Facial feature detection with Deep Learning

Juan Sebastián Cerón
Universidad de los Andes
Cra 1 Nº 18A - 12, Bogotá - Colombia
js.ceron10@uniandes.edu.co

Daniel Rodriguez
Universidad de los Andes
Cra 1 Nº 18A - 12, Bogotá - Colombia
da.rodriguez1253@uniandes.edu.co

Abstract

In this study we use deep learning to detect facial features. We try different network architectures with varied success. We comment on the pros and cons of our approach and advise on future experiments.

1. Introduction

Deep Learning marked the beginning of new problems in Computer Vision. The times were we had been working on identifying faces on images are on the past[1]. Now we want to be able to identify facial features as subtle as a dirty smile, or a 5'o clock shadow. Neural networks allow us to carry this goal by learning a tremendous amount of parameters that are specifically selected with the sole goal of better feature detection. One can identify several landmark networks that correspond to the incorporation of new architectural ideas such as full connections, depth of layers and dense interactions. Here, we use the celebrated ResNet network[2], which was developed to classify the ImageNet Database, to identify facial features. We also

2. Images used

The images used come from the Celeb dataset. It consists has 194k images containing 546k faces in total, covering 2622 labeled celebrities.

3. Approach

First, we employed ResNet 18 with pretrained weights as our baseline. Since ResNet was developed for ImageNet its last layer consists of 1000 outputs. To overcome this difficulty we added a new layer add the end so that it only produce 10 labels. However, Since, ResNet18 was not feasible with our computational resources, we used very simple networks as baseline. FC, ReLU, FC, ReLU, FC.

Class	Number of Identifications
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	height

3.1. Metric

Since we are on a MultiLabel Classification we need a specific loss function that is not mutually exclusive between labels. Here we used Binary Cross Entropy with Logits because it proved faster that Binary Cross Entropy. We employed only 2 functions although there are currently many in use.

4. Experiments and discussion

Our initial weights were selected so that null identification was the rule and not the exception. Hence, it is informative to see how many facial features the network identified on each specific class.

5. Conclusions

Facial Features are complex patterns that require social acuity to be understood. Neural Networks provide a first *real* approximation to this problem thanks to the many number of parameters and high quality images on new databases. We derive two main conclusions from our succesful and unsuccesful experiments. 1) Landmark networks were keenly adapted to solve specific database challenges. Although they perform very well on specific tasks, the complexity and flow of information between layers must be taken account when adapted to new classification problems. Also poor training on the new challenge, creates a highly complex network that is incapable of producing reasonable results. 2) Very simple networks with high number of parameters are able to mildy capture complex features. In particular, Convolutional Networks provide an excellent way to extract complex features from images. This shows that Deep Learning is a fabulous versatile approach with many more applications to come.

References

[1] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[2] L. Wang, W. Li, W. Li, and L. Van Gool. Appearance-and-relation networks for video classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1430–1439, 2018.