

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	4
3. Założenia	4
4. Motywacje	4
5. Zbiór danych	5
6. Przegląd rozwiązań	5
6.1. Metody klasyczne	5
6.2. Metody oparte o głębokie uczenie	6
7. Zaproponowane rozwiązanie	8
Bibliografia	9
Spis grafik	10



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

1. Wstęp teoretyczny

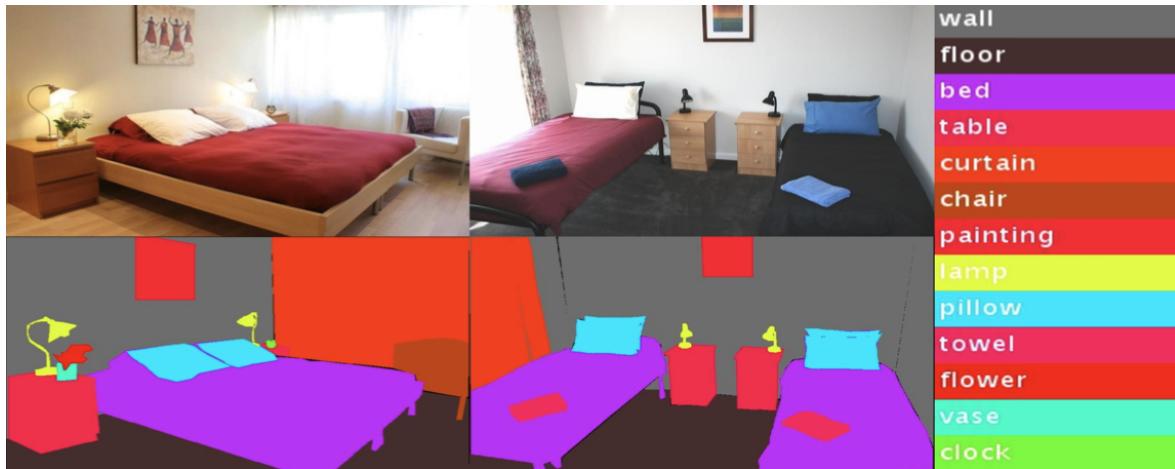
1.1. Klasyfikacja sceny

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 pierwszo-planowego, 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 wewnętrz klasowej 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. Zadanie segmentacji można rozszerzyć do zadania segmentacji instancji (ang. instance segmentation), czyli segmentacji klasycznej rozszerzonej o rozróżnienie poszczególnych obiektów w ramach tej samej klasy. W przypadku klasycznej wersji nie jesteśmy w stanie rozróżnić dwóch stojących obok siebie łózek, gdyż mapa segmentacji jest dla nich jednakowa. Segmentacja instancji pozwala natomiast takie rozróżnienie uczynić. Segmentacja semantyczna w dalszej części pracy będzie odnosić się do klasycznej wersji. Segmentacja instancji nie jest tematem pracy.



Rysunek 1.2. Segmentacja wewnętrz pomieszczeń [2].

2. Cel pracy

Celem pracy inżynierskiej są dwa zadania:

- segmentacja środowiska wewnętrz budynku
- klasyfikacja pomieszczeń

3. Założenia

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

4. Motywacje

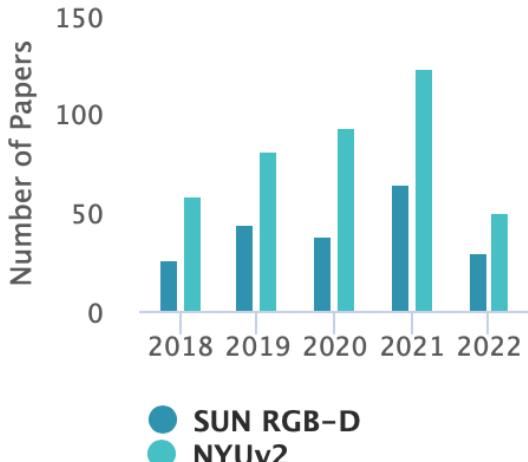
Istnieje wiele powodów, dla których temat pracy jest wart uwagi.

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

Nazwa	# Ilość	# Klas obiektów	# Klas scen	RGB-D	Rozdzielcość	# 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]

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.

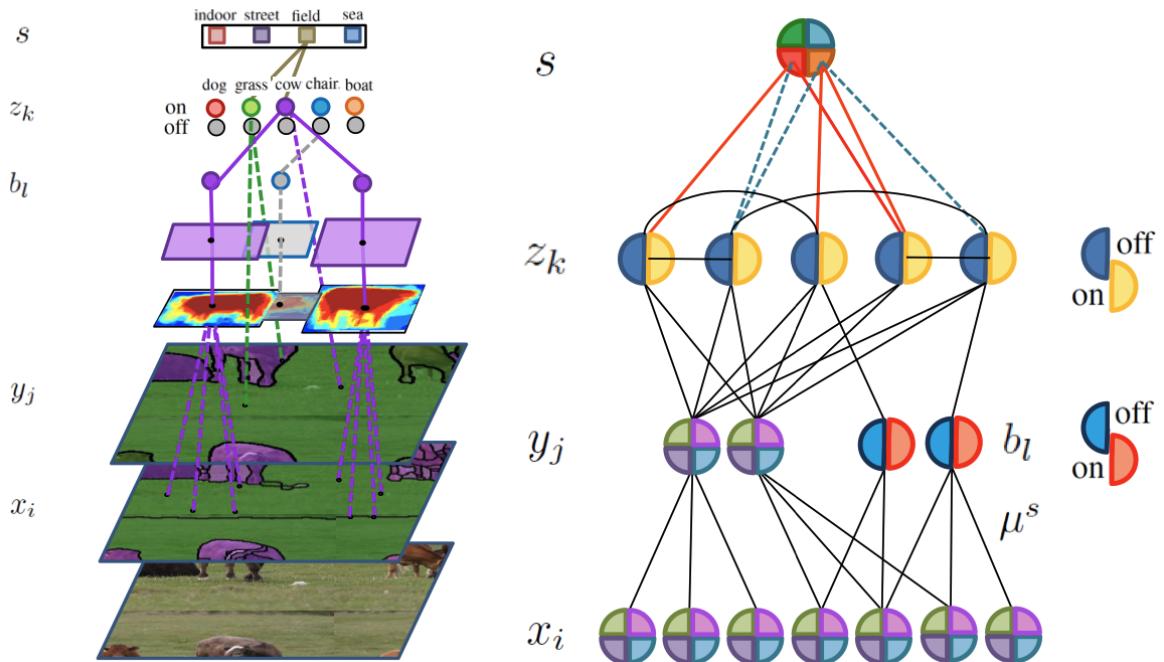
Po prześledzeniu wielu zbiorów danych udało się sprostać powyższym wymaganiom, uzyskując dwa podobne zbiory danych.

Porównanie zbiorów NYUv2 oraz SUN RGBD przedstawiono w tabeli 5.1. Mimo liczbowej 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 częściej cytowany niż SUN RGBD (rys. 5.1).

6. Przegląd rozwiązań

6.1. Metody klasyczne

Artykuł „Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation 2012 [5]” dobrze ilustruje relacje segmentacji obrazu oraz klasyfikacji sceny. Aby odpowiednio sklasyfikować scenę, należy obraz przedstawić w postaci cech, a następnie na tej podstawie dopiero wnioskować. Artykuł [5] udowadnia, iż możliwe jest przeprowadzenie wnioskowania na podstawie segmentacji (rys. 6.1).



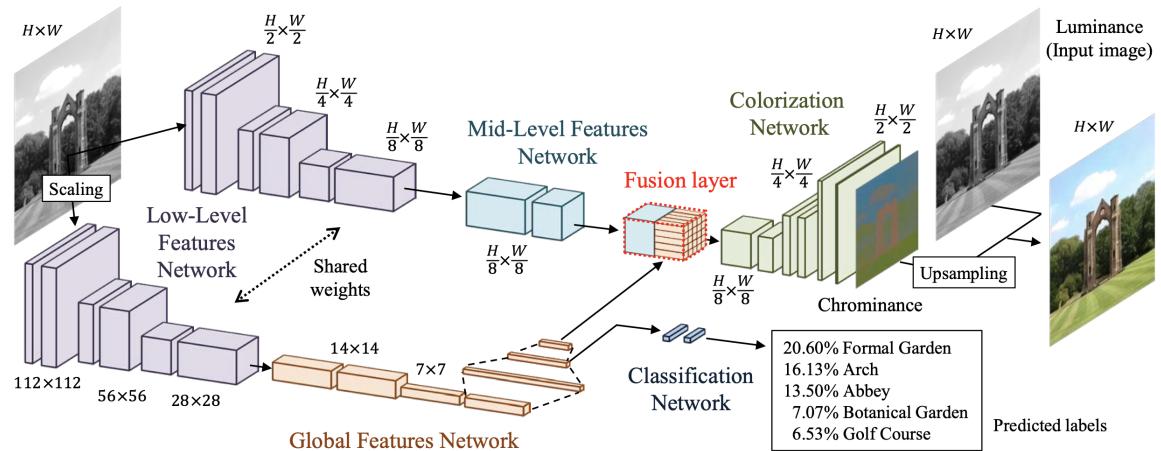
Rysunek 6.1. Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation 2012 [5].

6.2. Metody oparte o głębokie uczenie

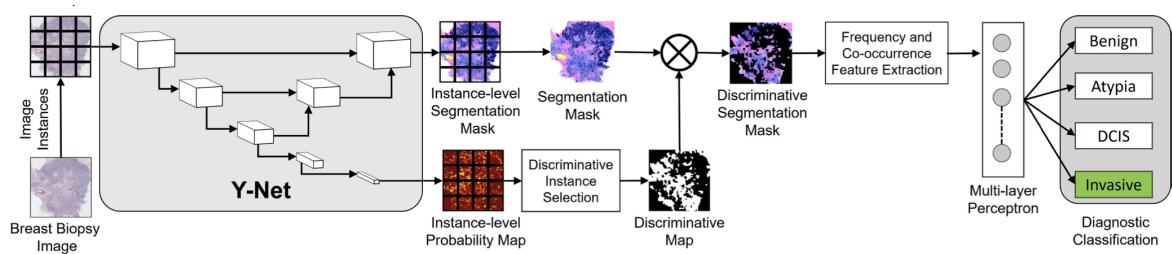
Współcześnie do zadań wizji komputerowej używa się głębokich sieci neuronowych z uwagi na ich duże zdolności generalizacji skomplikowanych przestrzeni. Celem każdej architektury jest odpowiednia ekstrakcja cech w sposób łatwo ekstrahowalny. Architektury różnią się zatem sposobem generalizacji, a dokładniej ułożeniem warstw i ich parametrów. W ramach przeglądu literatury pochyłono się nad różnymi metodami łączenia zadania segmentacji i klasyfikacji, ponieważ zadanie postawione w pracy, co do wiedzy autora, nie zostało wcześniej rozwiązane podobnymi metodami.

Pierwszy artykuł „Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification 2016 [6]” rozwiązuje problem kolorowania obrazków jednak, przekształcony może być użyty w pracy. Tego można dokonać odrzucając ostatnią warstwę konkatenacji w części segmentacji (rys. 6.2). Przedstawiona architektura symultanicznie ekstrahuje cechy globalne oraz średniego poziomu, które odpowiednio służą klasyfikacji oraz segmentacji.

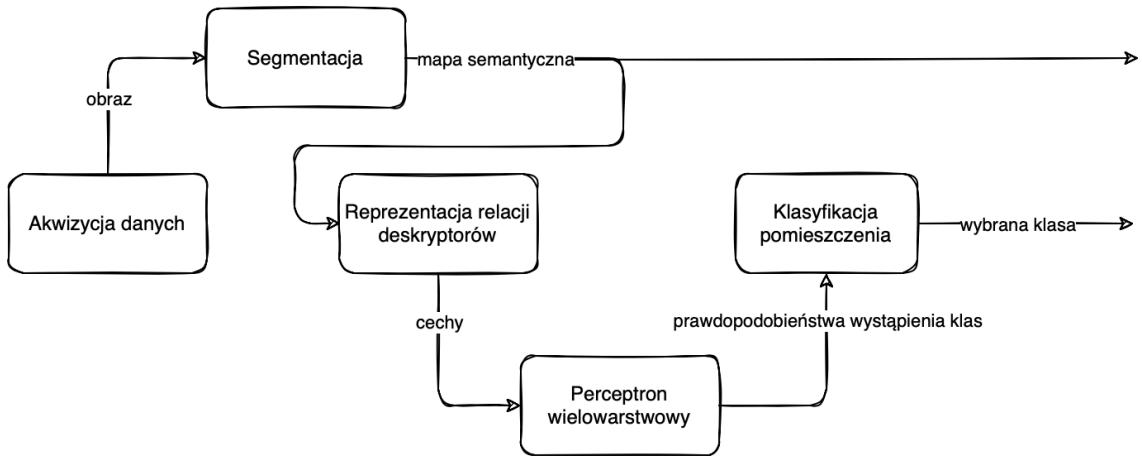
Kolejnym artykułem jest „Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images 2018 [7]”. Jest to standardowa architektura segmentacji U-Net rozszerzona o gałąź klasyfikacyjną (rys. 6.3). Rozwiązanie to jest na pewno ciekawe z punktu widzenia modularności rozwiązania.



Rysunek 6.2. Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification 2016 [6].



Rysunek 6.3. Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images 2018 [7].



Rysunek 7.1. Prototyp rozwiązania.

7. Zaproponowane rozwiązanie

W swoim rozwiążaniu chciałbym wykorzystać głębokie sieci neuronowe do realizacji celów pracy. Proponuję jako prototyp wykorzystać segmentację jako reprezentację cech wysokiej abstrakcji. Następnie na podstawie relacji opisanych wyżej cech zbudować klasyfikator w formie perceptronu wielowarstwowego. Na wyjściu zgodnie z rysunkiem 7.1 otrzymać mapę semantyczną oraz klasyfikację scen.

W kolejnych iteracjach pracy chciałbym zbudować inne architektury, a ostatecznie porównać je, wybierając architekturę o największej skuteczności.

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 rgbd 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.
- [6] S. Iizuka, E. Simo-Serra i H. Ishikawa, „Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification”, *ACM Transactions on Graphics (ToG)*, t. 35, nr. 4, s. 1–11, 2016.
- [7] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore i L. Shapiro, „Y-Net: joint segmentation and classification for diagnosis of breast biopsy images”, w *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, s. 893–901.

Spis grafik

1.1 Problem różnorodności wewnętrz klasowej oraz wieloznaczności semantycznej [1].	3
1.2 Segmentacja wewnętrz pomieszczeń [2].	4
6.1 Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation 2012 [5].	6
6.2 Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification 2016 [6]. .	7
6.3 Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images 2018 [7].	7
7.1 Prototyp rozwiązania.	8