Samodzielnie zaimplementowany model składa się z trzech warstw konwolucyjnych. Po każdej z nich następuje funkcja aktywacji ReLU, aby zapewnić nieliniowość, po której z kolei wykorzystano MaxPool2d, zmniejszający o połowę rozmiar mapy cech, co oprócz wpływu na czas obliczeń zmniejsza prawdopodobieństwo przetrenowania modelu. Każda kolejna warstwa konwolucyjna generuje dwukrotnie więcej map cech nich poprzednia (pierwsza z nich generuje 32).

Wartwy połączone zaczynają od spłaszczenia tensora, będącego wynikiem obliczeń modułu konwolucyjnego sieci. Następnie, warstwa liniowa zmniejsza wymiar spłaszczonego tensora do 256. W kolejnym kroku wywoływana jest ReLU, po której wywoływany jest Dropout. Jego celem jest odrzucenie losowo wybranej połowy neuronów. Jest to kolejne zabezpieczenie przed przetrenowaniem modelu – zmniejsza poleganie modelu na pojedynczych cechach. W ostatnim kroku, wyekstrahowana zostaje jedna cecha, która następnie jest konwertowana do prawdopodobieństwa wystąpienia tej cechy w danym przypadku.

Jako funkcję celu wybrano Binary Cross Entropy Loss, która, jak nazwa wskazuje, dobrze nadaje się do klasyfikacji binarnej. Co więcej do trenowania użyto optimizera Adam, aby zmniejszyć czas trenowania.

Jako dodatkowe zabezpieczenie przed przetrenowaniem zastosowano Early stopping, który jest wywołany przy przekroczeniu cierpliwości (ustawiona granica to 3 i zwiększa się za każdym razem, gdy strata na zbiorze walidacyjnym jest mniejsza od najlepszej uzyskanej do tej pory). W przypadku zaproponowanego modelu early stop zostaje aktywowany po zakończeniu trenowania 8 epoki.

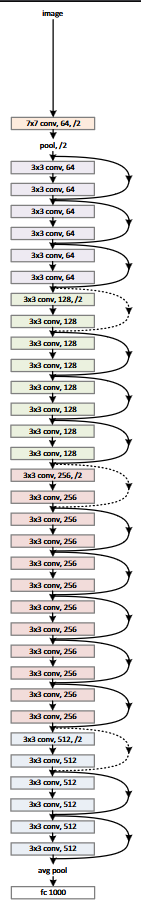
Wykaz przykładowych zdjęć z procesu treningu i testowania na końcu raportu.

Obraz zawierający tekst, zrzut ekranu, diagram, Wielobarwność

Opis wygenerowany automatycznieObraz zawierający tekst, zrzut ekranu, Prostokąt, diagram

Opis wygenerowany automatycznie

Rysunek Wyniki modelu 1 (z lewej celebA - accuracy 98,1%, z prawej widerface - accuracy 76,9%)

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



Rysunek 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.

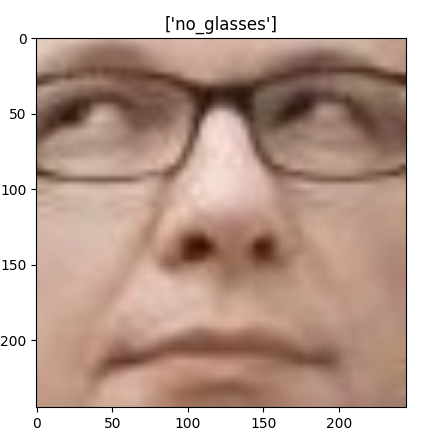
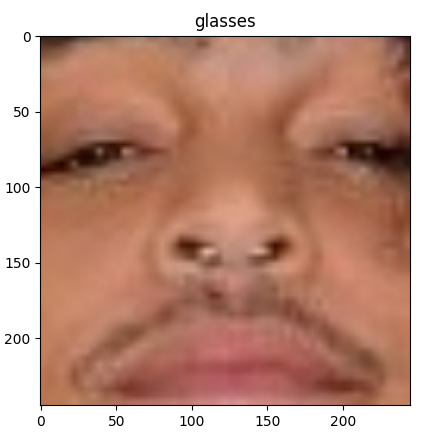
Dla zbioru testowego CelebA dokładność wyniosła 0,9965 a dla wybranych obrazów ze zbioru walidacyjnego wider (23 obrazy z okularami i 81 bez) 0,9904. Jako, że model nie był trenowany na obrazach z wider wszystkie przygotowane zdjęcia zostały użyte jako zbiór testowy.

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 Błędnie sklasyfikowany obraz ze zbioru testowego CelebA

Obraz zawierający tekst, Ludzka twarz, uśmiech, skóra

Opis wygenerowany automatycznie

Rysunek 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.

Dla modelu 1 testy zostały przeprowadzone na zdjęciach wyświetlonych na monitorze komputera z włączonym trybem nocnym.

Obraz zawierający Ludzka twarz, osoba, okulary/szklanki, ubrania

Opis wygenerowany automatycznieObraz zawierający osoba, ubrania, Ludzka twarz, Warga

Opis wygenerowany automatycznieObraz zawierający Ludzka twarz, uśmiech, osoba, brew

Opis wygenerowany automatycznieObraz zawierający osoba, uśmiech, Ludzka twarz, ubrania

Opis wygenerowany automatycznie

Rysunek Błędnie sklasyfikowane zdjęcia dla modelu 1 - z lewej CelebA, z prawej WIDER

Obraz zawierający Ludzka twarz, zrzut ekranu, skóra, osoba

Opis wygenerowany automatycznieObraz zawierający Ludzka twarz, osoba, człowiek, zrzut ekranu

Opis wygenerowany automatycznieObraz zawierający Ludzka twarz, zrzut ekranu, kobieta, osoba

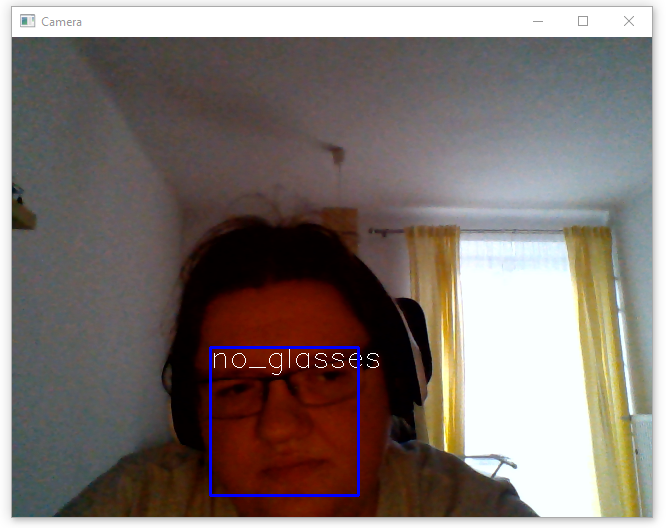
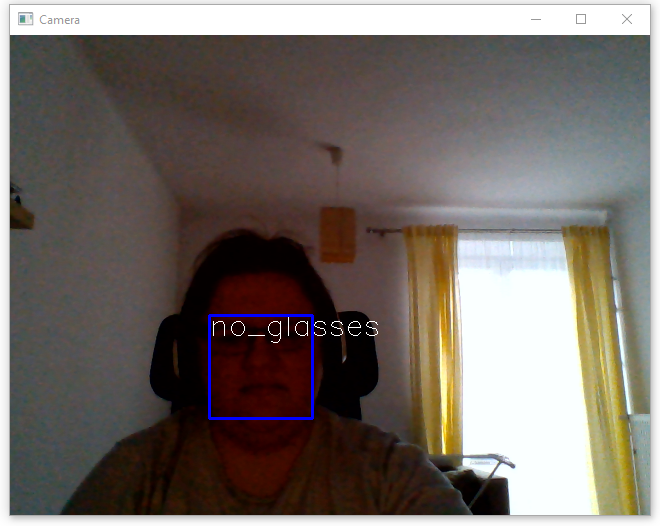
Opis wygenerowany automatycznieObraz zawierający Ludzka twarz, rzęsa, brew, zrzut ekranu

Opis wygenerowany automatycznieObraz zawierający ściana, osoba, Ludzka twarz, człowiek

Opis wygenerowany automatycznieObraz zawierający Ludzka twarz, brew, Czoło, szczęka

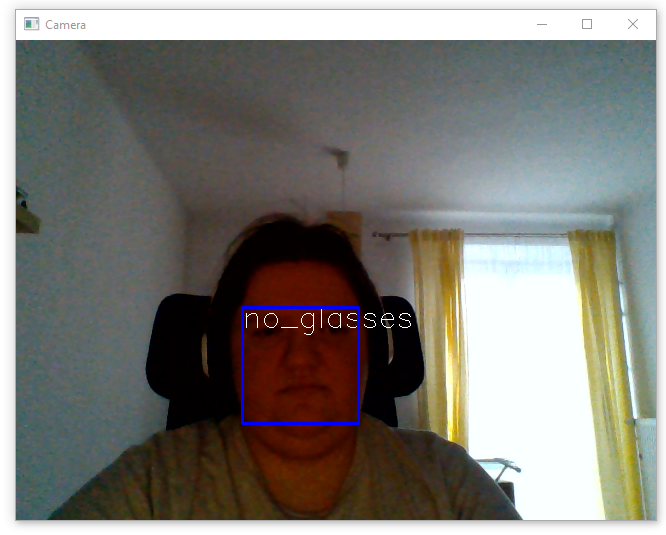
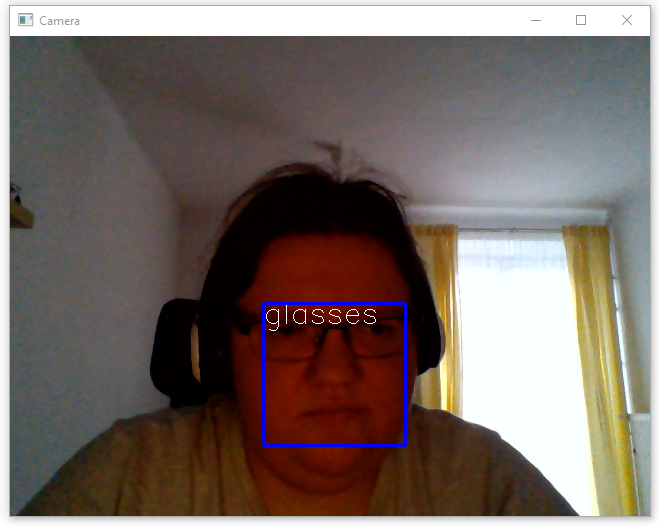
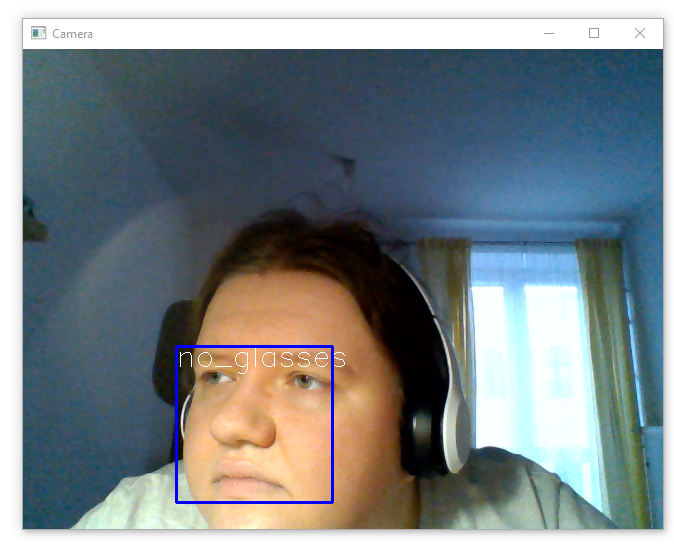
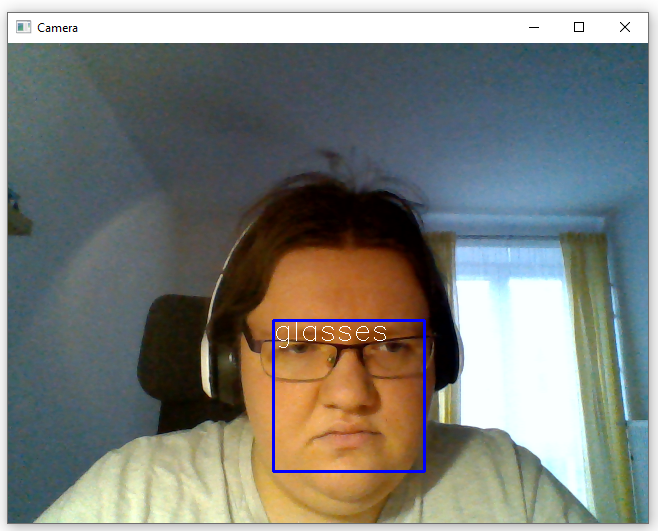
Opis wygenerowany automatycznie

Rysunek Niepoprawne rozpoznania modelu 1



Rysunek Niepoprawne rozpoznania modelu 2

Rysunek Poprawnie rozpoznania modelu 1



Rysunek Poprawne rozpoznania modelu 2