

Politechnika Warszawska

WYDZIAŁ ELEKTRONIKI
I TECHNIK INFORMACYJNYCH



Semantyczna analiza środowiska
przez robota usługowego

Sprawozdanie

Piotr Hondra

WARSZAWA 2022

Spis treści

1. Wstęp teoretyczny	3
1.1. Klasyfikacja sceny	3
1.2. Segmentacja obrazu	3
2. Cel pracy	3
3. Założenia	4
4. Motywacje	4
5. Zbiór danych	4
6. Przegląd rozwiązań	5
6.1. Metody klasyczne	5
Bibliografia	7
Spis grafik	8



Rysunek 1.1. Problem różnorodności wewnątrzklasowej oraz wieloznaczności semantycznej [1].

1. Wstęp teoretyczny

1.1. Klasyfikacja sceny

Zadanie klasyfikacji sceny polega na przyporządkowaniu kategorii miejsca w które przedstawia obraz. Istnieje duża różnica między klasyfikacją obrazka, a klasyfikacją sceny. Klasyfikacja obrazka jako taka zajmuje się przyporządkowaniem klasy obiektu pierwszoplanowego, np. czy na obrazie znajduje się pies czy kot. Klasyfikacja sceny natomiast musi wziąć pod uwagę wszystkie cechy obrazu, zarówno tła jak i pierwszego planu by określić odpowiednie miejsce.

Zadanie klasyfikacji sceny jest trudne ze względu na problem różnorodności wewnątrzklasowej oraz wieloznaczności semantycznej, co zostało przedstawione na rys. 1.1. Pierwszy z nich polega na fakcie, iż jedno miejsce może zostać przedstawione w bardzo różnej konfiguracji m.in. oświetlenia, ekspozycji, obiektów znajdujących się na obrazie. Drugi jest związany z występowaniem tych samych obiektów dla różnych klas scen.

1.2. Segmentacja obrazu

Zadanie segmentacji obrazu to przyporządkowanie każdemu pikselowi etykiety (rys. 1.2). W rezultacie obraz zostaje podzielony na homogeniczne regiony pod względem pewnych własności.

2. Cel pracy

Celem pracy inżynierskiej są dwa zadania:

- segmentacja środowiska wewnątrz budynku



Rysunek 1.2. Segmentacja wnętrza pomieszczeń [2].

- klasyfikacja pomieszczeń

3. Założenia

Praca zakłada wykonanie celów pracy w środowisku wewnątrz budynków, co więcej będzie to środowisko domowe. Ponadto inferencja zostanie przeprowadzona na robocie Tiago, który jest wyposażony w kamerę Kinect.

4. Motywacje

Istnieje wiele powodów dla których temat pracy jest wart uwagi.

Po pierwsze rozwiązanie może być wykorzystane w nawigacji robota. Wykrywanie przeszkód jest kluczowym aspektem możliwości poruszania się robota. Zostanie ono podjęte przez zadanie segmentacji. Należy zwrócić uwagę, że robot powinien zachowywać się ostrożniej w kuchni oraz w łazience. Ta informacja zostanie uzyskana poprzez klasyfikację sceny.

Innym zastosowaniem rozważanego rozwiązania jest pomoc dla osób niewidomych. Osoba niepełnosprawa mogłaby wówczas poruszać się po środowisku domowym z większą łatwością, mając na sobie kamerę oraz informację o otaczającej przestrzeni.

5. Zbiór danych

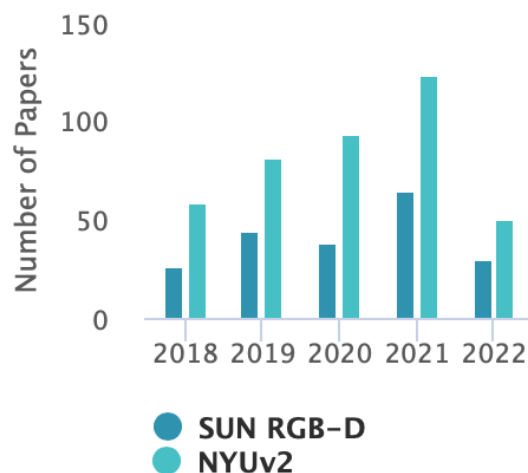
Zbiór danych powinien ściśle odpowiadać założeniom postawionym w pracy. Inferencja wymaga użycia kamery Kinect. Zatem zbiór danych powinien zawierać kategorie scen, segmentacje obrazów oraz najlepiej być ujętym przez kamerę Kinect wersji pierwszej.

Po prześledzeniu wielu zbiorów danych udało się sprostać powyższym wymaganiom, uzyskując dwa podobne zbiory danych.

Porównanie zbiorów NYUv2 oraz SUN RGBD przedstawiono w tabeli 5.1. Mimo liczbowej

Nazwa	# Ilość	# Klas obiektów	# Klas scen	RGB-D	Rozdzielczość	# Czujników	Nieposprzątane
NYUv2	1 449	894	26	✓	640 x 480	1	✓
SUN RGBD	10 335	800	47	✓	640 x 480	4	x

Tabela 5.1. Porównanie zbiorów danych [3],[4]

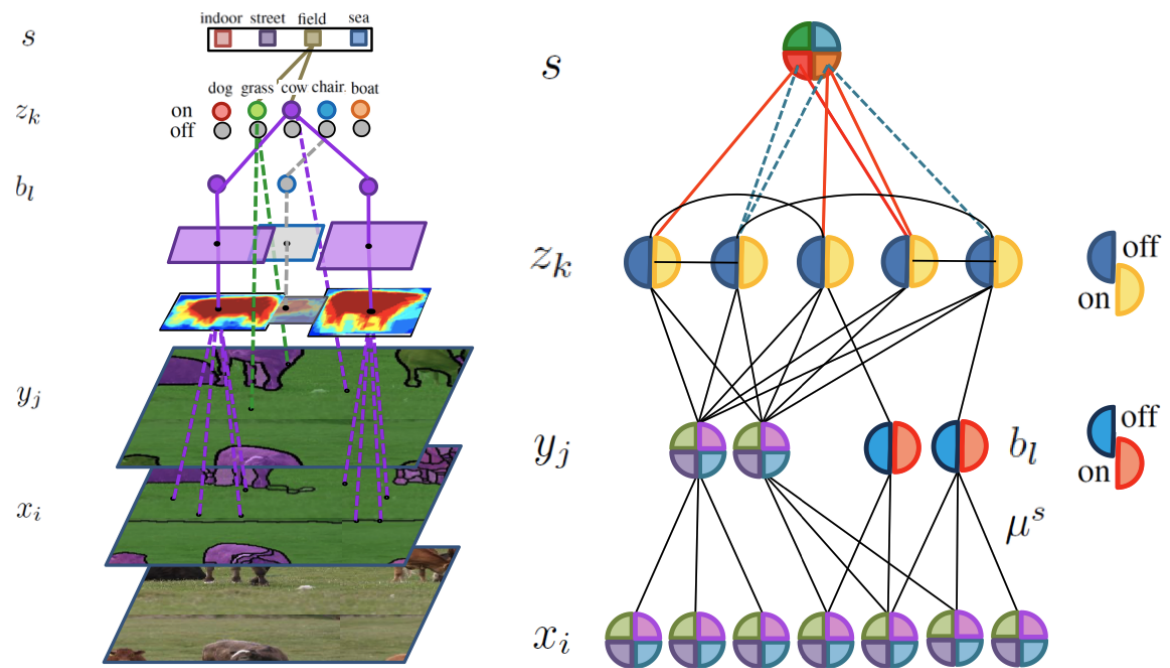


Rysunek 5.1. Szacowana liczba cytowań w latach 2018-2022 [paperswithcode.com]

przewagi SUN RGBD pod wieloma względami, ostatecznie wybrano NYUv2 z uwagi, że zbiór ten został zebrany dla pomieszczeń, w które nie są posprzątane. Fakt ten uznano, za kluczowy, iż uważano, że będzie przekładał się na lepsze rezultaty w naturalnych warunkach. NYUv2 jest też chętniej cytowany niż texttttSUN RGBD (rys. 5.1).

6. Przegląd rozwiązań

6.1. Metody klasyczne



Rysunek 6.1. Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation 2012 [5].

Bibliografia

- [1] D. Zeng, M. Liao, M. Tavakolian, Y. Guo, B. Zhou, D. Hu, M. Pietikäinen i L. Liu, „Deep learning for scene classification: A survey”, *arXiv preprint arXiv:2101.10531*, 2021.
- [2] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi i A. Agrawal, „Context encoding for semantic segmentation”, w *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, s. 7151–7160.
- [3] S. Song, S. P. Lichtenberg i J. Xiao, „Sun rgb-d: A rgb-d scene understanding benchmark suite”, w *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, s. 567–576.
- [4] N. Silberman, D. Hoiem, P. Kohli i R. Fergus, „Indoor segmentation and support inference from rgb-d images”, w *European conference on computer vision*, Springer, 2012, s. 746–760.
- [5] J. Yao, S. Fidler i R. Urtasun, „Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation”, w *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, s. 702–709.

Spis grafik

1.1	Problem różnorodności wewnątrzklasowej oraz wieloznaczności semantycznej [1].	3
1.2	Segmentacja wewnątrz pomieszczeń [2].	4
6.1	Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation 2012 [5].	6