

Interaktívna segmentácia obrazu pomocou Inside-Outside Guidance

Sabína Gregušová, Jan Šamánek, Adrián Tulušák
xgregu02, xsaman02, xtulus00

1 Úvod

V posledných rokoch sa začali rapídne rozvíjať práce zamerané na segmentáciu obrazu. Tento typ úlohy má potenciál nájsť uplatnenie v rôznych odvetviach, menovito v samoriadiacich vozidlách, analýze medicínskych či vzdušných snímok, alebo v editácii videí či fotiek, a mnohých iných. Zaujímavou podúlohou pre segmentáciu obrazu je práve interaktívna segmentácia obrazu. Cieľom všeobecnej segmentácie je identifikovať a vysegmentovať všetky objekty v obraze na základe príslušnej triedy, zatiaľ čo interaktívna segmentácia sa zameriava na oddelenie jedného, užívateľom vybraného objektu (*foreground*), od všetkého ostatného v obraze (*background*).

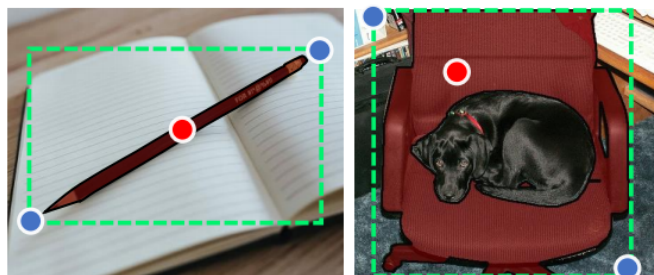
1.1 Existujúce riešenia

Jedným z prvých algoritmov, ktorý využíval deep-learning algoritmus bol *Deep Interactive Object Selection* pre interaktívnu segmentáciu obrazu predstavený v [5]. Tento článok jednoducho zhŕňa aj všetky predchádzajúce prístupy, napríklad dovtedy najznámejší [1], ktorý využíval algoritmus *Interactive Graph Cut*, na ktorý ďalej naviazali v článku [4] s optimalizovanou iteratívnou verziou. Tieto staršie algoritmy však veľmi záviseli na kvalite a hlavne množstve užívateľského vstupu, zatiaľ čo algoritmus *Deep Interactive Object Selection* veľmi zredukoval požadované množstvo užívateľského vstupu. Za užívateľský vstup prijímal *negatívne* (tam, kde sa objekt nachádza) a *pozitívneho* kliknutia (tam, kde sa objekt nenachádza), ktoré boli ďalej transformované do *Euklidovských distance máp*. Tento model bol natrénovaný pomocou FCN siete a dosahoval IoU až okolo 85% pri viac ako 6 kliknutiach.

Hoci tento algoritmus dosahoval dobré metriky, minimálne 6 kliknutí je pre bežného užívateľa stále dosť. Tento problém sa snaží riešiť algoritmus *Inside-Outside Guidance* v [6], ktorý môžeme považovať za jeden zo State-of-the art systémov, a na tomto prístupe je založená aj naša implementácia projektu.

2 Metóda

Náš tím si zvolil prístup *Inside-Outside Guidance* (ďalej iba IoG) [6], ktorého hlavným cieľom je zredukovať množstvo užívateľského vstupu a dosahovať metriky porovnateľné s predchádzajúcimi existujúcimi riešeniami. Podstatou IoG je najskôr získať *bounding box* pomocou dvoch kliknutí užívateľa (buď dvojica *horný ľavý roh*, *dolný pravý roh* alebo *horný pravý roh*, *dolný ľavý roh*) a zvyšné dve chýbajúce súradnice tohto obĺžnika je možné dopočítať. Užívateľ ďalej umiestni jedno kliknutie do vnútra objektu, ktorý chce vysegmentovať. Tento prístup ďalej umožňuje pridávať kliknutia aj po segmentácii obrazu a spresňovať ju, ak s ňou užívateľ nie je spokojný. Výhodou je, že bounding box nemusí úplne tesne obklopovať vybraný objekt. Kliknutia sú následne reprezentované ako v článku [3] pomocou *heatmaps*, kde bounding box predstavuje jednu heatmapu a zvyšné kliknutia druhú. Pri štandardnom RGB obrázku je má teda výsledný obrázok 5 kanálov (3 pre RGB a 2 pre heatmaps).



Obr. 1: Ukážka použitia prístupu *Inside-Outside Guidance* pre užívateľský vstup. Prevzaté z [6].

Pri tréňovaní siete nie je reálne, aby bol vstup získaný od skutočného užívateľa, ale musí byť náhodne vzorkovaný. Toto je implementované výberom bounding boxu z anotácie obrázku, ku ktorému je pripočítaný náhodný šum, aby vstup vyzeral ako od skutočného užívateľa. Kliknutie vo vnútri objektu je náhodne vygenerované vo vnútri objektu, s paddingom od okraja bounding boxu. Najväčšou výhodou tohto prístupu je schopnosť generalizovať aj iný typ predtým nevidených obrázkov bez potreby finetunovania.

3 Dataset

Momentálne existujú mnohé dostupné datasety určené pre segmentáciu obrazu, ktoré sa dajú adaptovať aj na interaktívnu segmentáciu. Medzi najčastejšie používané patrí *Pascal*, *Grabcut*, *Berkley* alebo *MS Coco*; a sú najčastejšie používané pre validáciu a porovnanie presnosti dnešných state-of-the-art systémov.

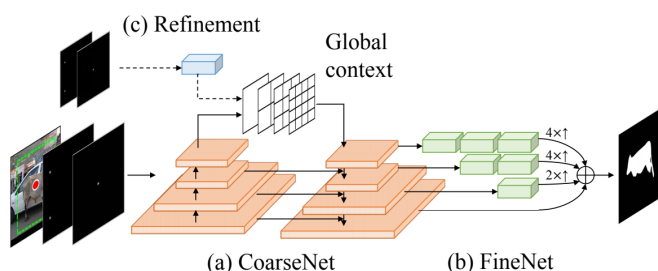
Pre náš projekt sme sa rozhodli použiť dataset *MS Coco* pre evaluáciu aj tréning, pretože obsahuje cez 80 objektov vo viac ako 200000 obrázkoch, ktoré zachytávajú bežné momenty zo života.



Obr. 2: Ukážka obrázkov z datasetu *MS Coco*. Prevzaté z [2].

4 Model

Model pre túto úlohu je založený na dvoch podsieťach: *CoarseNet* a *FineNet*. *CoarseNet*



Obr. 3: Ukážka konceptu architektúry pre túto úlohu. Prevzaté z [6].

5 Evaluácia

6 Praktická aplikácia a GUI

Použitá literatúra

[1] Boykov, Y.; Jolly, M.-P.: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Proceedings Eighth IEEE Interna-*

tional Conference on Computer Vision. ICCV 2001, doi:10.1109/iccv.2001.937505.

- [2] Lin, T.-Y.; Maire, M.; Belongie, S.; aj.: Microsoft COCO: Common Objects in Context. *Computer Vision – ECCV 2014*, 2014: str. 740–755, doi:10.1007/978-3-319-10602-1_48.
- [3] Maninis, K.-K.; Caelles, S.; Pont-Tuset, J.; aj.: Deep Extreme Cut: From Extreme Points to Object Segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, doi:10.1109/cvpr.2018.00071.
- [4] Rother, C.; Kolmogorov, V.; Blake, A.: "GrabCut". *ACM SIGGRAPH 2004 Papers on - SIGGRAPH '04*, 2004, doi:10.1145/1186562.1015720.
- [5] Xu, N.; Price, B.; Cohen, S.; aj.: Deep Interactive Object Selection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, doi:10.1109/cvpr.2016.47.
- [6] Zhang, S.; Liew, J. H.; Wei, Y.; aj.: Interactive Object Segmentation With Inside-Outside Guidance. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, doi:10.1109/cvpr42600.2020.01225.