Model użyty jako podstawa drugiego modelu (który przewiduje czy osoba ma okulary) to torchvision.models.resnet18(weights='IMAGENET1K\_V1') czyli wagi i 18 warstwowy model opracowany w trakcie pracy Deep Residual Learning for Image Recognition przez Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Schemat tego modelu ale dla 34 warstwowego modelu z powyższego artykułu zamieszczony jest po lewej gdzie czarne strzałki reprezentują równanieObraz zawierający Czcionka, typografia, tekst, kaligrafia

Opis wygenerowany automatycznie

a przerywane   
Obraz zawierający Czcionka, typografia, kaligrafia, tekst

Opis wygenerowany automatycznie jeśli ilość wymiarów nie jest taka sama a jeśli nie zachowują się tak jak czarne strzałki.

Przed uczeniem zresetowana została ostatnia warstwa: fc (fully connected) 1000 - pokolorowana na biało. Jest również przestawiana na zwracanie 2 wartości. Dane wejściowe zostały przetransformowane z użyciem torchvision.transforms według schematu  
transforms.Compose([  
 transforms.Resize(245),  
 transforms.ToTensor(),  
 transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])  
 ])

Obrazy ze zbiory testowego przeszły również augmentacje poniższym kodem:  
transforms.Compose([  
 transforms.RandomResizedCrop(245),  
 transforms.RandomHorizontalFlip(),  
 transforms.ColorJitter(brightness=0.6),  
 transforms.ToTensor(),  
 transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])  
 ])

Rysunek 1 Schemat ResNet-34 Żródło: Deep Residual Learning for Image Recognition



Rysunek 2 Obrazy treningowe po augmentacji

Warto jednak zauważyć, że wszystkie użyte obrazy zostały też wykadrowane z użyciem detektora twarzy udostępnionego razem z zadaniem.

Dokładność dla modelu drugiego (okulary) po pierwszym epochu treningu na zbiorze CelebA wynosiła 0.9729 na zbiorze treningowym i 0.9954 na zbiorze walidacyjnym. Najwyższą dokładność na zbiorze walidacyjnym (0.9964) osiągnięto po epochu 13.

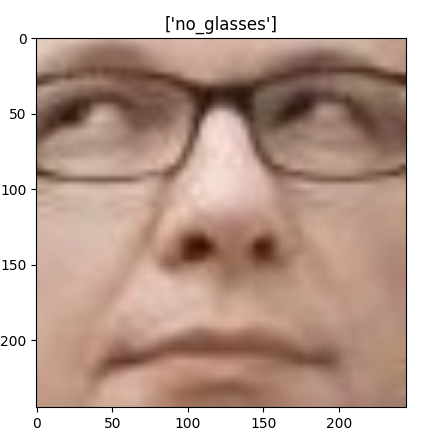
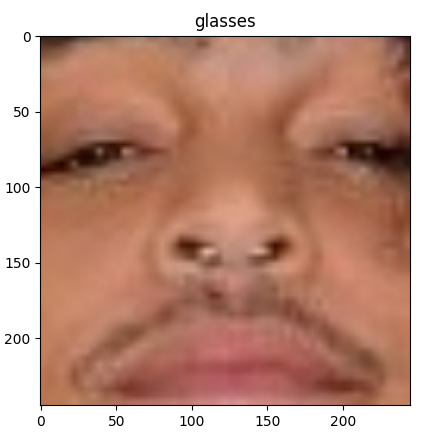
Dla zbioru testowego CelebA dokładność wyniosła 0,9965 a dla wybranych obrazów ze zbioru walidacyjnego wider (23 obrazy z okularami i 81 bez) 0,9904. Jako, że model nie był trenowany na obrazach z wider wszystkie przygotowane zdjęcia zostały użyte jako zbiór testowy.

Matryce błędów prezentują się następująco:

|  |  |  |  |
| --- | --- | --- | --- |
| wider | okulary | bez okularów | przewidziano |
| okulary | 22 | 1 |  |
| bez okularów | 0 | 81 |  |
| prawda |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| CelebA test | okulary | bez okularów | przewidziano |
| okulary | 969 | 44 |  |
| bez okularów | 20 | 17428 |  |
| prawda |  |  |  |

Trudno znaleźć cechę wspólną dla nieprawidłowo sklasyfikowanych obrazów jednak trudność w rozpoznawaniu osób z okularami może wynikać z mniejszej ilości ich zdjęć w danych treningowych (1823 z okularami i 31529 bez okularów)



Rysunek 3 Błędnie sklasyfikowany obraz ze zbioru testowego CelebA

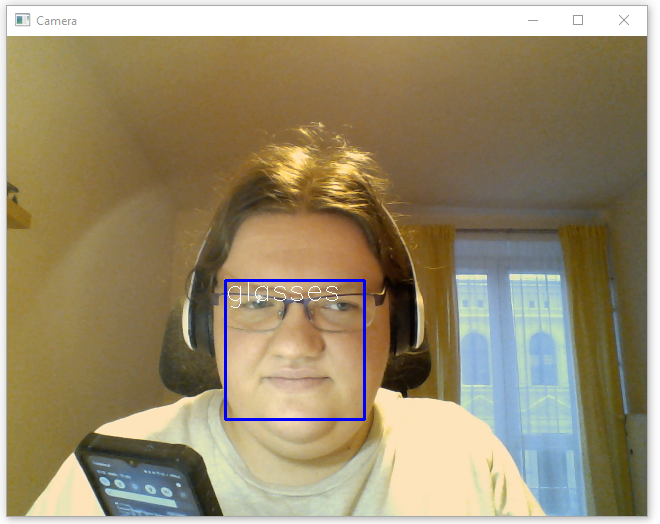
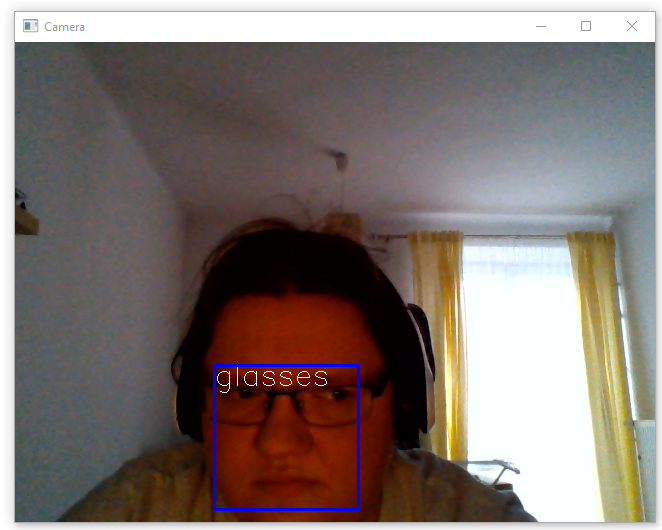
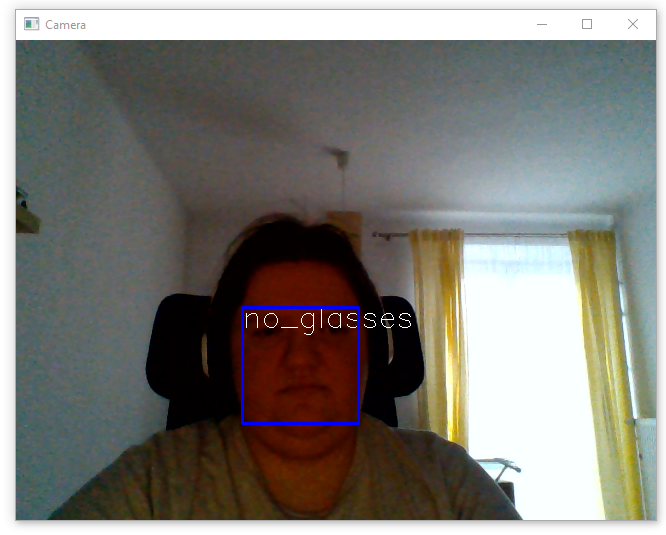
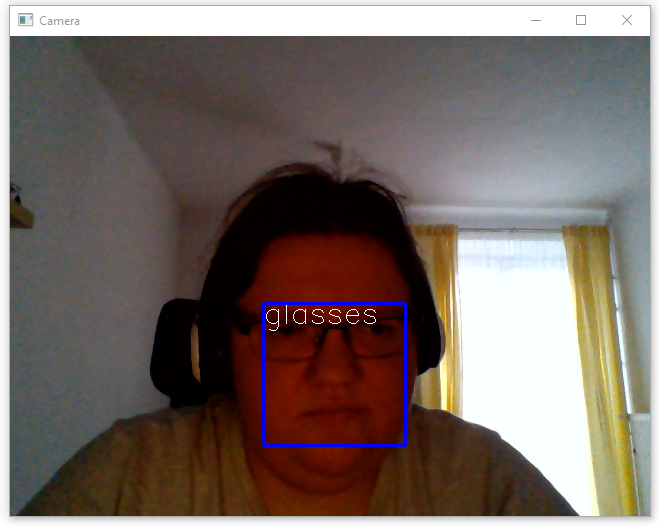
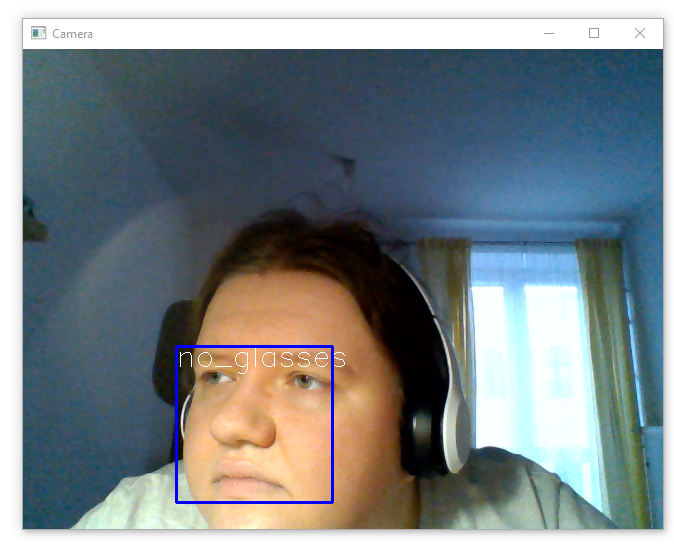
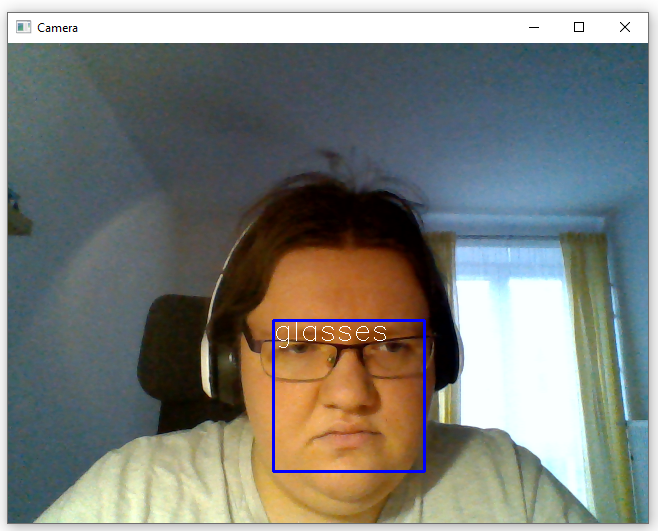
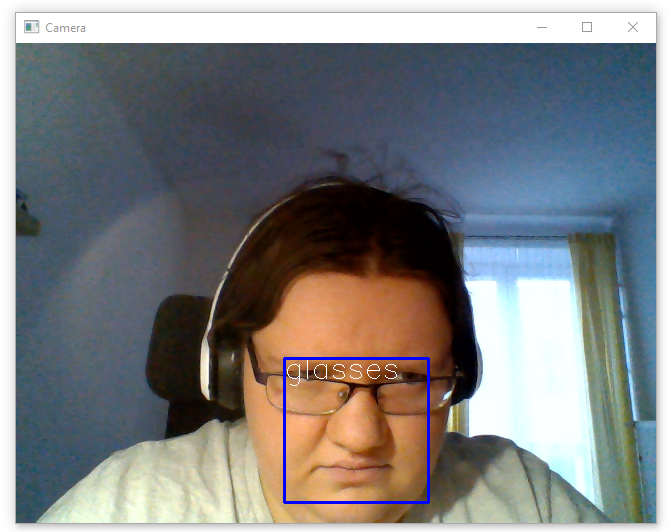
Obraz zawierający tekst, Ludzka twarz, uśmiech, skóra

Opis wygenerowany automatycznie

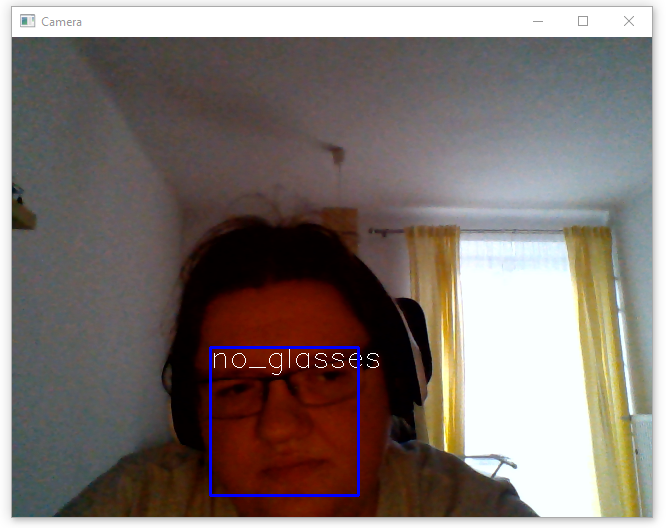
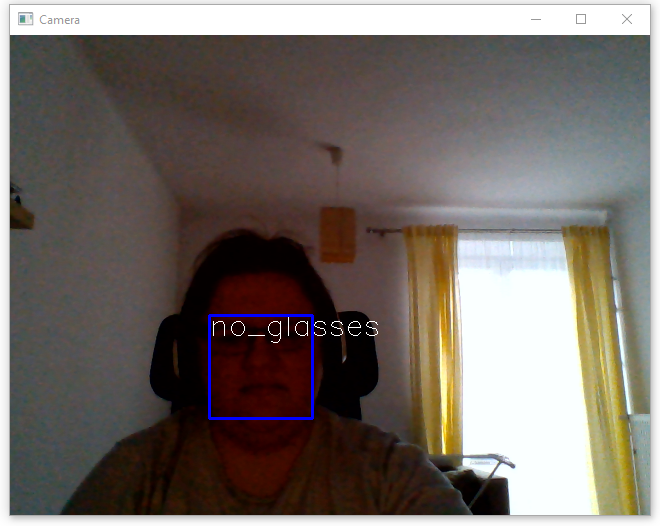
Rysunek 4 Błędnie sklasyfikowany obraz ze zbioru testowego wider

Nie trenowaliśmy modelu wykrywającego twarze: użyliśmy tego udostępnionego z zadaniem.

W testach z kamerą program radzi sobie dobrze, największe trudności ma w warunkach niskiego oświetlenia, gdzie ma tendencję przypisywania osobie braku okularów. Testy przeprowadzone zostały po 15:00 w zimie przy sztucznym oświetleniu wewnątrz.



Rysunek 5 Poprawne rozpoznania modelu 2



Rysunek 6 Niepoprawne rozpoznania modelu 2