Facial Keypoints Detection with CNN

Assignment 2

Image Based Biometrics 2019/20, Faculty of Computer and Information Science, University of Ljubljana

Kristjan Reba

***Abstract*—We present a problem of facial keypoints detection and why it is important. We use a convolutional neural network and show that good results can be achieved on a benchmark dataset. The root mean squared error achieved on a test set is equal to 2.1 pixels.**

I. Introduction

In recent years convolutional neural networks (CNN) [1], [2] have been able to surpass previous state of the art results on various data sets and problem domains [3]. In this paper we apply a CNN model to build a facial keypoints detector. Our goal is to show how simple it is to use CNN's and achieve satisfactory results on complex problems. Robust and accurate keypoints detection algorithm can greatly improve the accuracy of applications such as tracking faces in video, analysing facial expressions and biometric face recognition. Previous approaches for detecting keypoints relied on some variant of scale-invariant feature transform (SIFT) [4] which was sate of the art until CNN's were shown to give better results.

II. Methodology

To train and test our model we use the facial keypoints detection dataset from Kaggle [5] with the size of 7049 labeled images of faces. The model has to predict 15 keypoints, which means 30 values (x and y coordinates for each keypoint). We split the data into train set (55%), validation set(15%) and test set (30%). Below is an example of an image sampled from the dataset we use.

Figure 1. Image sampled from dataset.

We use a CNN architecture with 7M parameters. The model is trained for 50 epochs with the batch size of 256 data samples. For optimization of the model we used the Adam optimizer [6]. To evaluate the quality of our model we use the root mean squared error (RMSE) metric.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

III. Results

Our CNN achieves the RMSE on test set equal to **2.1** (just above 2 pixels of average error). On figure 2 we can see an example of correctly detected facial keypoints.

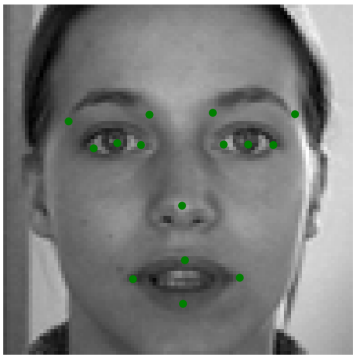

Figure 2. Example of correctly detected facial keypoints.

IV. Conclusion

We presented the problem of facial keypoints detection and build a model for the task based on convolutional neural network. We evaluated the performance of the model using RMSE metric. The model achieved the accuracy of 2.1 pixels. The results could be further improved by using larger dataset and more complex neural network architecture. We could expand the dataset using data augmentation techniques.

References

[1] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[4] D. G. Lowe *et al.*, "Object recognition from local scale-invariant features." in *iccv*, vol. 99, no. 2, 1999, pp. 1150–1157.

[5] "Kaggle competition: Facial keypoints detection," https://www.kaggle.com/c/facial-keypoints-detection, accessed: 2019-11-30.

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.