

Interaktivní segmentace bodových mračen

KNN - Konvoluční neuronové sítě

Zuzana Hrklová xhrklo00
Martin Kneslík xknesl02
Vojtěch Vlach xvlach22

Obsah

1 Zadanie	1
2 Úvod do problematiky	1
3 Existujúce prístupy	1
4 Dataset	1
5 Zvolený prístup	2
6 Hodnotiace prostredie	3
7 Experimenty	3
8 Dosiahnuté výsledky	4
9 Interaktivní aplikace	4
10 Záver	4
11 Rozdelenie práce v tíme	5

1 Zadanie

Libovolnou segmentační úlohu lze změnit na interaktivní tím, že na vstup sítě nedám jen obraz, ale i uživatelský vstup, třeba jako další "barevný" kanál s místy, které uživatel označil. Podobně to jde u bodových mrače. Můžete využít existující datasety (např. KITTI, NYU Depth V2, NYU Depth V2 - Kaggle), nebo si i můžete pujčit LIDAR Livox Horizon, případně nějakou RGB-D kameru typu Kinect. Ning Xu, Brian Price, Scott Cohen, Jimei Yang, and Thomas Huang. Deep Interactive Object Selection. CVPR 2016. <https://sites.google.com/view/deepselection>

2 Úvod do problematiky

Segmentácia obrazu je proces, ktorý rozdeľuje obraz na viacero oblastí, z ktorých každá predstavuje iný objekt alebo oblasť obrazu.

Využíva sa na rozpoznávanie dôležitých detailov v obraze či zníženie zložitosti. Odborníkom tento prístup umožňuje izolovať konkrétné časti obrazu, aby získali zmysluplné poznatky.

Interaktívna segmentácia bodových mračien kombinuje sémantickú segmentáciu s užívateľským vstupom pre dosiahnutie presnejších výsledkov. Užívateľ môže interaktívne ovplyvniť výsledky segmentácie vyberaním bodov alebo regiónov záujmu vo vizualizácii bodového mračna na doladenie segmentačného výstupu.

3 Existujúce prístupy

Súčasné metódy využívané na 3D segmentáciu sú zvyčajne trénované pomocou plne kontrolovaného učenia (fully supervised learning). To však vedie k vysokým nárokom na trénovacie dát a nízkej generalizácii v momente keď je natrénovaný model postavený voči novým dosiaľ nevideným dátam a triedam.

Metódy, ktoré ku segmentácii využívajú užívateľské vstupy sú často založené na princípe prevádzania medzi 2D a 3D reprezentáciou dát nakoľko pre získanie interaktívneho vstupu zobrazujú dané dátu užívateľovi v rámci 2D domény, z ktorej je potom potrebné vytiahnuť a previesť nové informácie do 3D reprezenzácie. Takéto prístupy si vyžadujú vlastné architektúry neurónových sietí a kardinálne niekoľko vstupných modalít.

Inšpiráciou našej práce je článok [4] reprezentujúci prvú state-of-the-art implementáciu interaktívnej segmentácie na čisto 3D reprezentáciu dát. V článku je predstavený prístup generovania a reprezentácie užívateľských vstupov potrebných na trénovanie modelu. Výsledný model je predtrénovaný pomocou Minkowski Engine[2], čo je PyTorch knižnica s podporou pre efektívne spracovanie dát v 3D priestore.

Vstupom siete je teda 3D scéna (xyz koordináty a rgb hodnoty) a dve binárne masky pre pozitívne a negatívne užívateľské vstupy - kliky. Výstupom siete je binárny softmax označujúci každý bod v sieti ako práve segmentovaný objekt alebo pozadie.

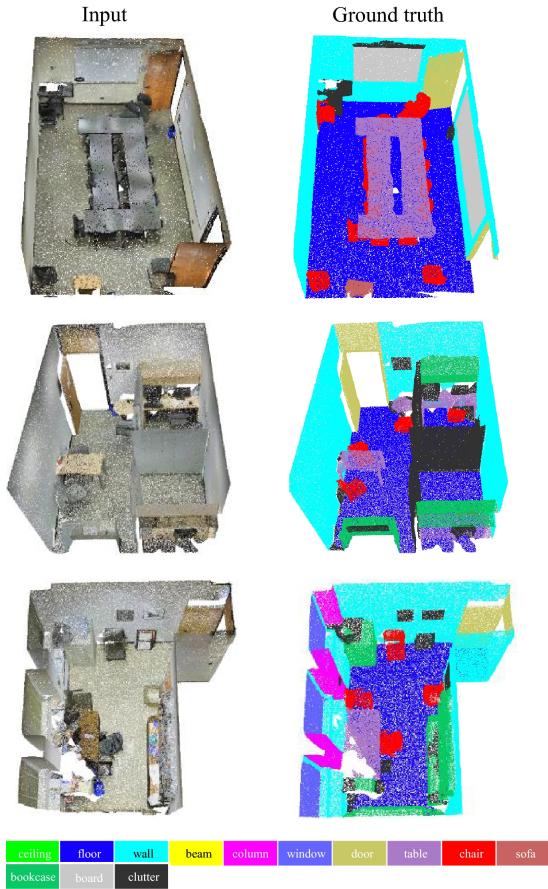
Výsledný model bol trénovaný na datasete ScanNetV2 a náhodne generovaných bodoch reprezentujúcich užívateľské vstupy. Výsledky poukazujú na jeho dobrú generalizáciu iných dosiaľ nevidených či už indoorových ale aj outdoorových datasetov.

4 Dataset

Pre trénovanie nami vytvorenej implementácie danej úlohy bol vybraný dataset S3DIS (Stanford 3D Indoor Scene Dataset) [1]. Dataset obsahuje 6 areálov, v rámci ktorých je obsiahnutých 271 miestností s 13 rôznymi sémantickými kategóriami.

Pred trénovaním je dataset predspracovaný skriptom `convert_dataset.py`.

Dataset bol následne rozdelený na 3 časti. Trénovací dataset tvorený z areálov 1-4, validačný dataset obsahujúci dátá areálu 6. Pre testovanie modelu bola využitá posledná časť datasetu dosiaľ nevidená našim modelom, predstavujúca areál 5



Príklad dát z používaného datasetu S3DIS

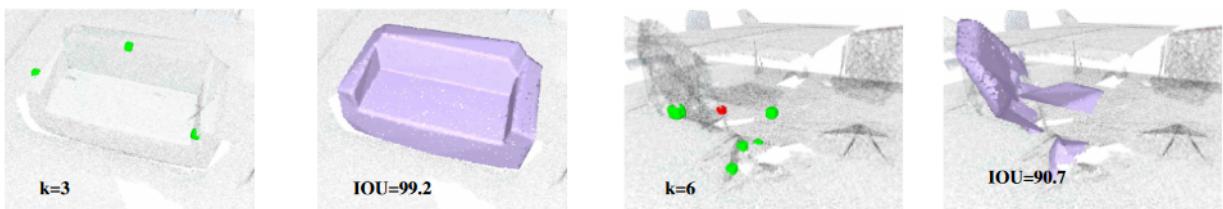
5 Zvolený prístup

Základné riešenie je prevzaté zo spomínaného článku [4] pomocou InterObject3d repozitára [3]. Pre jeho implementáciu bol zvolený Python, PyTorch a MinkowskiEngine [2]. Prevzatý model bol trénovaný na dataseite ScanNet a predstavuje 3D Voxel UNet s binárnej klasifikáciou využívajúcou sparse tensor.

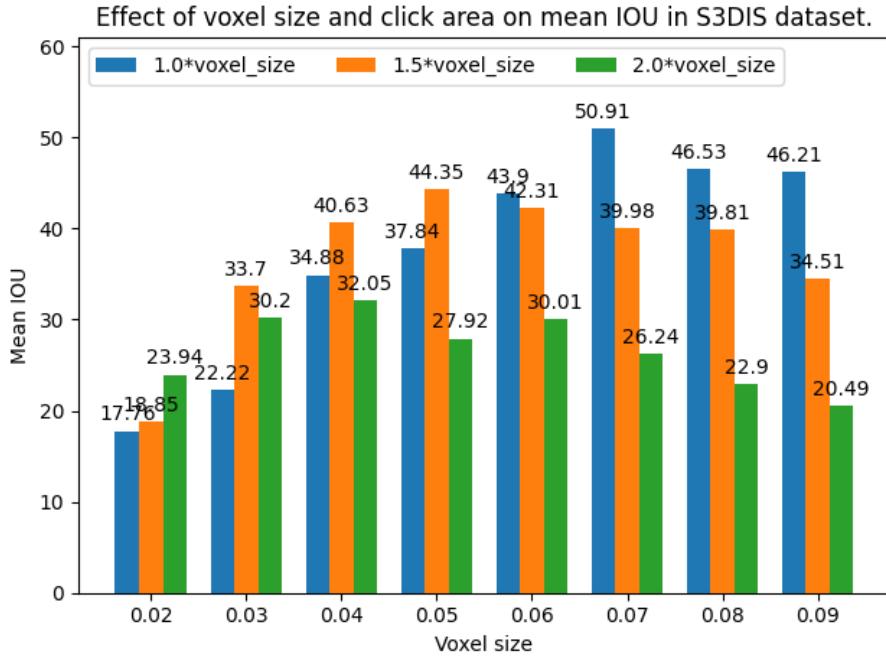
Sparse tensor je typ dátovej štruktúry využívaný pri viacdimenzionálnych dátach, obsahujúcich väčšie množstvo nulových hodnôt. Štruktúra obsahuje nenulové hodnoty spolu s ich príslušnými indexami čo vedie na zníženie pamäťových nárokov. MinkowskiEngine je knižnica vytvorená na efektívne spracovanie a vykonávanie sparse operácií ako konvolúcia či batch normalizácia, práve takýchto sparse tensorov v 3D.

Výsledné hodnoty modelu trénovanom na dataseite ScanNet zobrazujúce počet potrebných klikov pre dosiahnutie požadovanej presnosti pomocou metriky NOC @ k % IoU:

S3DIS Area 5		
80%	85%	90%
6.8	8.9	11.8



Výsledky segmentácie datasetu S3DIS pri použití k klikov[4]



IOU pro rôzne voxel size a click area. Barva znázorňuje pomér. např. oranžová pro voxel size 0.05 je 0.75.

Nami využívaný S3DIS dataset bol spracovaný pre jednoduchšiu a efektívnejšiu manipuláciu. V originálnom S3DIS datasete sú anotácie uložené v samostatných súboroch vo forme bodových mračien (pointcloudov). Naše spracovanie datasetu tieto samostatné súbory zlúči do jedného, čoho výsledkom vzniká súbor, v ktorom má každý bod v scéne priradenú svoju tiedu. Nami spracovaný dataset je naviac uložený binárne oproti pôvodnému uloženému v ASCII čo výrazne znižuje jeho pamäťové nároky.

Takto spracovaný dataset bol následne využitý na dotrénovanie modelu prevzatého z práce [4].

Vrámcí projektu bola vytvorená interaktívna aplikácia slúžiaca na vizualizáciu a verifikáciu správnosti modelu na reálnych užívateľských vstupoch. Do aplikácie je možné načítať bodové mračno reprezentujúce scénu na segmentáciu. Následne sa označí ľubovoľným počtom kliknutí objekt na segmentáciu a vloží model zvolený na segmentáciu. V scéne sa farebne označí výsledok segmentácie.

6 Hodnotiace prostredie

Výsledný model bol testovaný na doposiaľ nevidenej časti dát oddelenej z datasetu S3DIS, konkrétnie časťou areálu 5.

Pre vyhodnotenie presnosti modelu bola využitá metrika IoU (Intersection over Union) porovnávajúca výsledok segmentácie so skutočnými hodnotami (ground truth) pomocou rovnice:

$$\text{IoU} = \frac{\text{Plocha prieniku}}{\text{Plocha zjednotenia}}$$

Výsledná hodnota je v rozmedzí hodnôt 0-1, kde 0 značí, že žiadne prekrytie segmentovaných dát a skutočnými datami a 1 úplné prekrytie týchto dvoch množín.

Pre ďalšie vyhodnocovanie bola využitá metrika Number of Clicks: NOC @ k % IoU, ktorá predstavuje počet potrebných klikov pre dosiahnutie segmentácie s k% presnosťou IoU.

7 Experimenty

Experimentovanie pri trénovaní modelu bolo zamerané najmä na zistenie optimálneho počtu užívateľských vstupov - klikov na jeden objekt pre získanie čo najlepšej segmentácie.

Model bol najskôr natrénovaný s premenným počtom klikov na segmentovaný objek v rozmedzí 1-3. To či v danom momente pri ténovaní dostal model na vstup 1, 2 alebo 3 kliky bolo určené náhodne.

Tabulka 1: Výsledky IOU pro 3 body.

Model	IOU total	board	bookcase	ceiling	clutter	table	wall	window	...
Pretrained	41.87	36.70	40.36	69.70	37.22	40.25	48.08	26.39	...
OUR_voxel_0.5_click_0.1	39.50	38.15	40.71	65.60	35.24	38.96	45.54	24.72	...
OUR_voxel_0.7_click_0.7	44.10	34.31	38.76	69.14	41.54	40.73	45.78	26.46	...

Tabulka 2: Výsledky NoC k % IOU.

	80 % IOU	85 % IOU
InterObject3D_pretrained	12.96	14.76
OUR_voxel_0.5_click_0.1	10.29	11.77
OUR_voxel_0.7_click_0.7	9.5	11.01

Ako druhý v poradí bol natrénovaný model s využitím fixného počtu užívateľských vstupov.

Následne boli oba modely postavené oproti sebe a prebehlo vyhodnotenie ich schopností pomocou metrík spomínaných v predchádzajúcej sekcií 6.

V rámci experimentu jsme také testovali ideální hodnoty pro `voxel size` a `click area`. `Voxel size` značí, jak daleko od sebe musí být body ve vstupním point cloudu, aby se vepsali do jiného voxelu. `Click area` se aplikuje před voelizací na výběr okolí vybraného bodu (kliku). Tato velikost ovlivňuje kolik voxelů si uchová informaci o kliku.

Ideální poměr těchto parametrů znázorňuje obrázek TODO ID.

8 Dosiahnuté výsledky

V rámci pokusů o trénování se nám povedlo natrénovat dva slibné modely. Jejich rozdíl je ve `voxel size` a `click area`. Zjistili jsme, že právě tyto parametry jsou zásadní pro dosažení ideální úspěšnosti. Každý dataset je totiž vytvořený jinou technologií a tyto parametry určují hlavně velikost vstupu a signifikanci kliků.

Dosiahnuté výsledky jsou vidět v Tabulce 1 (8) a Tabulce 2 (8). Podle dosažených výsledků lze pozorovat zlepšení oproti původnímu modelu.

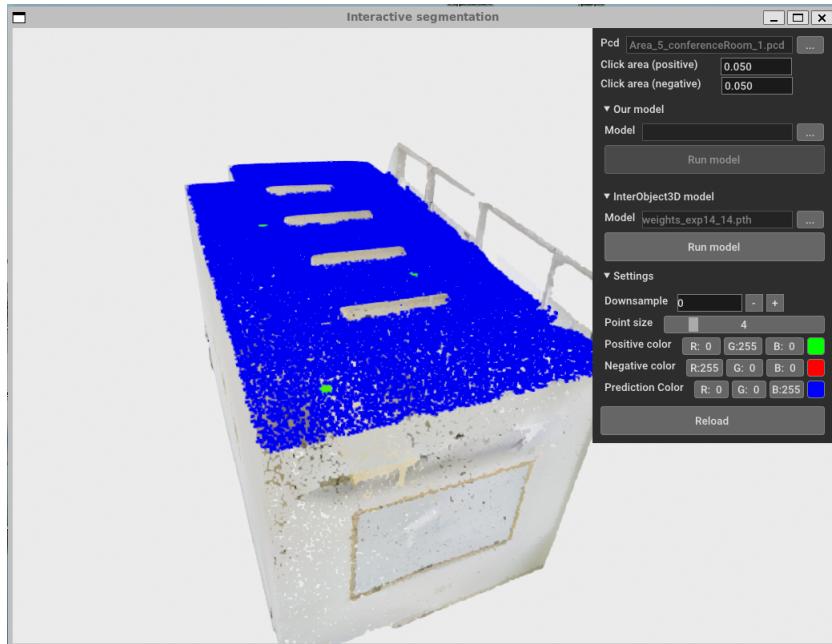
9 Interaktivní aplikace

Pro účel ověření funkčnosti modelu jsme vytvořili jednoduchou demo aplikaci, která umí vybrat model a point cloud, zadat kliky, spustit inferenci modelu a zobrazit výsledky. Více viz screenshot níže.

10 Záver

Za použití předtrénovaného modelu a neviděného datasetu se nám podařilo dotrénovat 2 nové modely na novou testovací sadu. Nevelké zlepšení na dané sadě je pravděpodobně způsobené robustností původního modelu, který měl už při trénování za cíl být univerzální mezi sémantickými třídami.

Hlavním přínosem této práce nejsou ani tak konkrétní výsledky jako spíše zmapování úlohy, vytvoření interaktivní demo aplikace a připravení půdy pro další experimenty či výzkum.



Screenshot z naší demo aplikace, který používá natrénovaný model.

11 Rozdelenie práce v tíme

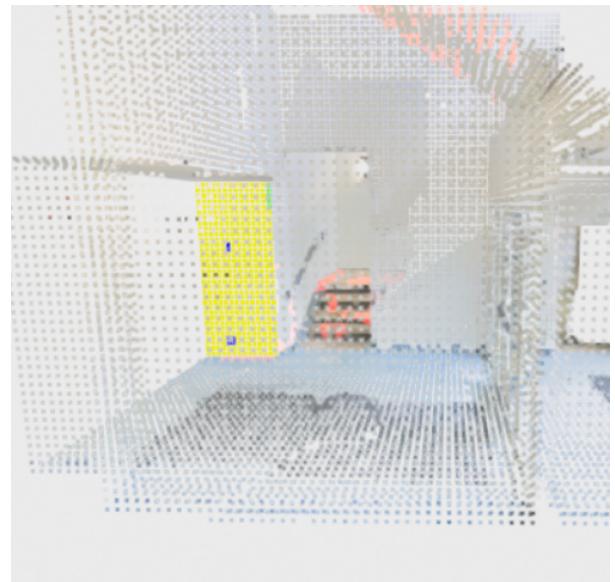
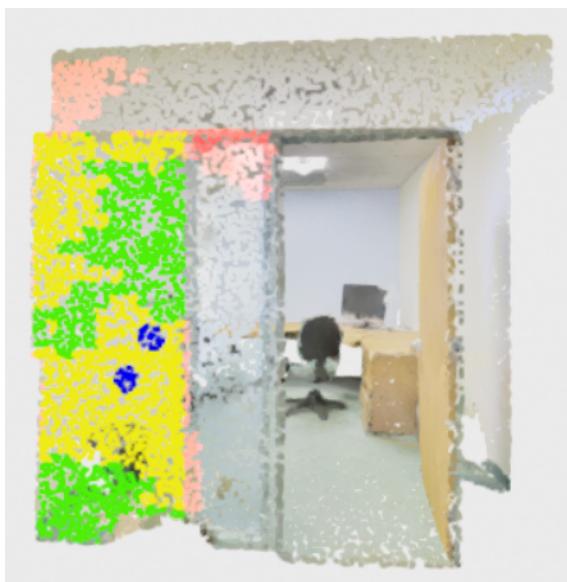
- Vojtěch Vlach (xvlach22): trénovanie, experimenty, testovanie
- Martin Kneslík (xknesl02): spracovanie datasetu, dataloader, interaktívna aplikácia
- Zuzana Hrkľová (xhrklo00): trénovanie, testovanie, dokumentácia

Verejný repozitár projektu:

https://github.com/vlachvojta/KNN_3D_segmentation?tab=readme-ov-file

Příklady point cloudů

(modrá = klik, zelená = ground truth, žlutá = správná predicke, červená = chybná predikce)



Reference

- [1] ARMENI, I., SENER, O., ZAMIR, A. R., JIANG, H., BRILAKIS, I. et al. 3D Semantic Parsing of Large-Scale Indoor Spaces. In: *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. 2016.
- [2] CHOY, C., GWAK, J. a SAVARESE, S. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, s. 3075–3084.
- [3] KONTOGIANNI, T. *InterObject3D* [<https://github.com/theodorakontogianni/InterObject3D>]. [cit. 2024-03-30].
- [4] KONTOGIANNI, T., CELIKKAN, E., TANG, S. a SCHINDLER, K. Interactive Object Segmentation in 3D Point Clouds. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. 2023, s. 2891–2897. DOI: 10.1109/ICRA48891.2023.10160904.