

POLITECHNIKA WROCŁAWSKA  
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



---

# Widzenie komputerowe

---

Sprawozdanie z laboratorium

AUTOR

**Igor Lewandowski**

nr albumu: **260357**

kierunek: **Inżynieria systemów**

*14 styczeń 2023*

### Streszczenie

Zaproponowano własną sieć neuronową do klasyfikacji wycinków obrazków w zależności, czy znajduje się na nich samochód. Jakość sieci zestawiono z podobną siecią rozszerzoną o dwie warstwy ukryte, porównując wyniki klasyfikacji na zbiorze danych uczących oraz zbiorze danych DOTA. Okazało się, że rozszerzony model nieznacznie lepiej wykonuje zadanie klasyfikacji.

## 1 Wstęp – opis problemu

Problemem zadania laboratoryjnego jest klasyfikacja wycinków obrazów uchwyconych za pomocą drona, w zależności czy znajduje się w nim samochód osobowy, czy nie. Zostało ono rozwiązane za pomocą modelu sieci neuronowej, która na wyjściu zwraca etykietę jednej z klas.

## 2 Opis rozwiązania

### 2.1 Dane

Zbiór danych uzyskano z wycinków wybranych klatek z filmów wykonanych dronem udostępnionych mi przez dr inż. Jarosława Drapalę, znad ulicy Młodych Techników, Mostu Grunwaldzkiego oraz parkingu blisko Politechniki Wrocławskiej. W wyniku uzyskano 242 obrazy zawierające samochody osobowe i 464 obrazów bez samochodów osobowych zapisanych w formacie `png`. Etykiety zostały przyporządkowane poprzez zapisanie obrazów w osobnych folderach, następnie utworzeniu macierzy  $X$  oraz  $y$ , gdzie macierz  $X$  zawiera odczytane za pomocą pakietu `opencv` zdjęcia, a macierz  $y$  odpowiadającą temu zdjęciowi etykietę, która następnie jest zakodowana metodą "1 z n", tak jak w tabeli 2.1. Tak zakodowane etykiety odpowiadają wyjściu sieci neuronowej opisanej w punkcie 2.2.

Etykieta	Kodowanie
"car"	[1 0]
"no car"	[0 1]

Zdjęcia zostały wstępnie przetworzone. Zmodyfikowano ich wymiary do 100 x 100, ponieważ wycinki różniły się między sobą wielkością.

Dokonano wzbogacenia danych (ang. *data augmentation*) poprzez losową rotację obrazów, losowemu przybliżeniu, obróceniu w pionie, czy zmianie jasności obrazu. Zabiegi te doprowadziły do uzyskania zbioru o liczności 2906 obrazów ze zbalansowanymi etykietami dwóch klas.

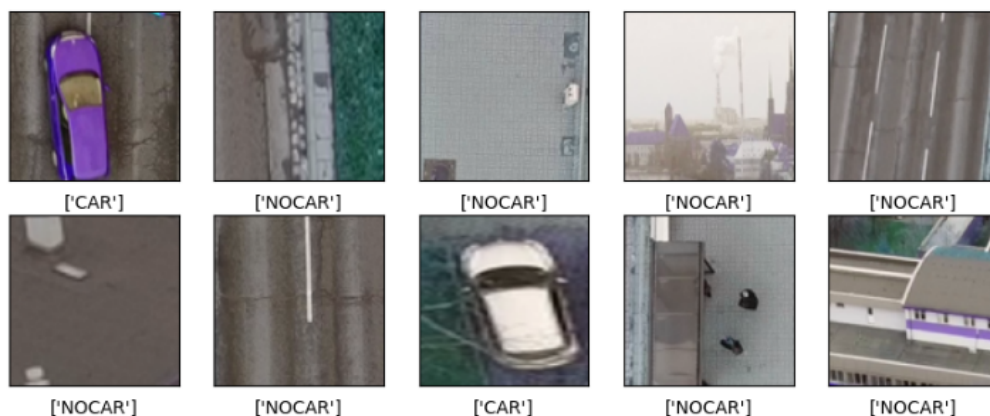
Przykładowe obrazy przedstawia rys. 1.

Następnie w celu przeprowadzeniu testów działania modelu w trudniejszych warunkach skorzystano ze zbioru danych DOTA, który zawiera zdjęcia z Google Earth, jak i zdjęcia udostępnione przez Cyclomedia B.V.. Etykiety zostały przypisane w ten sam sposób, w jaki zostały one przypisane do wycinków obrazów z drona.

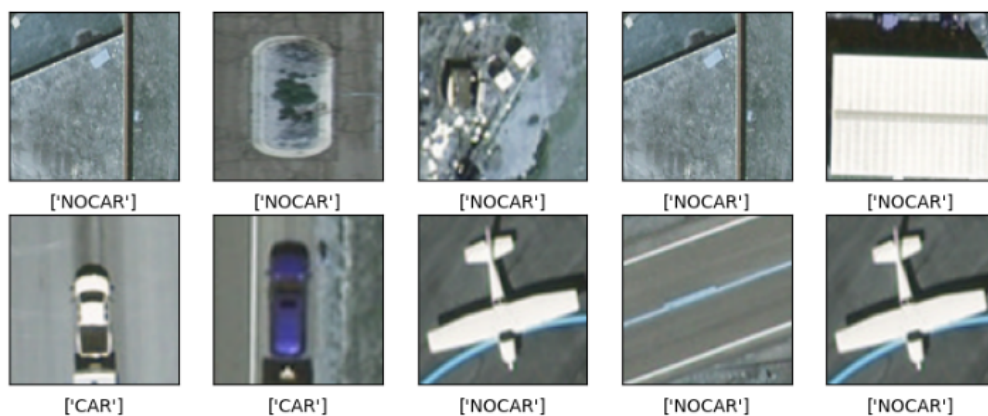
Przykładowe obrazy ze zbioru danych DOTA przedstawia rys. 2.

### 2.2 Algorytm rozpoznawania

Zadanie rozpoznawania wykonywać będzie głęboka sieć neuronowa własnego projektu (klasyfikator A). Jest to nieznaczna modyfikacja sieci głębokiej przedstawionej na zajęciach laboratoryjnych, którą wykorzystywaliśmy do klasyfikacji na zbiorze `keras fashion mnist`. Strukturę tej sieci przedstawia rysunek 3.



Rysunek 1: Przykłady zdjęć z etykietami ze zbioru wycinków klatek z drona.



Rysunek 2: Przykłady zdjęć z etykietami ze zbioru danych DOTA.

Liczba jej wyjść jest równa liczbie klas. Aktywacją ostatniej warstwy jest zatem funkcja *softmax*. Wszystkie pozostałe parametry konfiguracyjne sieci są identyczne jak użyte na laboratorium. Kod sieci udostępniłem w archiwum. Dla porównania korzystam z sieci podobnej do wcześniejszej, ale rozszerzonej o dwie warstwy ukryte. (klasyfikator B) rys 4.

. Jakość rozpoznawania oceniam z wykorzystaniem macierzy pomyłek.

### 3 Badania symulacyjne

#### 3.1 Cel i plan badania

Celem badawczym jest sprawdzenie, jak dobrze nauczy się sieć w porównaniu z jej rozszerzonym wariantem. Dodatkowo celem jest sprawdzenie, jak dobrze te sieci działają w trudniejszych warunkach, korzystając ze zbioru danych DOTA, z którego obrazy nie były w zbiorze treningowym. Obie

Model: "sequential\_6"

Layer (type)	Output Shape	Param #
flatten_6 (Flatten)	(None, 30000)	0
dense_12 (Dense)	(None, 32)	960032
dense_13 (Dense)	(None, 2)	66

=====  
Total params: 960,098  
Trainable params: 960,098  
Non-trainable params: 0  
=====

Rysunek 3: Struktura proponowanej sieci neuronowej (klasyfikator A)

Model: "sequential\_5"

Layer (type)	Output Shape	Param #
flatten_5 (Flatten)	(None, 30000)	0
dense_20 (Dense)	(None, 128)	3840128
dense_21 (Dense)	(None, 64)	8256
dense_22 (Dense)	(None, 32)	2080
dense_23 (Dense)	(None, 2)	66

=====  
Total params: 3,850,530  
Trainable params: 3,850,530  
Non-trainable params: 0  
=====

Rysunek 4: Struktura rozszerzonej sieci neuronowej (klasyfikator B)

sieci uczyły się na tych samych danych tyle samo epok.

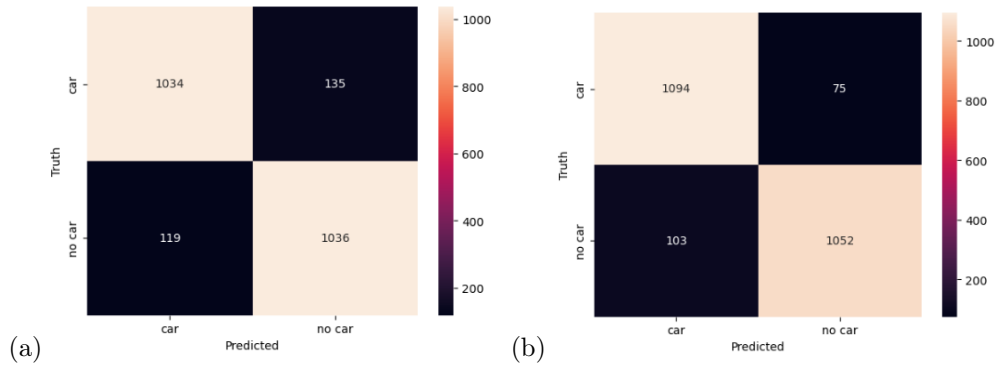
### 3.2 Wyniki eksperymentów

Porównując obie sieci neuronowe można stwierdzić, że klasyfikator A na zbiorze testowym dobrze zklasyfikował 488 na 582 obrazów, natomiast klasyfikator B 507 na 582 obrazów. W trudniejszych warunkach na zbiorze danych DOTA klasyfikator A dobrze zklasyfikował 78 na 114 obrazów, a klasyfikator A 74 na 114 obrazów. Warto jednak zauważyć, że klasyfikator A wykrył tylko 7 samochodów na 40, oraz 3 błędnie, natomiast klasyfikator B 24. Niedokładność klasyfikatora B polega głównie na klasyfikacji wycinka na którym nie ma samochodu jako taki, który je zawiera. Taka sytuacja występuje głównie przy wycinkach zawierających samoloty lub obiekty przypominające kształtem samochody, czego przykład widać na rys. 8.

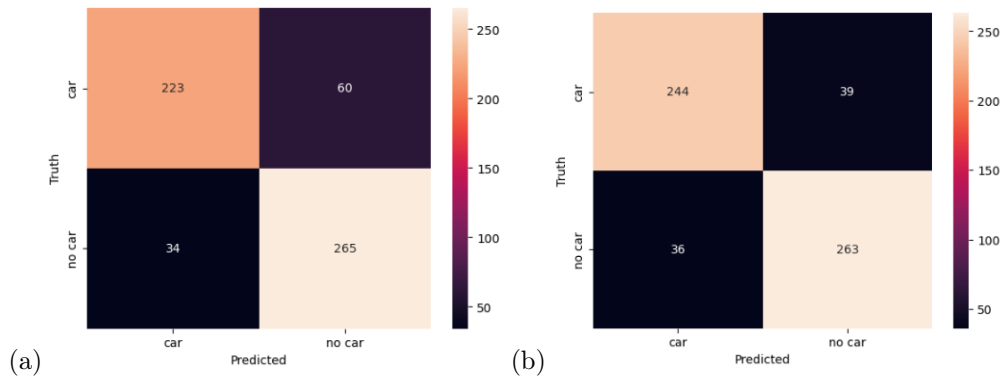
Macierze błędów dla danych treningowych przedstawione są na rys. 5.

Macierze błędów dla danych testowych przedstawione są na rys. 6.

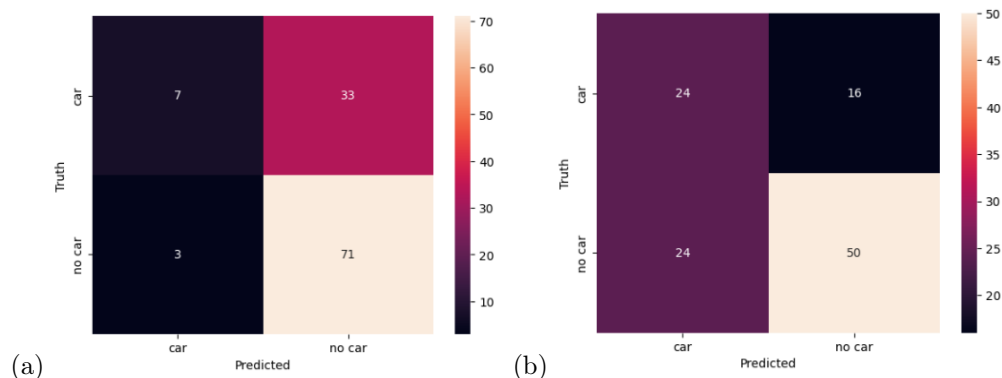
Macierze błędów dla danych ze zbioru danych DOTA przedstawione są na rys. 7.



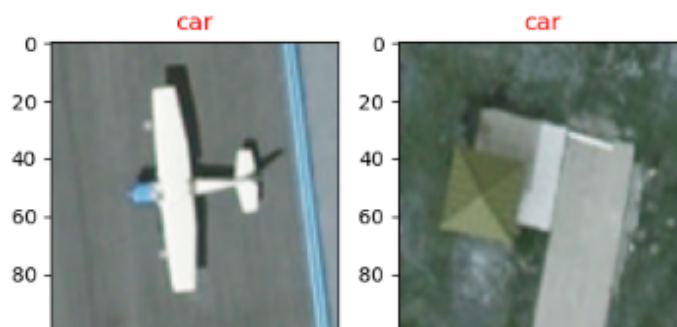
Rysunek 5: Macierze błędów dla danych treningowych: (a) Klasyfikator A (b) Klasyfikator B



Rysunek 6: Macierze błędów dla danych testowych : (a) Klasyfikator A (b) Klasyfikator B



Rysunek 7: Macierze błędów dla danych ze zbioru danych DOTA: (a) Klasyfikator A (b) Klasyfikator B



Rysunek 8: Przykłady złych predykcji na zbiorze danych DOTA.

## 4 Wnioski

Zaproponowana przeze mnie sieć z jedną warstwą ukrytą dokonuje klasyfikacji na zbiorze danych, na których była uczona w dopuszczającym stopniu, natomiast w trudniejszych warunkach na zbiorze danych DOTA sieć ta nie radzi sobie z zadaniem klasyfikacji. W porównaniu z nią, sieć z trzema warstwami ukrytymi sprawuje się nieco lepiej na zbiorze danych, na których się uczyła, natomiast nieco gorzej na danych ze zbioru DOTA. Warto jest jednak zaznaczyć kontekst, w którym klasyfikator ten się myli, czyli wykrywanie samochodów wtedy, gdy na wycinku obrazu znajduje się samolot lub inny obiekt o kształcie przypominającym samochód. Można więc stwierdzić, że dodanie dwóch warstw ukrytych poprawiło jakość klasyfikatora.

## Literatura

- [1] Ding, Jian and Xue, Nan and Xia, Gui-Song and Bai, Xiang and Yang, Wen and Yang, Michael and Belongie, Serge and Luo, Jiebo and Datcu, Mihai and Pelillo, Marcello and Zhang, Liangpei: Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges. IEEE

- 
- Transactions on Pattern Analysis and Machine Intelligence **10.1109/TPAMI.2021.3117983**, 1–1 (2021)
- [2] Xia, Gui-Song and Bai, Xiang and Ding, Jian and Zhu, Zhen and Belongie, Serge and Luo, Jiebo and Datcu, Mihai and Pelillo, Marcello and Zhang, Liangpei: DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June (2018)
- [3] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, Qikai Lu: Learning RoI Transformer for Detecting Oriented Objects in Aerial Images. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June (2018)