Wykrywanie środka twarzy na zdjęciu

Michał Janik

8 listopada 2020

1 Problem

Skontruować model do przewidywania współrzędnych (x,y) na zdjęciu, które oznaczą środek twarzy osoby na nim się znajdującej.

2 Dataset

Do wykonania tego zadania wybrałem dataset **CelebA**. Zawiera ponad 200 tysięcy obrazów twarzy ludzich. Co ważne, zawiera zdjęcia ludzkich twarzy w różnych pozycjach i środki twarzy nie pokrywają się ze środkiem zdjęć. Nie zawiera bezpośrednio zlabelowanego środka twarzy, posłużymy się zatem **współrzędną nosa**, która dobrze ją odzwierciedla.
Współrzędne x i y normalizujemy: dzielimy dla każdego zdjęcia przez jego szerokość i wysokosc tak, aby $x, y \in [0, 1]$.
Dataset dzielimy na train/valid/test sety, odpowiednio w proporcjach 8:1:1. Stosuję augmentację obrazów poprzez zmianę jasności obrazów(nie używam standardowych opcji jak cropowanie/rotacja/zoomowanie, ponieważ zmieniałoby to punkt środka twarzy).

3 Rozwiązanie

3.1 Haar Cascades

Wykrywanie obiektów przy użyciu kaskadowych klasyfikatorów Haar jest skuteczną metodą wykrywania obiektów zaproponowaną przez Paula Violę i Michaela Jones aw 2001 roku. Jest to podejście oparte na uczeniu maszynowym, w którym funkcja kaskadowa jest szkolona z wielu pozytywnych i negatywnych obrazów. Jest ona następnie wykorzystywana do wykrywania obiektów na innych obrazach.
Wykorzystuję wbudowany w bibiotekę 'opencv' wytrenowany klasyfikator 'haarcascade_frontalface_default.xml'. Zdjęcia podajemy w bazowej rozdzielczości.

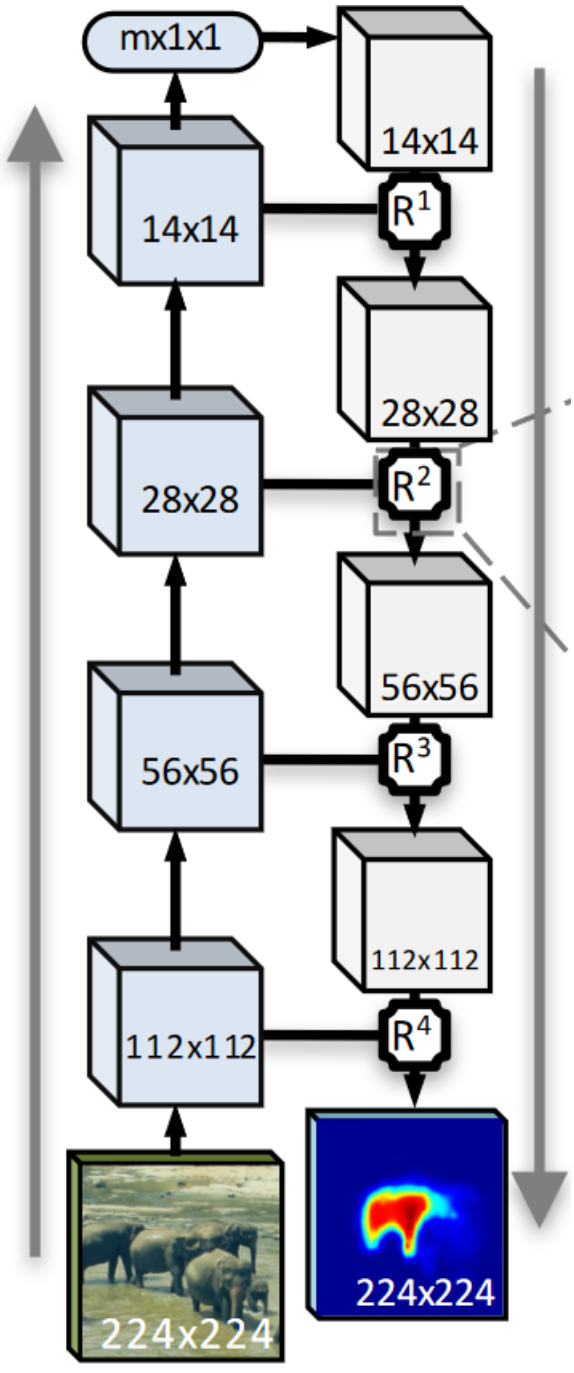
3.2 Prosty model bazujący na ResNet

Głębokie sieci konwolucyjne są SOTA w analizie obrazu, więc były oczywistym wyborem do tego problemu. Jako 'feature extractor' użyłem pretrenowowanego modelu ResNet50V2 (1). Finalny model nakłada na Resnet(bez warstw FC) GlobalPooling, a nastepnie warstwy FC. Ostatnia warstwa składa się z dwóch neuronów, na które nałożone jest funkcja aktywacyjna 'sigmoid'.

3.3 Bottom-up/Top-down

Problemem wcześniejszego modelu jest łatwa utrata informacji o lokalizacji środku twarzy przez nakładanie GlobalPoolingu. Potencjalnym rozwiązaniem jest zastosowanie architektury 'bottom-up/top-down', czyli serii warstw

konwolucyjnych stopniowo zmniejszających wymiary(szerokość i wysokość) pośrednich reprezentacji, po której występuje seria warstw dekonwolucyjnych zwiększających wymiary. Bazując na pracach (2), (3), i (4) stosujemy 'skip-connections': kolejne wyniki warstw konwolucyjnych są dołączane do wyników warstw dekonwolucyjnych jak na rysunku:



Wyjściem jest mapa o wymiarach jak obraz wejściowy, która reprezentuje 'rozkład środka twarzy' (0 - tło, środka twarzy nie ma; 1-twarz). Dane wyjściowe byłyby wtedy jednak zbyt niezbalansowane(prawie same 0), zatem w niewielkim otoczeniu środka twarzy również ustawiamy wyjście na 1. Aby dostać współrzędne centrum twarzy, wybieramy takie x i y, dla których odpowiedni neuron o współrzędnych x i y ma największy output.

4 Trenowanie

- Optimizer: Adam($beta_1 = 0.9$, $beta_2 = 0.999$,),
- Learning rate: $5 \cdot 10^{-4}$,
- Dropout: 0.5

  Modele przez 50 epok trenujemy z zamrożonymi wagami Resnet'a, a następnie dokonujemy fine-tuning z odmrożonymi 20 ostatnimi warstwami przez dodatkowe 20 epok. Dla modelu 3.2 jako loss function wziąłem prosty MSE. Dla modelu 3.3, aby zminimalizować problem niezbalansowania użyłem Focal Loss (5).

5 Wyniki

MSE na test-set		
Haar	prosta głęboka sieć	bottom-up/top-down
0.0184	0.0040	0.0019

Literatura

[1] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian Identity Mappings in Deep Residual Networks. arXiv:1612.03144, 2016.

[2] Olaf Ronneberger and Philipp Fischer and Thomas Brox U-Net: Convolutional Networks for Biomedical Image Segmentation [*On the electrodynamics of moving bodies*]. arXiv:1505.04597, 2015.

[3] Tsung-Yi Lin and Piotr Dollár and Ross Girshick and Kaiming He and Bharath Hariharan and Serge Belongie Feature Pyramid Networks for Object Detection arXiv:1612.03144, 2017.

[4] Pedro O. Pinheiro and Tsung-Yi Lin and Ronan Collobert and Piotr Dollàr Learning to Refine Object Segments arXiv:1603.08695, 2016.

[5] Tsung-Yi Lin and Priya Goyal and Ross Girshick and Kaiming He and Piotr Dollár Focal Loss for Dense Object Detection arXiv:1708.02002, 2018.