

Semantic image segmentation

Matej Berezný <xberez03@stud.fit.vutbr.cz>
Ondrej Valo <xvalo00@stud.fit.vutbr.cz>

29. prosince 2021

1 Úvod

Cieľom projektu je vytvoriť program pre sémantickú segmentáciu obrázkov. Ako nástroj pre daný účel sme sa rozhodli využiť jednu z mnohých architektúr neurónových sietí kde po konzultácii s Ing. Michal Hradiš Ph.D. bola nakoniec zvolená architektúra U-Net. Dodatočne bolo toto zadanie rozšírené o pridanie rôznych modifikácií pre danú architektúru a jej testovanie pri rôznych hyper-parametroch. Využitá dátová sada je popísaná v sekcii 2, ďalej nasleduje sekcia 3 ktorá popisuje augmentácia dát. Nasledujúca sekcia 4 sa zaoberá samotnou architektúrou a tréningom NN. Na koniec sa nachádzajú sekcie ohľadom vyhodnotenia výsledkov 5 a sekcia záver 6.

2 Dataset

Ako dataset pre účely segmentácie bol zvolený CityScapes dataset [COR⁺16]. Keďže sa jednalo o kategorickú segmentáciu, pôvodných 30 tried pre každý typ objektu na obrázkoch bolo zredukovaných na 7 kategórií bližšie definovaných v tabuľke 1. Takisto boli z dôvodu časových a HW obmedzení pôvodné obrázky 8-násobne zmenšené (2048x1024 -> 256x128).

3 Augmentácia

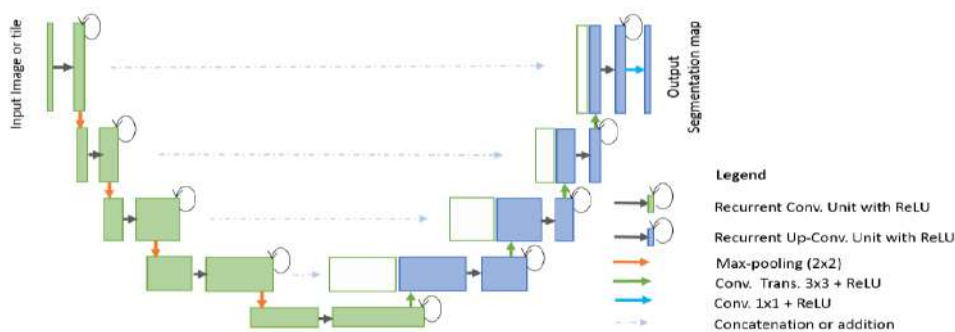
Pre zvýšenie celkovej presnosti neurónovej siete a zároveň zníženie rizika overfittingu bola použitá tzv. "on the fly" augmentácia, čiže každá trénovacia vzorka bola svojim spôsobom unikátna. Pre tieto účely bolo použitých niekoľko augmentačných metód nachádzajúcich sa v `dataset/augmentations.py`. Konkrétne sa jedná o horizontálne otáčanie, elastické transformácie, gamma korekcia a rôzne modifikácie jasu, saturácie a samotných farieb.

4 Trénovanie

Všetky architektúry použité v tomto projekte sa nachádzajú v `model/unet.py` a samotné trénovanie má na starosti `model/trainer.py`. Okrem samotnej architektúry U-Net boli implementované aj jej modifikácie.

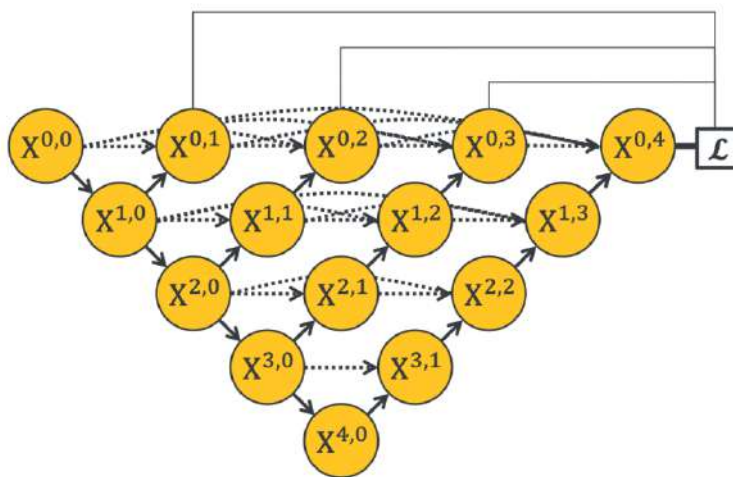
Prvou z nich bola výmena typického `DoubleConvolution` bloku za tzv. `RCNNBlock` [AHY⁺18], ktorý pozostával z jednej 1x1 konvolúcie nasledovanej dvoch rekurentných konvolúcií s 3x3 kernelom.

Ďalšou modifikáciou bolo obohatenie dekodujúcej časti U-Net architektúry o tzv. `AttentionBlock`, ktorý pozostával z dvoch 1x1 konvolúcií bez aktivačnej funkcie a jednej 1x1 konvolúcie s sigmoidou ako aktivačnou funkciou.



Obrázek 1: U-Net s recurrent blokom

Poslednou z modifikácií U-Net architektúry bolo použitie tzv Nested U-Net [ZSTL18], kde sú navzájom pospájané jednotlivé podsiete pozostávajúce z vlastného enkodéru-dekodéru pomocou tzv "dense"prepojení.



Obrázek 2: Nested-UNet

Ako ztrátová funkcia bola v konečnom dôsledku použitá DiceLoss, avšak bolo experimentované aj s kombinovanou ztrátovou funkciou pozostávajúcou s CategoricalCrossEntropy a DiceLoss. Ako optimalizátor bol využitý Adam, s hodnotou learning rate testovanou na 0.01, 0.001 a 0.0002.

5 Vyhodnotenie

Celkovo bolo natrénovaných 5 modelov, každý pre inú architektúru. Prvotne sme stanovili optimálny parameter "learning rate"ktorý bol v našom prípade 0.0002. Následne bol každý z modelov trévaný po dobu max. 200 epoch, pričom trévanie niektorých modelov sa ukončilo predčasne, keďže nedošlo k zlepšeniu validačnej ztrátovej funkcie po dobu 100 epoch. Pre vyhodnotenie účinnosti jednotlivých modelov sme použili Intersection over Union (IoU) metriku [RTG⁺19]. Výstupom každého modelu bolo 8 kategórií (spolu s pozadím, ktoré sme vo výsledkoch neuvádzali) detailnejšie popísaných v tabuľke 1.

Kategórie segmentácie	
Index	Kategória
1	Cesta
2	Príroda
3	Objekt
4	Obloha
5	Stavba
6	Človek
7	Vozidlo

Tabulka 1: Kategórie segmentácie



Obrázek 3: Výsledok segmentácie

Na obrázku 3 je možné pozorovať výstupnú masku získanú po segmentácii z pôvodného obrázku. Aj keď je samotná maska relatívne presná voči vstupnému obrázku, má určité nedostatky v podobe miernych nepresností medzi hranicami kategórií (prevažne budovy a príroda), alebo chýbajúce časti objektov (červená farba) a drobné detaily ľudí (modrá farba). Podľa našich predpokladov jedným z hlavných dôvodov stojacim za týmito nedostatkami môže byť príliš malý vstupný obrázok (8-násobné zmenšenie), spôsobujúci zhoršenie viditeľnosti niektorých črt v obrázku.

IoU skóre pre jednotlivé kategórie a modely								
Model	Celkové	1	2	3	4	5	6	7
U-Net	80.5%	97.2%	86.3%	51.1%	91.7%	85%	66%	86.1%
AttU-Net	80.5%	97.3%	86.2%	51%	91.8%	84.7%	66.5%	85.8%
NestedU-Net	78.2%	97%	84.2%	45.5%	91.6%	82.9%	62.1%	83.9%
R2U-Net	61.7%	91.9%	74.2%	35.5%	81.8%	68%	40.8%	39.4%
AttR2U-Net	50.4%	72%	73.4%	21.4%	51.2%	51.5%	26.3%	56.9%
Modely z oficiálnych benchmarkov pre cityscapes dataset								
U-Net++	75.5%	-	-	-	-	-	-	-
Ghost-UNet	74%	-	-	-	-	-	-	-
SAIT SeeThroughNet	93.2%	98.9%	94.3%	81.1%	96.2%	95%	90.1%	96.7%

Tabulka 2: IoU skóre pre jednotlivé kategórie a modely

Z výsledkov segmentácie uvedených v tabuľke 2 je možné vyčítať, že vo všeobecnosti modifikácie U-Netu (až na AttU-Net) dosahovali horšie výsledky v porovnaní s pôvodnou architektúrou pri použití rovnakých hyper-parametrov. Ako bolo badateľné na obrázku 3, problémy so segmentáciou objektov (kategória 3) a ľudí (kategória 6) pretrvávali u všetkých obrázkov a prejavili sa všeobecným znížením IoU skóre oproti ostatným kategóriám.

Modely natrénované v tomto projekte si vedli relatívne dobre, dosahujúc vyššiu celkovú presnosť oproti U-Netom z iných zdrojov (U-Net++, Ghost-Net), avšak dosiahli nedostatočné výsledky v porovnaní s inými architektúrami (SAIT SeeThroughNet). V konečnom dôsledku je možné z získaných výsledkov zhrnúť, že architektúra U-Net nie je zrovna najefektívnejšia pre segmentáciu CityScapes datasetu.

6 Záver

Výsledkom projektu je komplexný nástroj na trénovanie segmentačných modelov rôznych architektúr pre CityScapes dataset, obsahujúci preprocessing, trénovacie skripty, augmentáciu a validačný skript. Pre dosiahnutie čo najlepších výsledkov boli testované rôzne hyper-parametre, rôzne modifikácie architektúry U-Net a množstvo augmentačných metód. Ako celkový víťaz skončil pôvodný U-Net obohatený o Attention blok a pridanie rekurentného bloku do architektúry sa ukázalo ako neefektívna voľba ústiaca v pokles celkovej presnosti o 20%.

Reference

- [AHY⁺18] Md. Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *CoRR*, abs/1802.06955, 2018.
- [COR⁺16] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. *CoRR*, abs/1604.01685, 2016.
- [RTG⁺19] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 658–666, 2019.
- [ZSTL18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CoRR*, abs/1807.10165, 2018.