

Facial Keypoints Detection Using Convolutional Neural Network

Qinyao He
CST34
Tsinghua University
2012010548

Xiaocheng Yang
CST34
Tsinghua University
2013XXXXXX

Abstract

In this report we building an neural network for carrying out facial keypoints detection problem. We have tried several ways, but a problem of non convergence still need to be resolved, which till now we have no idea about that.

1. Introduction

This task comes originally from Kaggle¹, a online website holding many data science competitions. It's aim is to determine the location of several certain key points on face, given a photograph contains a human face.

This key points detection can form the building block for many applications. Such as:

- * tracking faces in images and video
- * analysing facial expressions
- * detecting dysmorphic facial signs for medical diagnosis
- * biometrics / face recognition

Detecting facial keypoints is a very challenging problem. Facial features vary greatly from one individual to another, and even for a single individual, there is a large amount of variation due to 3D pose, size, position, viewing angle, and illumination conditions. Computer vision research has come a long way in addressing these difficulties, but there remain many opportunities for improvement.

In this specific competition², we are required to localize 15 facial key points given a images. Each predicted point is specified as (x, y) real-valued pair in the space of pixel indices. The 15 keypoints is:

left_eye_center, right_eye_center, left_eye_inner_corner,
left_eye_outer_corner, right_eye_inner_corner,
right_eye_outer_corner, left_eyebrow_inner_end,

left_eyebrow_outer_end, right_eyebrow_inner_end,
right_eyebrow_outer_end, nose_tip, mouth_left_corner,
mouth_right_corner, mouth_center_top_lip,
mouth_center_bottom_lip

so 30 real value is going to be predicted for each input image.

Kaggle given a training data set of 7049 image with size of $96 * 96$, color is gray scale and represented as an integer from 0 to 255 for each pixel. And each training images is labeled with 30 real value for position of the 15 key points. In some examples, some of the target keypoint positions are missing.

Also, a test set of 1783 images without label is given, for competitors to submit their predicted result to Kaggle for score.

Performance is graded by MSE(mean square error) for those 30 real values. A lower MSE indicate a better prediction.

2. Related Work

3. Technical Approach

4. Network Structure

5. Experiment

6. Conclusion

References

- [1] A. Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002.
- [2] A. Alpher and J. P. N. Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003.
- [3] A. Alpher, J. P. N. Fotheringham-Smythe, and G. Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.
- [4] Authors. The frobnicable foo filter, 2014. Face and Gesture submission ID 324. Supplied as additional material `fg324.pdf`.
- [5] Authors. Frobnication tutorial, 2014. Supplied as additional material `tr.pdf`.

¹<https://www.kaggle.com/>

²<https://www.kaggle.com/c/facial-keypoints-detection>