

# SRGAN for enhancing melanoma images to improve recognition

**Martin Gano**

Faculty of Information Technology  
Brno University of Technology  
Email: xganom00@stud.fit.vutbr.cz

**Marek Hrivnak**

Faculty of Information Technology  
Brno University of Technology  
Email: xhrivn02@stud.fit.vutbr.cz

**Jozef Horvath**

Faculty of Information Technology  
Brno University of Technology  
Email: xhorva11@stud.fit.vutbr.cz

*Rozpoznanie malígneho melanómu pomocou vizuálneho zhodnotenia sa čoraz viac stáva úlohou vhodnou pre strojové učenie. Keď zanedbáme otázku zodpovednosti a morálnu stránku veci, neurónové siete dokážu na základe dostatočne presného obrázka dosiahnuť v tejto úlohe vyššiu presnosť ako špecializovaní lekári. Problém nastáva pri využití menej kvalitných fotografií, ktoré sú vykonané napríklad mobilným telefónom a odoslané na automatickú detekciu. V prípade nižšieho rozlíšenia je nutné vhodným spôsobom rekonštruovať jemné štruktúry, ktoré sú kľúčové v klasifikácii kožných lézií na malígne a benígne. Metóda, ktorú využívame je schopná rekonštruovať jemné textúry aj za cenu nižšieho dosiahnutého PSNR, čo vo výsledku pomôže zlepšiť presnosť klasifikácie na menších dátach.*

## 1 Introduction

Hlavným problémom ktorý práca rieši je nízka kvalita fotografií určených pre detekciu malígneho melanómu. Najlepšie systémy pre detekciu malígneho melanómu publikované na stránke organizácie *Kaggle* dosahujú úspešnosť okolo 95% pre obrázky vo veľmi dobrej kvalite s veľkosťou 224x224. Pri obrázkoch s menším rozlíšením je výsledná úspešnosť výrazne menšia. V tejto práci sme si ako meradlo úspešnosti vybrali neurónovú sieť známu ako *EfficientNet*, pretože porovnateľnú presnosť ako state-of-the-art architektúry a zároveň je 8.4-krát menšia a 6.1-krát rýchlejšia ako napr. neurónová sieť s architektúrou *ImageNet* (2). Model sme natrénovali na datasete poskytnutom organizáciou *Kaggle*<sup>1</sup>, ktoré boli poskytnuté za účelom natrénovania súťažného modelu. V danej súťaži (ale až po jej ukončení) sa nám podarilo s architektúrou *EfficientNet* získať 90.43%, pričom víťaz dosiahol 94.90%. Dáta využité na tréning mali veľkosť 224x224, model bol natrénovaný pred začiatkom experimentovania so Super Resolution GAN

(SRGAN)(1) a v priebehu experimentovania už neprebiehala žiadne dotrenovávanie, cieľom bolo zvýšiť jeho presnosť len na základe predspracovania dát. Keďže hlavným cieľom projektu bolo vytvoriť odolný systém, ktorý sa správa stabilne bez ohľadu na zdroj a veľkosť dát, pre vyhodnotenie klasifikátora ale aj pre tréning a a meranie úspešnosti sme využili iný dátový zdroj<sup>2</sup> ako pre samotný tréning *EfficientNet* modelu. Kým sme si pripravili vlastnú dátovú sadu obsahujúcu obrázky kožných lézií, vyskúšali sme natrénovať aj model pomocou dátovej sady *DIV2K*<sup>3</sup>. S týmto modelom sme ďalej experimentovali a porovnávali jeho úspešnosť s úspešnosťou modelu natrénovaného pomocou obrázkov kožných lézií.

## 2 Dátové sady

Počas práce na projekte boli využité tri rôzne dátové sady. Prvá dátová sada ktorá bola poskytnutá organizáciou *Kaggle* bola zozbieraná projektom *International Skin Imaging Collaboration (ISIC)* z rôznych kliník zaoberajúcimi sa danou problematikou. Táto dátová sada obsahovala okrem samotného obrázku aj ďalšie informácie o pacientovi, ktoré sme v rámci nášho projektu nevyužili. Obrázky boli vo formáte *TFRecord* alebo *JPEG*, a ich veľkosť je 1024x1024. Z dôvodu nedostatku časových a výpočtových zdrojov sme sa rozhodli zmenšiť veľkosť obrázkov na 224x224 a stratiť tak časť informácie. Tieto dáta boli využité pre natrénovanie klasifikátora malígneho melanómu.

Ďalšou využitou dátovou sadou bola sada *DIVERse 2K resolution (DIV2K) dataset* (3). Táto dátová sada bola pôvodne využitá na natrénovanie prvej verzie SRGAN a základné vyhodnotenie tejto metódy. Dôvodom bola jednoduchosť práce s danou dátovou sadou a zároveň sme v rámci projektu porovnali ako funguje model natrénovaný

<sup>1</sup><https://www.kaggle.com/c/siim-isic-melanoma-classification/data>

<sup>2</sup><https://s3-us-west-1.amazonaws.com/udacity-dlnfd/datasets/skin-cancer/>

<sup>3</sup><https://data.vision.ee.ethz.ch/cvl/DIV2K/>

na špeciálnych dátach pre melanóm v porovnaní s rôznymi obrázkami.

Posledná dátová sada bola zozbieraná v rámci súťaže 2017 ISIC Challenge on Skin Lesion Analysis Towards Melanoma Detection. Podobne ako prvá dátová sada je aj táto sada poskytnutá projektom ISIC. Obsahuje len obrázky kožných lézií a ich označenie. Veľkosť obrázkov je 256x256 a táto dátová sada bude slúžiť na tréning SRGAN a vyhodnotenie klasifikátora.

Pre zväčšenie počtu prvkov trénovacej sady sme sa rozhodli rozložiť jednotlivé obrázky kože na 4 rovnako veľké časti (128x128) a k nim vygenerovať pomocou bikubickej interpolácie 4 krát menšie obrázky. Následne bol vstup pre generatívnu neurónovú sieť obrázok s menším rozlíšením a očakávaný obrázok bol s väčším rozlíšením.

### 3 Super-Resolution GAN

SRGAN v našej práci je natrénovaná na 4-násobné zväčšenie obrázku. Autori metódy predstavili tzv. perceptuálnu chybovú funkciu. Podobne ako navrhol Johnson a spol. (4), aj v našej práci získavame vypočítavame chybu extrakcie aktivácií z predtrénovanej neurónovej siete. V našom prípade sa jedná o VGG19, predtrénované váhy sme prebrali z registra projektu TF, váhy boli predtrénované na dátovej sade ImageNet. Pre generátor sme zvolili ResNet a experimentálne sme ukázali, že je výhodnejšie generátor predtrénovať na vhodnej dátovej sade a pri tréningu SRGAN inicializovať váhy generátora na predtrénované hodnoty.

### 4 Experimenty

Experimentami sme sa snažili potvrdiť hypotézu, že ak sme natrénovali sieť na väčších dátach než sú evaluačné dáta a využijeme sofistikovanú metódu na zväčšanie evaluačných dát, dokážeme zvýšiť presnosť klasifikácie oproti malým obrázkom. Pri experimentoch sme využili natrénovaný model EfficientNet na klasifikáciu a natrénovaný generátor s architektúrou ResNet, ktorý sme natrénovali ako súčasť SRGAN. Pre vyhodnotenie sme napokon použili dátovú sadu ktorá obsahovala 2000 vzoriek z toho 1500 benígnych a 500 malígnych. Pomer sme prebrali z testovacej sady na Kaggle, v skutočných dátach je pomer malígnych vzoriek menší a najpresnejšie (ale zbytočné) modely by mohli klasifikovať všetko ako benígne.

Na obrázku 1 vidíme ako funguje vizuálne porovnanie obrázka s nízkym rozlíšením a super-resolvovaného obrázka. V rámci našej práce sme nevyhodnocovali PSNR oproti iným dátovým sadám ako kožné lézie, autori metódy však upozorňujú na fakt, že metóda nie je najlepšia z pohľadu PSNR. Ako sme už spomenuli, ponúka možnosť rekonštrukcie jemných textúr v obrázku. Napr. na obrázku 1 dosahuje super-resolvovaná verzia PSNR 24.43, pričom pri bikubickej interpolácii je PSNR len 17.95. Na druhej strane metódy ako WDSR(5) alebo EDSR(6) by v podobnom prípade mohli dosiahnuť lepšie výsledky z pohľadu PSNR.

Obrázok 2 ukazuje ako sa líši obrázok vygenerovaný generátorom natrénovaným na špeciálnych dátach od ori-



Obr. 1. Obrázok ukazuje obrázok s nízkym rozlíšením a obrázok vygenerovaný SRGAN natrénovanej na DIV2K



Obr. 2. Obrázok porovnáva obrázok so zníženým rozlíšením, super-resolvovaný obrázok a originálny obrázok rovnako veľký ako SR obrázok

ginálneho obrázku. Môžeme vidieť, že obrázok nie je vyhladený ale zachováva svoje kontúry, to môže potencionálne pomôcť klasifikátoru správne rozhodnúť.

Na záver sme porovnali predspracovanie dát pomocou generátora natrénovanom na všeobecných obrazových dátach DIV2K, a generátor natrénovaný na špeciálnych dátach ISIC. Ukázalo sa, že z pohľadu klasifikácie sa obrázok so zníženou veľkosťou a následne zväčšením pomocou DIV2K generátoru ešte viac poškodil z pohľadu klasifikácie. Po diskusii sme zhodnotili, že obrázok bol síce viac vyhladený a vizuálne ostrejší, avšak dôležité detaily neboli doplnené dostatočne správne. Generátor natrénovaný na DIV2K pravdepodobne nie je schopný doplniť detaily ktoré sú zásadné pri klasifikácii rakoviny kože, na druhej strane sa stretol len s veľmi odlišnými dátami a preto sa to aj dalo očakávať. Odporúčame na ďalšie experimentovanie inicializáciu siete. V prípade, že by sme využili verejne dostupné váhy natrénované na DIV2K alebo ImageNet pre generátor s architektúrou ResNet, boli by sme schopný ušetriť množstvo zdrojov. V našom projekte sme neoverili, či by to bolo aj realizovateľné pri udržaní kvalitného výsledku. Nasledujúca tabuľka ukazuje aké presnosti dosiahol klasifikátor v závislosti od zmenšenia, resp. poučiria SRGAN na predspra-

covanie. LR znamená, že predspracovanie použité nebolo. SR ISIC znamená, že predspracovanie bolo použité pomocou generátora natrenovaného na ISIC dátach a SR DIV2K zasa generátor natrenovaný na DIV2K dátach. Presnosť ktorú sme dosiahli na originálnej dátovej sade s pôvodným rozlíšením je 86.45%.

Zmenšenie obrázka	LR	SR ISIC	SR DIV2K
4x	79.25%	82.91%	62.80%
2x	81.05%	84.22%	72.75%

Tabuľka ukazuje, že využitie SRGAN je prospešné ak je GAN natrenovaná na dátach ISIC, v prípade DIV2K presnosť ešte výrazne zhoršuje.

## 5 Záver

Vránci projektu sme ukázali možnosť predspracovania dát pred klasifikáciou. Výhoda predspracovania je v našom prípade možnosť využiť väčšie dáta, pretože nepotrebujeme anotované dátové sady. Zároveň sme potvrdili, že PSNR nie je jedinou dôležitou metrikou pre vyhodnotenie kvality vylepšovania rozlíšenia. Prínosom do problematiky bola implementácia trénovateľnej siete s možnosťou vykonávania ďalších experimentov. Implementovaný framework obsahuje aj klasifikátor a ponúka možnosť klasifikovať s predspracovaním. Nepodarilo sa nám stihnúť experimentovať s klasifikátormi, chybovými funkciami a nepodarilo sa nám porovnať našu metódu s ďalšími metódami, ktoré nefungujú na princípe GAN.

## Literatúra

- [1] Christian Ledig et al. *Photo-realistic single image super-resolution using a generative adversarial network*. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
- [2] Mingxing Tan and Quoc V. Le *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. International Conference on Machine Learning, 2019.
- [3] Eirikur Agustsson and Radu Timofte *NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study*. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.
- [4] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei *Perceptual losses for real-time style transfer and super-resolution..* European conference on computer vision. Springer, Cham, 2016.
- [5] Jiahui Yu et al. *Wide activation for efficient and accurate image super-resolution..* arXiv preprint arXiv:1808.08718, 2018.
- [6] Lim, Bee, et al. *Enhanced deep residual networks for single image super-resolution..* Proceedings of the IEEE

conference on computer vision and pattern recognition workshops. 2017.