EmotiveLens: face detection with focal loss training criterion and small data

Kevin Song[1] & Anthony Van[2]
[1]Stanford Biomedical Informatics Graduate Program, [2]Stanford Department of Electrical Engineering
(kmsong@stanford.edu, avan994@stanford.edu)

Summary

Computer-vision-based facial emotional classifiers have valuable applications in telemedicine (specifically, psychiatry and mental healthcare in the digital age). In addition, computer vision innovations have not been readily applied to the field of telemedicine.

Here, we proposed and *partially* built a webcam-enabled emotional expression classifier by fully implementing our initial object detector (i.e., by training a ResNet CNN with a novel focal loss criterion). We hope our open source implementation exists as a proof-of-concept that leads to potential exploration and optimization of this technology space for future development.

Data

Object detection dataset: frontal face images (N = 450, each 896 x 592 pixels in jpeg format) were collected by Markus Weber at Caltech. Color images were collected of 27 unique people under various lighting, expressions, and backgrounds. Images were previously annotated with boundary box coordinates around individual faces.

Emotional classification dataset (not used): The AR Face Database (collected by Aleix Martinez and Robert Benavente at Ohio State) consisted of over 4,000 color images (of 768 x 576 dimensions) from 126 people. Images were previously annotated according to the following feature labels: neutral expression, smile, anger, scream, lighting conditions, and presence/absence of occlusions.

Features

Our algorithms' features were RGB pixel values from images fed as inputs to our CNNs.

Models

Object detection task: we trained a RetinaNet model (ResNet trained with focal loss) on our object detection dataset. Focal loss was previously proposed[1] by Lin, et al. at Facebook AI Research.

Emotional classification task (not performed): we proposed training a VGG-19 CNN with ImageNet weights on our emotional classification dataset, freezing the first five layers. Training would have been performed one-versus-all, such that labeled smiling images would be considered as positive cases, and non-smiling images would considered as negative cases.

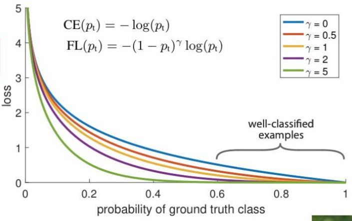

Figure 1: Focal loss. Our RetinaNet model was trained with an additional factor added to the standard cross-entropy training criterion. As γ increases, more focus is put on harder, misclassified examples.[1]

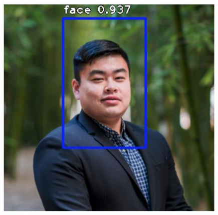

Figure 2: Boundary boxing. Our object detector draws boundary boxes around faces detected in images and assigns a probability of the boxed region as belonging to the "face" class.

Results

Object detector performance: we obtained a mean training accuracy of 94.3% (IoU, N = 450).

Discussion

"If you do it right 51% of the time you will end up a hero." — Alfred P. Sloan

Our object detector did not have extraordinarily great performance. Our primary roadblock in our project was sourcing boundary-box-annotated data that was properly formatted for loading into the RetinaNet model. A larger dataset was publicly available (FDDB, N > 5000), though its boundary box annotations were presented as elliptical coordinates. These coordinates, when trigonometrically converted to boundary box dimensions, did not lead to viable classifier performance. As our sample size was super-small, we did not expect substantially high performance or accuracy. However, our object detector's usage of a focal loss training criterion could have contributed to our observed training accuracy, given the small size of our training set.

Future

Given another six months, access to more time, and larger, better data, we would have ideally been able to implement this project end-to-end. Computational power was not an issue for our project, though it may be a larger issue when training on larger datasets, or for more epochs. With larger, more representative training sets, we would have theoretically achieved better model performance.

References

1. T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," arXiv:1708.02002 [cs], Feb. 2018.