

Rozpoznawanie cech i detekcja twarzy

Celem zadania było wytrenowanie i zbadanie dwóch modeli opartych na sieciach splotowych. Jeden z nich ma za zadanie identyfikować płeć osoby na zdjęciu, zaś drugi określać obecność lub też nie okularów na twarzy człowieka.

1. Klasyfikator płci

Architektura modelu:

```
GenderCNN(  
  (cnn): Sequential(  
    (0): Conv2d(3, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
    (1): ReLU()  
    (2): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)  
    (3): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
    (4): ReLU()  
    (5): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)  
  )  
  (fc): Sequential(  
    (0): Linear(in_features=16384, out_features=256, bias=True)  
    (1): ReLU()  
    (2): Linear(in_features=256, out_features=1, bias=True)  
  )  
)
```

Rys. 1 Architektura sieci GenderCNN

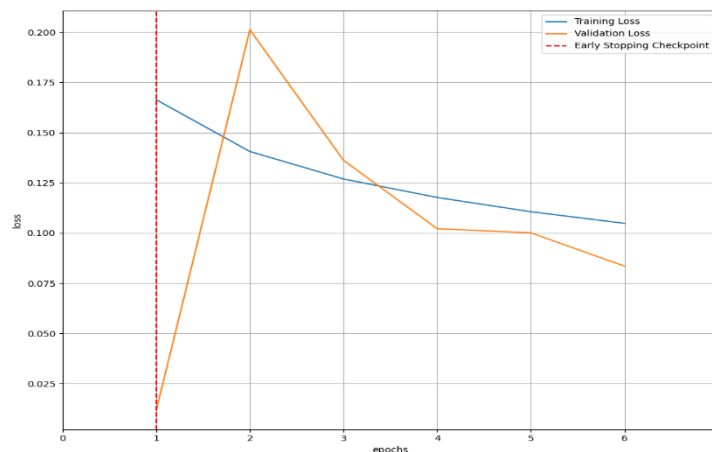
Sieć przyjmuje na wejściu obraz przekształcony do rozmiaru 64x64 zapisany w 3 kanałach, zaś jako wyjście zwraca jedną liczbę, która po interpretacji pozwala na określenie klasy danego obrazu.

Sieć składa się z dwóch części: pierwszą tworzą sekwencje składające się z warstwy splotowej o rozmiarze jądra 3x3 z funkcją aktywacji ReLu oraz warstwy poolingowej, zaś druga stanowi połączenie warstw liniowych i ReLu.

Do treningu wykorzystano funkcję straty BCEWithLogitsLoss (Binary Cross Entropy with Logits Loss), która jest powszechnie stosowana w klasyfikacji binarnej oraz dla wyjść z sieci, które niekoniecznie są prawdopodobieństwami. W porównaniu z entropią krzyżową wykorzystaną do wyżej przedstawionej architektury, wyniki uzyskane przez BCEWithLogitsLoss były zdecydowanie dokładniejsze i nauka przebiegała bardziej efektywnie.

Zaś jako optymalizator wybrano optymalizator Adam (Adaptive Moment Estimation), opartą na metodzie spadku gradientu, która także jest z powodzeniem stosowana w klasyfikacji binarnej. Zaletą tego rozwiązania jest adaptacja szybkości uczenia się dla każdego parametru indywidualnie.

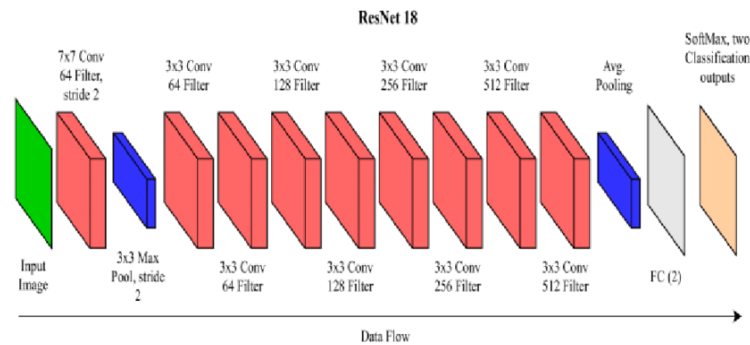
Aby zapobiec przeuczeniu wykorzystano mechanizm early stopping na zbiorze walidacyjnym. Podejście to pozwala na porzuceniu nauki, kiedy model nie wykazuje postępów.



Rys. 2 Wykres zmian wartości funkcji celu na zbiorach

2. Klasyfikator obecności okularów

Jako drugi model zaadaptowano sieć ResNet18, który dodatkowo został wzbogacony o warstwę liniową na wyjściu pozwalającą na przewidzenie binarnej etykiety. Sieć ta składa się z 18 warstw, w których skład wchodzi warstwy splotowe, łączące i w pełni połączone.



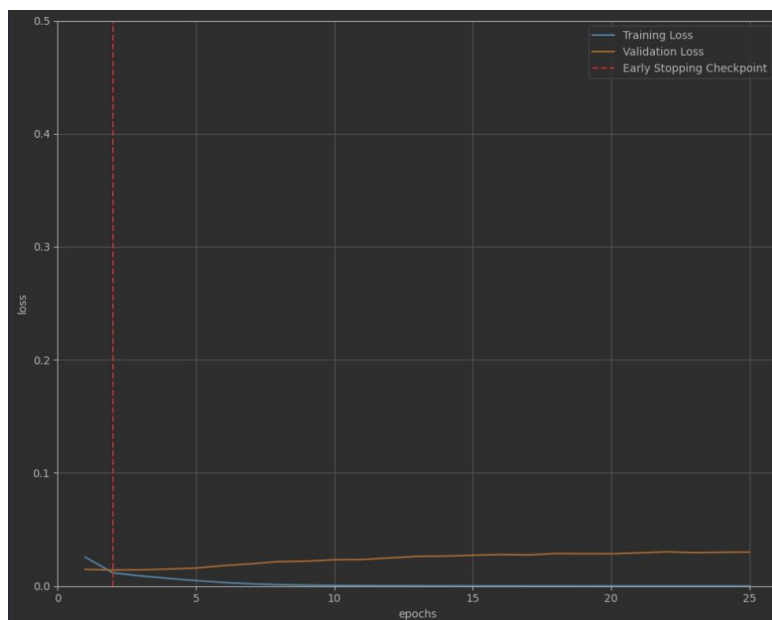
Rys. 3 Architektura resnet18

Do trenowania wykorzystano technikę fine tuning, dotrenowując model na danych zbioru celebA. Zastosowano również early stopping, aby powstrzymać model przed nadmiernym dopasowaniem do danych treningowych.

W tym przypadku również wykorzystano funkcję straty BCEWithLogitsLoss, zaś optymalizator SGD (Stochastic Gradient Descent), który jest wydajny obliczeniowo, szczególnie w przypadku dużych zbiorów danych, z jakim mamy do czynienia w tym przypadku.

Dane na wejściu sieci zostały odpowiednio przekształcone do rozmiaru 244x244. Zaś jako wyjście sieć zwraca jedną liczbę, która po przetworzeniu przez funkcję sigmoidalną daje prawdopodobieństwa przynależności do klas.

W przypadku atrybutu "glasses" zbiór CelebA okazał się być niezbilansowany – podzbiór treningowy zawiera 10521 zdjęć z i 152249 obiektów bez okularów. Aby pozbyć się tego problemu zwielokrotniono 15-krotnie podzbiór z okularami, otrzymując tym samym 168336 obrazów ludzi noszących okulary.

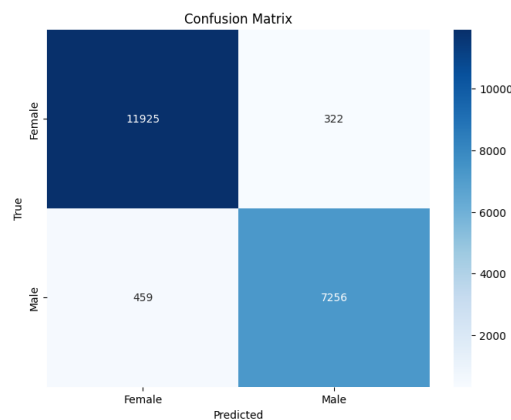


Rys. 4 Wykres zmian wartości funkcji celu na zbiorach

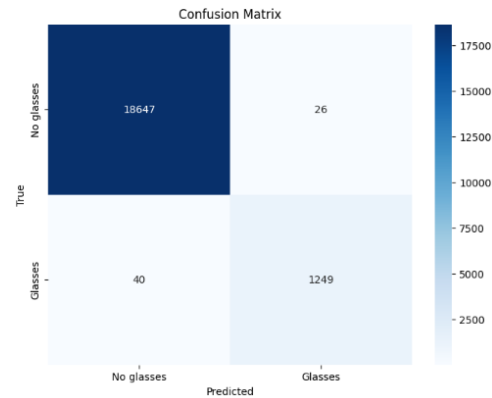
3. Wyniki testów na zbiorze CelebA

Na zbiorze testowym CelebA klasyfikator płci uzyskał wartość Accuracy o wartości 96.09%, zaś model bazujący na sieci ResNet18 osiągnął dokładność 99.67%.

Poniżej przedstawiono macierz pomyłek:

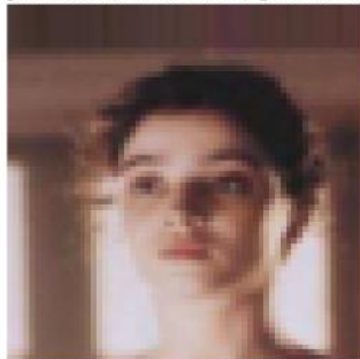


Rys. 5 Macierz pomyłek dla modelu GenderCNN



Rys. 6 Macierz pomyłek dla modelu ResNet18

Real gender: Female Predicted gender: Female



Real gender: Male Predicted gender: Female



Rys. 7 Przykłady klasyfikacji modelu GenderCNN

Real eyeglasses: No glasses Predicted eyeglasses: No glasses



Real eyeglasses: Glasses Predicted eyeglasses: No glasses



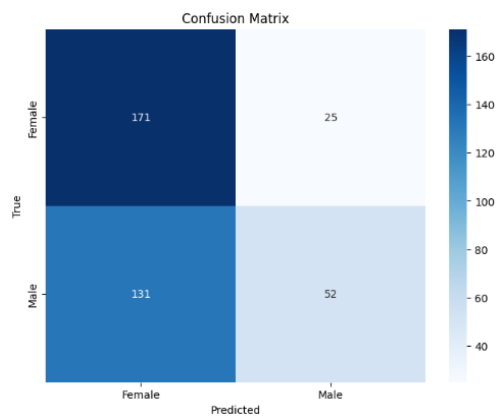
Rys. 8 Przykłady klasyfikacji modelu resnet18

Oba modele uzyskały satysfakcjonujące wyniki, jednakże model resnet18 posiadający zdecydowanie więcej warstw okazał się niemal niezawodny. Odpowiednie przekształcenie obrazu dla sieci resnet, na pierwszy rzut oka pozwala na łatwiejsze wyodrębnienie postaci od tła. W przypadku działania sieci

stworzonej od podstaw, większość przypadków źle zaklasyfikowanych nie jest jednoznaczna nawet dla człowieka.

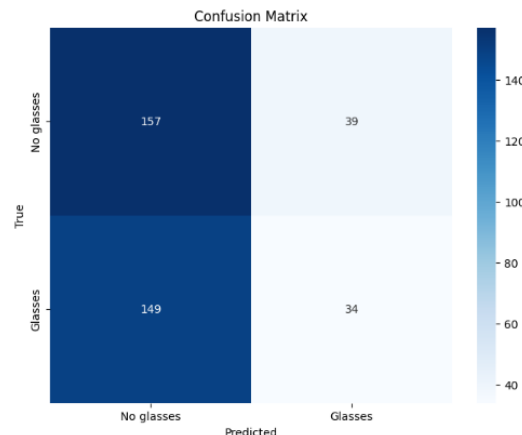
4. Wyniki testów na zbiorze WIDERFace

Na zbiorze testowym WIDERFace klasyfikator płci uzyskał wartość Accuracy o wartości 58.84%, zaś model bazujący na sieci ResNet18 osiągnął dokładność 50.40 %.



Rys. 9 Macierz pomyłek modelu GenderCNN

Real gender: ['Female', 'Female', 'Female']
Predicted gender: ['Female', 'Female', 'Female']



Rys. 10 Macierz pomyłek modelu resnet18

Real gender: ['Male', 'Male', 'Female']
Predicted gender: ['Female', 'Female', 'Female']



Rys. 11 Przykłady klasyfikacji modelu GenderCNN

Real eyeglasses: ['No glasses', 'No glasses', 'Glasses']
Predicted eyeglasses: ['Glasses', 'Glasses', 'Glasses']



Real eyeglasses: ['No glasses', 'No glasses', 'No glasses', 'Glasses']
Predicted eyeglasses: ['No glasses', 'No glasses', 'No glasses', 'No glasses']



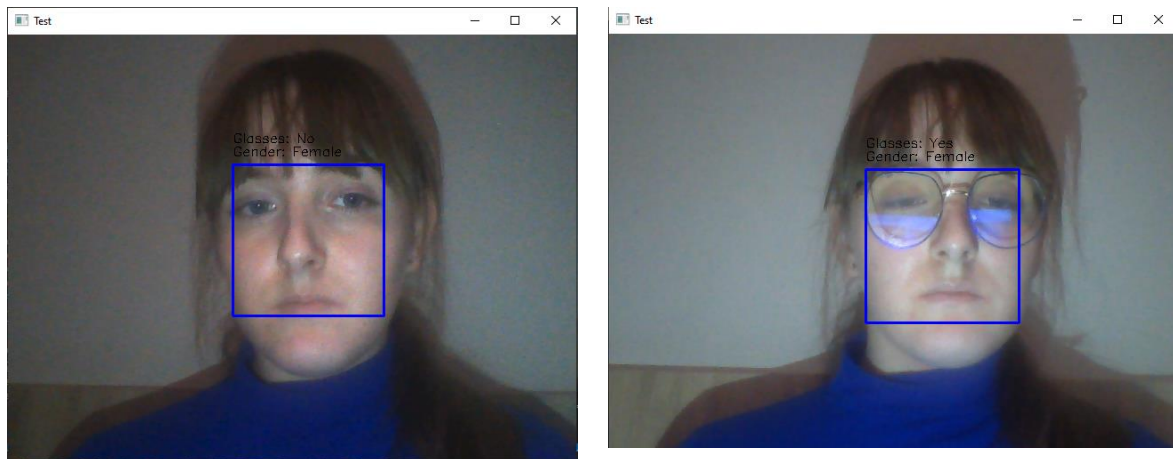
Rys. 12 Przykłady klasyfikacji modelu resnet18

Testowy zbiór WIDERFace powstał na bazie bibliotecznej bazy danych, ograniczając się jedynie do 102 zdjęć. Do każdej twarzy dodano adnotacje informującą o płci oraz posiadaniu okularów. Modele po kolei klasyfikowały poszczególne twarze, wynik interpretowany jest jako atrybut poszczególnych osób czytając od lewej strony.

Modele uzyskały dokładność porównywalną z losową klasyfikacją, rzędu 50%. Przyczynę można się doszukiwać w wielkości wyciętych twarzy podanych na wejściach sieci. Wycinki mogły okazać się zbyt małe, które dodatkowo po przekształceniu mogły stracić część ważnych informacji i tym samym powodować błędną klasyfikację. Również zbiór treningowy zawierał więcej obrazów kobiet, co mogło też wpłynąć na faworyzowanie tej grupy podczas testowania.

5. Wyniki z programu testowego

Weryfikacja skuteczności działania modeli zastosowanych do detekcji cech postaci z obrazu odczytanego z kamery dokonana jest jedynie w sposób wizualny. Poniżej zamieszczono kilka testów modeli.



Rys. 12 Przykładowe dane wyjściowe z programu testowego

Program w określonych odstępach czasu z obrazu kamery obiera obraz, który zostaje poddany jako wejście do modeli detekcji twarzy. Uzyskany wycinek ujęcia kolejno podany zostaje jako wejście do modeli. Powyżej przedstawiono prawidłowo oznaczone przykłady. Jednakże, podczas sprawdzania działania programu widać, że model dotyczący okularów potrzebował lepiej oświetlonego ujęcia kamery aby poprawnie określić atrybut. Z kolei model identyfikujący płeć jest bardziej wrażliwy na przesunięcia. Niewielka zmiana pozycji powoduje już błędną klasyfikację.