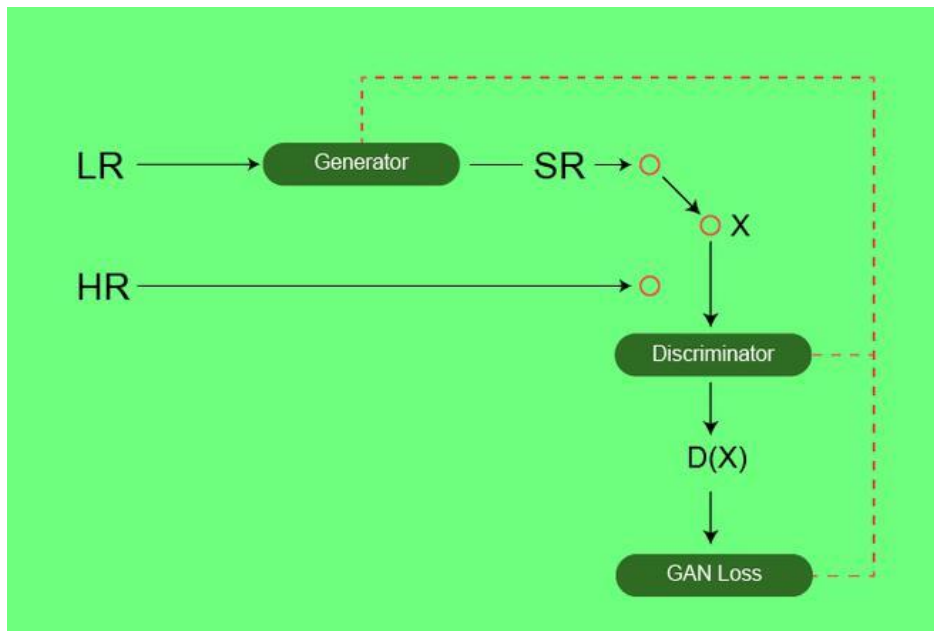
Klasické CNN modely, ako SRCNN, EDSR alebo VDSR, produkujú často príliš **hladké** výsledky, ktoré strácajú vysokofrekvenčné detaily ako napríklad textúry. Tento

problém vyplýva z použitia pixelových strát (L1, L2), ktoré uprednostňujú priemerné hodnoty pred realistickosťou obrazu.

3.3 GAN prístupy

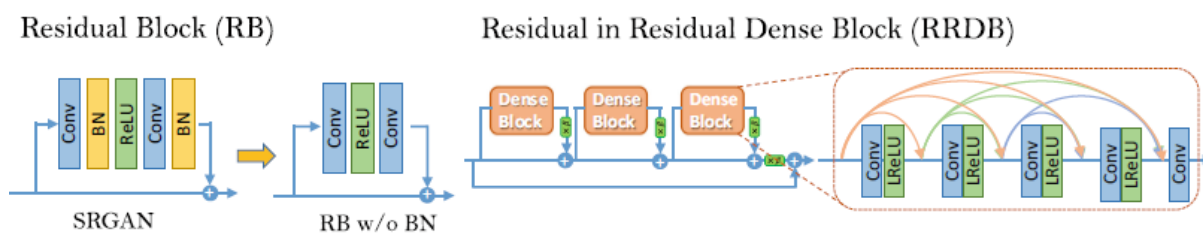
Generatívne adversariálne siete (GAN) priniesli zlepšenie v kvalite SR výstupov tým, že sa začali zameriavať na perceptuálnu kvalitu namiesto iba štatistickej presnosti. Prvá GAN sieť bola predstavená v roku 2014 Ianom Goodfellowom a jeho tímom [9], čím sa položili základy pre revolučný prístup v oblasti hlbokého učenia. Tento koncept rýchlo ovplyvnil rôzne domény počítačového videnia, vrátane zvyšovania rozlíšenia obrazu (Image Super-Resolution), čím nastolili nový štandard v tejto oblasti. GAN sieť sa skladá z generátora a diskriminátora, kde generátor generuje dáta a diskriminátor sa snaží uhádnuť či sú dáta vygenerované alebo reálne z datasetu. Existuje veľa typov GAN sietí, ako napríklad Vanilla GAN, Conditional GAN, Deep convolutional GAN, Self-attention GAN, Laplacian pyramid GAN, StyleGAN, CycleGAN, alebo DiscoGAN, avšak GAN (SRGAN), ESRGAN a Real-ESRGAN sa považujú za kľúčové míľniky v oblasti zvyšovania rozlíšenia obrazu (Image Super-Resolution) a patria medzi najčastejšie používané modely.

SRGAN [10] (Super-Resolution Generative Adversarial Network) bola prvá GAN sieť pre zvýšenie rozlíšenia od Lediga et al. (2017), ktorý skombinoval generátor (založený na hlbokoj CNN) a diskriminátor na produkciu realistických HR obrazov. Použitie percepčnej straty a adversariálneho tréningu umožnilo lepšie zachytenie detailov, avšak výsledky niekedy obsahovali artefakty. Adversariálna strata (adversarial loss) núti riešenie približovať sa prirodzenému obrazovému priestoru pomocou diskriminátorovej siete, ktorá je trénovaná na rozlišovanie medzi SR obrázkami a pôvodnými reálnymi obrázkami.



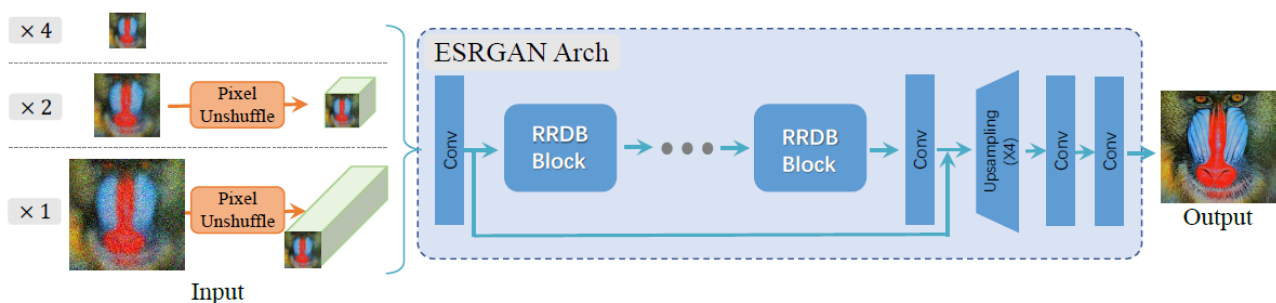
Obr. 3 SRGAN architektúra [11]

ESRGAN (Enhanced SRGAN) [12] zlepšil predchádzajúci model pridaním reziduálnych v dvojitych blokoch (RRDB) a odstránením normalizácie v generátore. RRDB kombinujú viac reziduálnych spojení a husté prepojenia v rámci blokov. Odstránenie vrstvy batch normalization prítomné v SRGAN, zlepšilo tok informácií a stabilitu tréningu siete. Tento prístup zvyšuje perceptuálnu kvalitu a redukuje artefakty, čím zvyšuje kvalitu obrazu. Sieť je populárna najmä v praxi (napr. v AI upscale pre videohry, anime).



Obr. 4 Architektúra ESRGAN siete [12]

Real-ESRGAN [13] je moderná sieť, ktorá je rozšírená pre reálne dáta. Zohľadňuje reálne degradácie (napr. šum, rozmazanie, kompresiu) namiesto idealizovaných modelov. Real-ESRGAN používa vylepšený U-Net diskriminátor s spektrálnou normalizáciou (spectral normalization), ktorý poskytuje väčšiu diskriminačnú silu a stabilnejší tréning v porovnaní so štandardným diskriminátorom ESRGAN. Tento dizajn zlepšuje odhaľovanie rozdielov medzi reálnymi a generovanými obrazmi, najmä pri komplexných degradáciách. Je tu taktiež zavedený pixel-unshuffle ako časť svojej architektúry generátora. Táto technika, ktorá je opakom pixel shuffle, znižuje priestorové rozmery obrazu a zvyšuje počet kanálov na vstupe do hlavnej siete. To zefektívňuje spracovanie a umožňuje lepšie zachytiť detaily pred samotným upsamplingom, čím sa líši od **ESRGAN**.



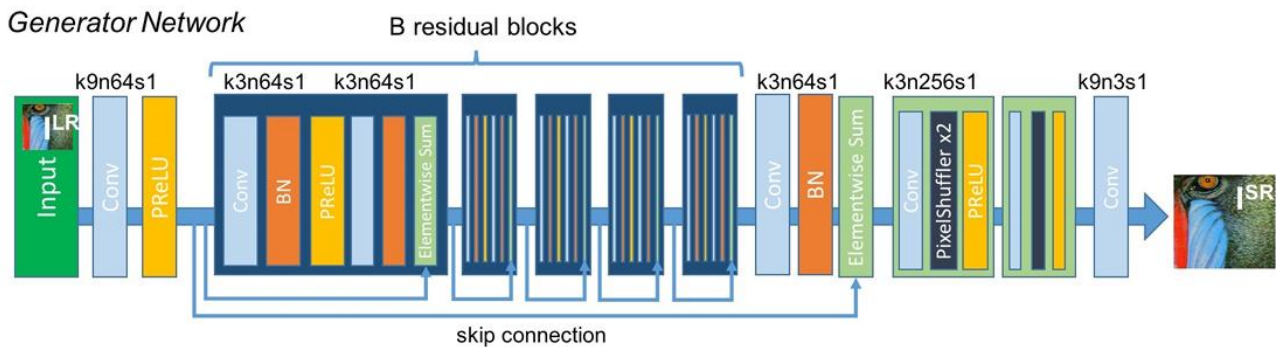
Obr. 5 Real-ESRGAN architektúra generátora a pixel unshuffle. [13]

4 Popis architektúry SRGAN

V tejto časti si bližšie popíšeme architektúru SRGAN (Super-Resolution Generative Adversarial Network), ktorá bola po prvý krát navrhnutá výskumníkmi zo spoločnosti Twitter (teraz X). Architektúra SRGAN je podobná GAN sieti. Skladá sa z generátora a diskriminátora. Generátor vytvára dáta na základe pravdepodobnostného rozdelenia a diskriminátor sa snaží určiť, či dáta pochádzajú zo vstupného datasetu alebo ich vytvoril generátor. Generátor sa následne snaží optimalizovať generované dáta tak, aby oklamal diskriminátor. [11]

4.1 Generátor

Architektúra generátora sa skladá z reziduálnej siete namiesto hlbokých konvolučných sietí (DEEP CNN), pretože reziduálne siete sa trénujú oveľa ľahšie a umožňujú vytvárať hlbšiu architektúru, ktorá vedie k lepším výsledkom. Reziduálne siete používajú skip connections (preskokové prepojenia), čo sú špeciálny typ prepojení, vďaka ktorým sieť produkuje lepšie výsledky.



Obr. 6 Architektúra generátora [10]

Sieť obsahuje B reziduálnych blokov (16), prebrané z ResNet. V reziduálnom bloku sa nachádzajú dve konvolučné vrstvy s malými 3x3 filtrami a 64 feature mapami. Potom za nimi nasleduje batch-normalizácia a ako aktivačná funkcia sa používa Parametric ReLU (PReLU).

Na rozdiel od klasických konvolučných sietí, SRGAN zvyšuje rozlíšenie obrazu pomocou dvoch sub-pixel konvolučných vrstiev (3x3, 256 filtrov) s PixelShuffle technikou zväčšujú rozlíšenie (napr. 2x alebo 4x).

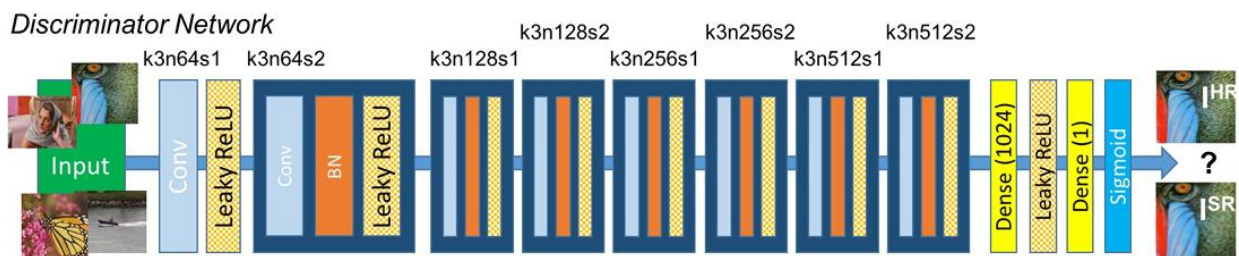
Použitie Parametric ReLU namiesto napr. LeakyReLU umožňuje, aby sieť sama naučila hodnotu parametra (alfa), čím sa zlepšuje presnosť modelu bez výrazného zvýšenia výpočtovej náročnosti.

Počas trénovania sa obrázok vo vysokom rozlíšení (HR) najprv zmenší na nízke rozlíšenie (LR). Generátor sa potom snaží túto LR verziu previesť späť do SR. Výsledný obrázok sa následne odovzdá diskriminátoru, ktorý sa pokúša rozlíšiť, či ide o reálny HR obrázok z datasetu alebo výstup zo SRGAN. Na základe toho vzniká adversariálna

strata (adversarial loss), ktorá sa spätne šíri do generátora a pomáha ho vylepšiť tak, aby vytváral realistickejšie výstupy.

4.2 Diskriminátor

Úlohou diskriminátora je rozlišovať medzi skutočnými vysokorozlíšenými (HR) obrázkami a generovanými superrozlíšenými (SR) obrázkami. Architektúra diskriminátora použitá v tejto práci je podobná architektúre DCGAN [14] a ako aktivačná funkcia sa používa LeakyReLU.



Obr. 6 Architektúra diskriminátora [10]

Sieť sa skladá z osem konvolučných vrstiev s 3x3 filtermi, pričom počet filtrov sa postupne dvojnásobne zvyšuje od 64 až po 512. Na znižovanie rozlíšenia obrazu sa používajú stride konvolúcie vždy, keď sa zdvojnásobuje počet výstupných máp (angl. feature maps). Po konvolučných vrstvách nasledujú dve husté (dense) vrstvy, medzi ktorými je aplikovaná funkcia LeakyReLU. Na výstupe je použitá sigmoidová aktivačná funkcia, ktorá dáva výslednú pravdepodobnosť toho, či je vstupný obrázok reálny (z datasetu HR) alebo generovaný (SR).

4.3 Tréningový proces

SRGAN používa tzv. percepčnú stratu (LSR – perceptual loss), ktorá je váženým súčtom dvoch stratových zložiek a to: obsahovej straty (content loss) a adversariálnej straty (adversarial loss). Vďaka kombinácii týchto dvoch stratových zložiek generátor

zlepšuje jeho schopnosti vytvárať vizuálne realistické výstupy. Percepčná stratová funkcia je súčet obsahovej straty a adversariálnej, môžeme ho vidieť na obrázku 7.

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3}l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

Obr. 7 Percepčná stratová funkcia [10]

4.3.1 Obsahová strata

MSE (mean square error) strata na úrovni pixelov je najčastejšie používaný optimalizačný cieľ pri úlohách super-resolution obrázkov, na ktorom sú založené mnohé najmodernejšie prístupy, používa sa napríklad pri architektúre SRResNet. Hoci takýto prístup dosahuje vysoké hodnoty PSNR (Peak signal-to-noise ratio) riešeniam optimalizovaným pomocou MSE často chýba vysokofrekvenčný obsah, čo vedie k vizuálne nepekným výsledkom s príliš hladkými textúrami. Namiesto klasických pixelových strát autori použili stratovú funkciu bližšiu percepčnej podobnosti, presnejšie VGG stratu na základe ReLU aktivačných vrstiev predtrénovanej 19-vrstvovej siete VGG.

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2$$

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

Obr. 8 Vzorce pre obsahovú stratu vo SRGAN [10]

4.3.2 Adversariálna strata

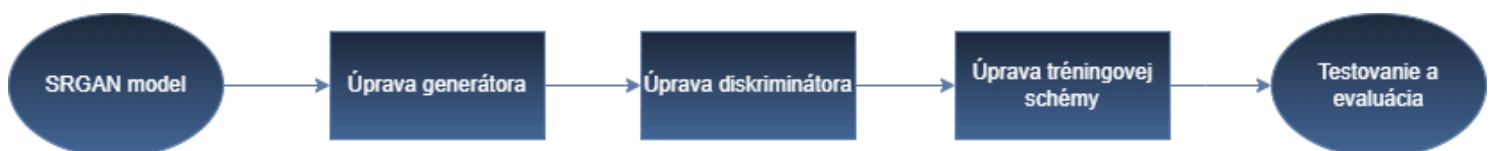
Adversariálna strata je stratová funkcia, ktorá núti generátor vytvárať obrázky podobnejšie skutočným vysoko-rozlíšeným obrázkom. Robí to prostredníctvom diskriminátora, ktorý je trénovaný na rozlišovanie medzi reálnymi HR obrázkami a generovanými SR výstupmi.

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

Obr. 9 Vzorec pre adversariálnu stratu

5 Návrh riešenia

V tejto časti sa budeme zaoberať návrhom riešenia a jeho možnou implementáciou. Návrh riešenia pozostáva z výberu GAN modelu na super-rezolúciu (SR), úprava jeho generátora, diskriminátora, úprava tréningovej schémy a nakoniec testovanie a evaluácia daného modelu oproti originálu.



5.1 SRGAN model

Ako východiskový bod pre implementáciu a experimentovanie sme zvolili verejne dostupnú implementáciu SRGAN modelu: <https://github.com/tensorlayer/srgan>.

Tento model bol navrhnutý v práci Ledig et al. (2017) ako prvý, ktorý demonštroval schopnosť GAN vytvárať fotorealistické obrázky s vysokým rozlíšením.

Výhodou tohto prístupu je existencia fungujúceho kódu, ktorý umožňuje analyzovať, upravovať a rozširovať model podľa vlastného návrhu. Zvolená implementácia využíva framework TensorFlow a je dostatočne modulárna na to, aby umožnila zásahy do architektúry siete.

Ako už bolo napísané, SRGAN sa skladá z dvoch častí: generátor a diskriminátor. Generátor je tvorený konvolučnými vrstvami, reziduálnymi blokmi a sub-pixel konvolúciami (pixel shuffle) pre zväčšenie rozlíšenia. Diskriminátor je klasická CNN, ktorá klasifikuje obrázky ako reálne alebo generované.

5.2 Úprava generátora

Jednou z hlavných oblastí úprav je generátor, ktorý tvorí jadro SRGAN siete. Jedna z možností by bolo pridať ďalšie reziduálne bloky, čím by sa zvýšila kapacita siete na zachytávanie detailov, upraviť ich vnútornú štruktúru, napríklad zmenou typu aktivačnej funkcie (napr. z ReLU na Leaky ReLU) alebo optimalizáciou skip connections pre efektívnejší tok gradientov. Ďalšou možnosťou je experimentovanie s inými formami upsamplingu, ako napríklad transponované konvolúcie [15] namiesto tradičného pixel shuffle, čo by mohlo zlepšiť plynulosť prechodov a redukovať artefakty. Okrem toho vylepšenie inicializácie váh, alebo prispôsobenie vrstvenia siete môže prispieť k stabilnejšiemu a efektívnejšiemu tréningu, čím sa zvýši celková kvalita výstupov.

5.3 Úprava diskriminátora

Diskriminátor, ktorý hodnotí realističnosť generovaných obrazov, je možné taktiež upraviť. Existuje možnosť zavedenia PatchGAN diskriminátora, ktorý posudzuje realističnosť v lokálnych oblastiach namiesto celého obrazu, by mohlo viesť k detailnejšiemu zachyteniu štruktúr a textúr. Zmena veľkosti receptive field, napríklad pridaním väčších konvolučných jadier, alebo rozšírenie siete o hlbšie vrstvy umožní diskriminátoru lepšie rozpoznať komplexnejšie črty, čím salepší jeho schopnosť odlíšiť reálne HR obrázky od generovaných (SR).

5.4 Úprava tréningovej schémy

Tréningový proces SRGAN môže byť optimalizovaný úpravou tréningovej schémy. Pôvodný SRGAN namiesto klasických pixelových strát, ako napríklad MSE, využíva stratovú funkciu založenú na percepčnej podobnosti, konkrétne VGG stratu odvodenú z ReLU aktivačných vrstiev predtrénovanej 19-vrstvovej siete VGG. Táto voľba umožnila lepšie zachytenie detailov a textúr, čím sa posunula kvalita generovaných obrazov nad rámec štatistickej presnosti. S ohľadom na túto inováciu by sme vedeli navrhnúť ďalšie optimalizácie tréningovej schémy. Okrem toho môžeme experimentovať s alternatívnymi loss funkciami, ktoré by mohli nahradiť alebo doplniť

existujúce straty a poskytnúť robustnejší prístup k dátam, čím by sa zvýšila odolnosť voči anomáliám.

5.5 Testovanie a evaluácia dosiahnutých výsledkov

Všetky spomenuté úpravy otvárajú priestor pre experimentálne zlepšenie SRGAN siete, pričom každá zmena môže priniesť výhody, ako zníženie artefaktov, zlepšenie detailov alebo stabilnejší tréning. Ich kombinácia a testovanie na rôznych dátových súboroch, ako DIV2K, by mohlo posunúť výkon modelu nad rámec pôvodného návrhu od Lediga et al. (2017).

Po implementácii a úpravách pôvodnej SRGAN siete s našimi modifikáciami je dôležité porovnať jej výkon s originálnym modelom, aby sme objektívne zhodnotili, či sme dosiahli zlepšenie, alebo naopak, kde sa môžu skrývať nedostatky.

Na kvantitatívne hodnotenie vieme použiť štandardné metriky PSNR (Peak Signal-to-Noise Ratio) a SSIM (Structural Similarity Index), ktoré sú bežne využívané pri hodnotení kvality zvyšovania rozlíšenia. PSNR nám ukáže, ako veľmi sa generovaný vysokorozlíškový obraz líši na úrovni pixelov od originálu, pričom vyššie hodnoty naznačujú lepšiu presnosť. SSIM zase hodnotí podobnosť v štruktúre, vrátane jas, kontrastu a textúr, čo viac korešponduje s tým, ako ľudské oko vníma kvalitu obrazu. Porovnanie týchto metrik na rovnakej dátovej sade, napríklad DIV2K, nám umožní posúdiť, či zmeny v generátore (ako napríklad pridanie reziduálnych blokov či úprava upsamplingu) a diskriminátore (napríklad využitie PatchGAN) priniesli očakávané zlepšenia.

Ďalej môžeme využiť metriku LPIPS (Learned Perceptual Image Patch Similarity), ktorá vychádza z predtrénovaných sietí (napr. VGG) a hodnotí percepčnú podobnosť obrázkov. Táto metrika je vhodná na posúdenie, či zmeny stratovej funkcie pomáhajú vytvárať vizuálne realistickejšie výsledky, bližšie k tomu, čo vníma človek. Nižšia hodnota LPIPS by v tomto prípade znamenala lepší výsledok oproti originálnemu SRGAN.

Kvantitatívne výsledky budú doplnené vizuálnou kontrolou generovaných obrázkov, s dôrazom na detaily, ako sú ostré hrany, textúry a zložité vzory. Pri tejto analýze zhodnotíme prípadné artefakty, mätúce efekty alebo rozmazanie, ktoré sa môžu pri SRGAN objaviť. Môžeme tiež využiť prieskum, ktorý pomôže zachytiť subjektívny dojem z kvality, keďže GAN siete sa zameriavajú na to, aby to bolo čo najviac vizuálne dobre vyzerajúce pre ľudské oko. Pre lepšiu prehľadnosť výsledky vieme doplniť o grafy a príklady obrázkov, aby bolo jasné, či modifikácie priniesli viditeľné zlepšenie.

Testovanie môže prebehnúť na rovnakej dátovej sade a za rovnakých podmienok, teda s rovnakými parametrami, ako sú learning rate (učiaci pomer) a počet epoch, aby sme zabezpečili porovnanie. Výsledky vieme zhrnúť v tabuľkách a vizualizovať, kde porovnáme priemerné hodnoty PSNR, SSIM a LPIPS pre oba modely na viacerých obrázkoch. Takýto komplexný prístup nám umožní porovnať pokrok, ale aj identifikovať oblasti, v ktorých je možné ešte metódu vylepšiť.

6 Záver

Zvýšenie rozlíšenia a rekonštrukcia obrazu patria medzi náročné, no zároveň veľmi perspektívne úlohy v oblasti počítačového videnia. V tomto dokumente sme sa zamerali na analýzu problému super-rezolúcie (SR) a návrh možného riešenia. V úvodných častiach sme si priblížili, čo super-rezolúcia znamená, prečo je výpočtovo aj prakticky náročná a aké výzvy sú s ňou spojené. Následne sme prešli prehľad doterajších prístupov, od klasických interpolácií, cez konvolučné neurónové siete až po pokročilejšie generatívne modely, ako sú GAN siete. V návrhu riešenia sme sa podrobnejšie venovali výberu konkrétneho modelu, jeho úprave, ako aj spôsobu tréningu a vyhodnocovania výstupov.

V aktuálnej fáze práce sme si definovali návrh riešenia, vrátane architektúry siete a spôsobu testovania. Ďalšie kroky budú zamerané na dôkladnejšiu analýzu siete, ladenie parametrov, zlepšenie výstupnej kvality a prípadné rozšírenie modelu o nové prvky. Cieľom bude dosiahnuť lepšiu vizuálnu kvalitu výstupov a overiť, či navrhnuté úpravy prinášajú zmysluplné zlepšenie oproti existujúcim riešeniam.

Použitá literatúra

1. A Complete Guide to Image Super-Resolution in Deep Learning and AI. [online] Dostupné z: <https://www.digitalocean.com/community/tutorials/image-super-resolution>
2. Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, Ray E. Sheriff, Ce Zhu, Real-world single image super-resolution: A brief review, Information Fusion, Volume 79, 2022, Pages 124-145, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2021.09.005> . <https://www.sciencedirect.com/science/article/pii/S1566253521001792>
3. D. Glasner, S. Bagon and M. Irani, "Super-resolution from a single image," 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 349-356, doi: 10.1109/ICCV.2009.5459271.
4. Sung Cheol Park, Min Kyu Park and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," in IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36, May 2003, doi: 10.1109/MSP.2003.1203207.
5. Ong Si Ci, Image Super Resolution: A Comparison between Interpolation & Deep Learning-based Techniques to Improve Clarity of Low-Resolution Images [online] Dostupné z: <https://medium.com/htx-s-s-coe/image-super-resolution-a-comparison-between-interpolation-deep-learning-based-techniques-to-25e7531ab207>
6. B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 1132-1140, doi: 10.1109/CVPRW.2017.151.

7. J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.
8. Maganti Jahnvi, D. Rajeswara Rao, Amballa Sujatha, A Comparative Study Of Super-Resolution Interpolation Techniques: Insights For Selecting The Most Appropriate Method, Procedia Computer Science, Volume 233, 2024, Pages 504-517, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2024.03.240>.
9. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. Commun. ACM 63, 11 (November 2020), 139–144. <https://doi.org/10.1145/3422622>
10. C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.
11. Super Resolution GAN (SRGAN) [online]. Dostupné z: <https://www.geeksforgeeks.org/machine-learning/super-resolution-gan-srgan/>
12. Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European conference on computer vision (ECCV) workshops. 2018.
13. X. Wang, L. Xie, C. Dong and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 1905-1914, doi: 10.1109/ICCVW54120.2021.00217.
14. M. Mehralian and B. Karasfi, "RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks," 2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium, Kish Island, Iran, 2018, pp. 31-38, doi: 10.1109/AIAR.2018.8769811.
15. What is Transposed Convolutional Layer? [online]. Dostupné z: <https://www.geeksforgeeks.org/machine-learning/what-is-transposed-convolutional-layer/>