Key Facial Points Detection with Image Processing Techniques

Elçin İbişoğlu[#1]

Istanbul Technical University,
Istanbul 34467, Turkey
[1]ibisoglu20@itu.edu.tr

***Abstract*— In this study, facial key points are detected by applying some image processing techniques and deep learning methods. Key Facial Points [1] dataset is used to detect key points on faces. A Convolutional Neural Network (CNN) algorithm is created for this project. Train and validation losses are obtained and Root Mean Square Error (RMSE) evaluation metric is utilized. The results show that the smallest RMSE is obtained with the predictions of the points on eyes. The largest RMSE is observed in the case of the mouth.**

***Keywords*— Key Facial Points, CNN, Image Processing, RMS.**

I. Introduction

The technology has been developed in the growing and changing world. Image processing is one of the fields which achieved notable development. Using image processing techniques in accessible technologies allows people to benefit from those kinds of technologies frequently. Thanks to that, the usage area of image processing techniques has become wider. The detection of facial points are crucial to build lots of different applications such as face filters, emotion recognition, pose recognition, and age detection [2]. Also, the most popular applications are used filters such as animal filters or switching faces. In order to apply such filters successfully on faces, the key points of faces should be detected accurately. Various studies have been done to detect facial points. There are some difficulties to catch key points on the face since the structures of people's faces are different from each other. In this study, a Convolutional Neural Network (CNN) algorithm is created to predict facial key points.

The rest of the paper is organized as follows. The utilized methods in the algorithm and approaches in the project are presented in Section 2, while experimental results and datasets information are provided Section 3. Finally the conclusion is drawn in Section 4.

II. Approaches

In this part of the study, the steps to detect key facial points are provided. In order to ensure detailed information and to compare rms results in Section 3, a dataset which is also examined in Section 3 is processed in two different ways. In the first method, all columns of the data set are used. For this method, there are 31 columns in the dataset. In the second method, the facial points are examined separately for each part of faces. These are determined as key points on eyes , key points on eyebrows, key point on nose, and key points on mouth.

A. Data Preprocessing

The first step of the process is examining the dataset and preparing the data for regression. As mentioned in the previous part, all columns in the dataset are used for the first method. It provides obtaining 15 key facial points on a face. In the fig.1, a random image from the dataset and its key points are shown as follows.

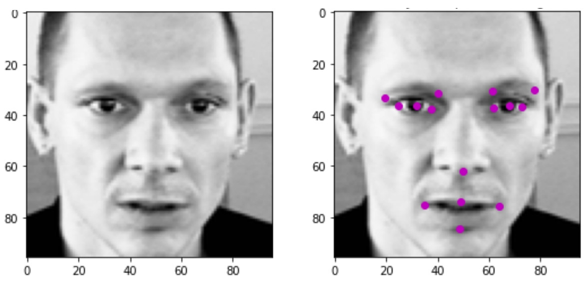

Fig. 1-Image with 15 Facial Points

For the second method, only corresponding columns for each part are used. The unnecessary feature columns are dropped from the dataset. The Table 1. shows the number of feature columns and obtained key points for corresponding parts.

TABLE I
THE NUMBER OF COLUMNS AND KEY POINT FOR EACH PART

Part of Face/Amount	The Number of Feature Columns	The Number of Key Facial Points
Eyes	12	6
Eyebrows	8	4
Nose	2	1
Mouth	8	4

The images with key points for eyes, eyebrows, nose, and mouth are respectively shown in fig. 2 as follows.

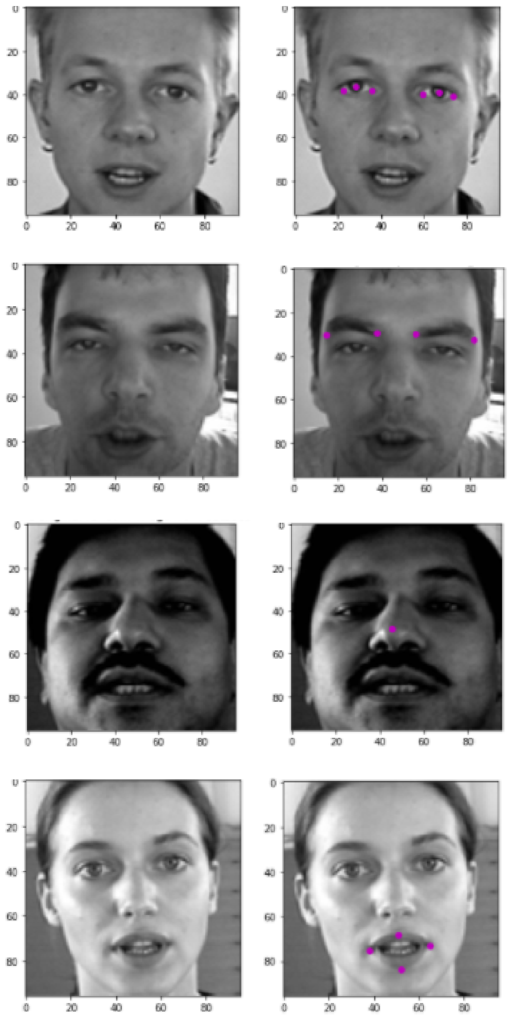

Fig. 2-Keypoints with Differents Parts of Face

As seen from the figure, there are 3 different points given for each eye on a face. Similarly, two different points for each eyebrow for a face. There exists only one point for the nose part while the mouth has 4 different points.

B. Data Augmentation

First of all in this part of the study, images provided with the dataset are flipped horizontally. It results in a change in x coordinate values while y coordinate values are the same. This process doubles the size of the dataset. Then, the brightness of images are increased randomly. After that, the images are flipped vertically. Those two steps are also used to enlarge the size of the dataset. Increasing the number of images in the dataset provides more examples for training and validation. Also, it supports more accurate predictions.

C. Train-Test Splitting

After the data augmentation part, the dataset is splitted into two parts as train and test. The test size ratio is determined as 0.1. The test data is used to predict facial points at the end of the project. Results are given and discussed in Section 3. Then, the remaining part is divided into train and validation parts. The validation size ratio is also decided as 0.1.

D. CNN Algorithm

For prediction of facial key points, a CNN algorithm is created in this project. In the architecture of CNN, 4 sequential layers are prefered. After each convolution, batch normalization and ReLU activation function are utilized. In addition to that, Xavier initialization is used to give a chance to model for initializing better weights. Max pooling is applied to decrease input dimension after each layer. Being prone to overfit is prevented by using dropout. Also, the Adam optimizer is prefered for optimization and loss is used as Mean Square Error (MSE). The results are discussed in Section 3.

III. EXPERIMENTS AND RESULTS

In this project, the key facial points dataset [1] is used to predict facial key points on different faces.

A. Dataset

The dataset [1] that covers 2140 rows and 31 columns is used for this project. The last column of the dataset includes pixel values for each image. This column is determined as target and remaining columns are feature columns. In the dataset, there are two coordinate values (x,y) for each feature. For different features given in the dataset, each coordinate is distributed in a column. For instance, left eye center x is provided in the first column and left eye center y is in the second column. The 15 facial points' x and y values are supplied with the first 30 columns in this dataset where the last column is about image pixels. For the first method, the dataset used as a whole. However, only the relevant features are used in the second method of the project. The irrelevants columns are dropped as mentioned in Section 2. Face parts and corresponding relevant feature names are provided in Table 2.

TABLE II
THE NAME OF COLUMNS FOR FEATURE GROUPS

Part of Face	Name of Columns in Feature Groups	
Eyes	Left eye center x Right eye center x Left eye inner corner x Left eye outer corner x Right eye inner corner x Right eye outer corner x	Left eye center y Right eye center y Left eye inner corner y Left eye outer corner y Right eye inner corner y Right eye outer corner y
Eyebrows	Left eyebrow inner end x Left eyebrow outer end x Right eyebrow inner end x Right eyebrow outer end x	Left eyebrow inner end y Left eyebrow outer end y Right eyebrow inner end y Right eyebrow outer end y
Nose	Nose tip x	Nose tip y
Mouth	Mouth left corner x Mouth right corner x Mouth center top lip x Mouth center bottom lip x	Mouth left corner y Mouth right corner y Mouth center top lip y Mouth center bottom lip y

The names of columns for facial parts are shown in Table 2.

B. Evaluation Method

The evaluation metric is decided as Root Mean Square Error (RMSE) for this project. RMSE is obtained by evaluating the square root of Mean Square Error (MSE). The formula for MSE [3] is shown in Equation 1 where i is from 1 to N.

$$\text{MSE} = \frac{1}{N}\sum(y_i - \widehat{y_i})^2 \qquad (1)$$

In the formula given in Equation 1, the number of samples are shown with N, y_i represents the target point from the test set, and $\widehat{y_i}$ is the model prediction results for these points.

C. Results

For the first method, the train and validation loss curves are obtained and shown in Figure 3.

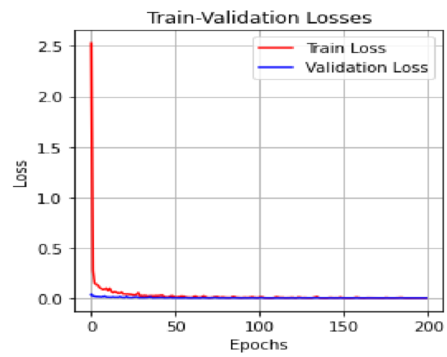

Fig. 3-Train and Validation Losses for 15 Facial points

The train loss is 0.0020 where validation loss equals 0.0021 for training model result of 15 facial key points. The RMSE result of these 15 points is evaluated by using the test set. The RMSE is calculated as 0.8475. A random test result of predicted 15 points is given in Figure 4.

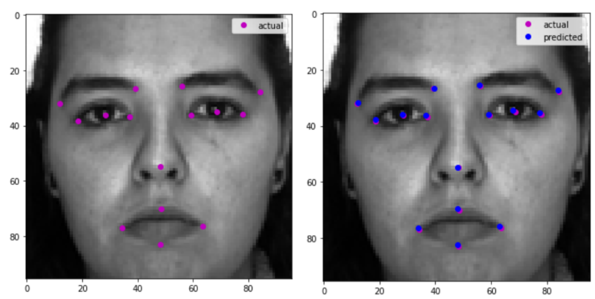

Fig. 4- Prediction Result for 15 Points

For the second method, the loss curves results of the eyes part is shown in Figure 5.

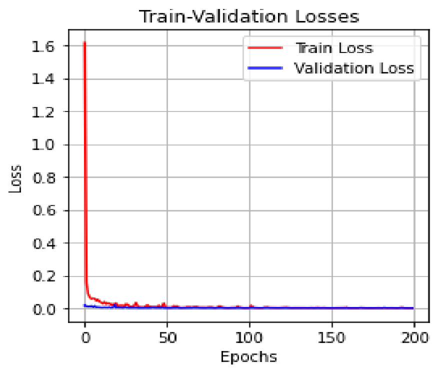

Fig. 5-Train and Validation Losses for Points of Eye

The train and validation losses are obtained as 0.0011 and 0.0013, respectively. RMSE value is 0.5540. A random test result of predicted points on eyes are shown in Figure 6 where magenta represents the actual places of points and blue points are the predicted points.

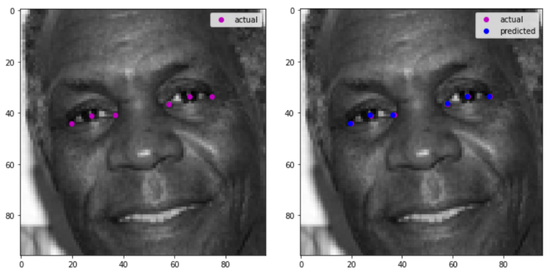

Fig. 6- Prediction Result for Key Points on Eyes

For eyebrows part of the face, the losses are obtained as 0.0022 in training and 0.0017 in validation. RMSE is evaluated as 0.7750.

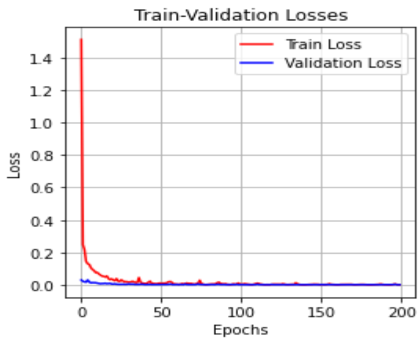

Fig. 7-Train and Validation Losses for Points of Eyebrow

The test result for a random image from the test set is provided in the following Figure 8.

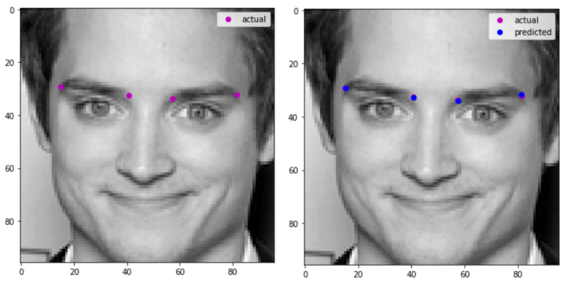

Fig. 8- Prediction Result for Key Points on Eyebrows

For the nose part of the dataset, the train and validation results are 0.0029 and 0.0012 respectively. The results are shown in Figure 9.

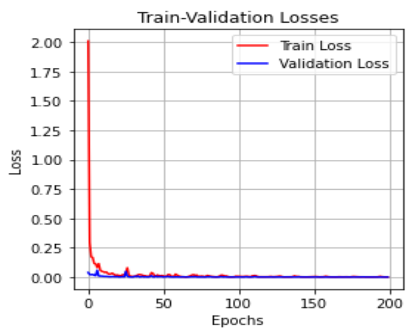

Fig. 9-Train and Validation Losses for the Point of Nose

RMSE is obtained as 0.7214. The test result for a random image from the test set can be seen in the following Figure 10.

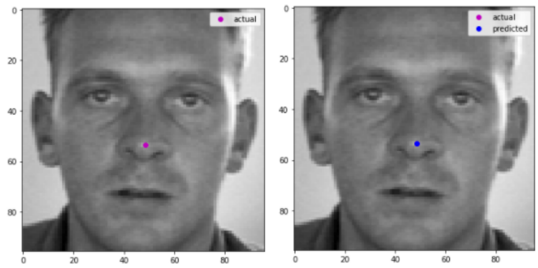

Fig. 10- Prediction Result for Key Point on Nose

In the mouth part of the dataset, 0.0046 and 0.0017 are obtained as train and validation losses respectively. The curves with respect to epoch number are shown in Figure 11.

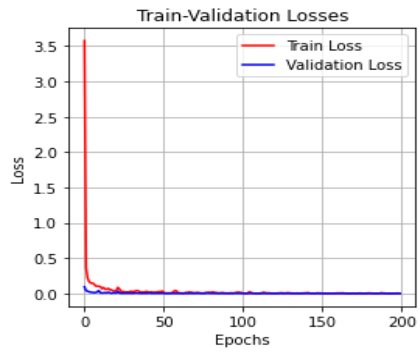

Fig. 11-Train and Validation Losses for Points of Mouth

For this part, RMSE result is calculated as 0.9372. The prediction result for a random image from the test set is provided in Figure 12.

Fig. 12- Prediction Result for Key Points on Mouth

When the results of all cases are examined, it is seen that the smallest RMSE is obtained as 0.5540 in the eyes part of the dataset. In addition to that, the smallest loss results are also observed in this part of the face. The biggest error and losses are in the mouth case. The biggest error is observed as 0.9372. The overall RMSE is calculated as 0.8475 by using the first method. The RMSE result of the nose part is smaller than the result of eyebrows with 0.053.

IV. Conclusions

The detection of facial key points are important to use different applications. The key points are predicted in different ways. A CNN algorithm is created by the purpose of prediction. All key points on a face are predicted in the first method of the project. In the second approach, the face is divided into 4 parts as eyes, eyebrows, nose, and mouth. For that approach, the relevant feature columns are used for each part of the face and results are discussed. The evaluation metric is prefered as RMSE since the project is a regression task. The loss is determined as MSE loss. The results show that the error and losses are the smallest in the eyes part of the face. The RMSE of eyes are calculated as 0.5540. The train and validation losses are 0.0011 and 0.0013, respectively. The largest error and losses are obtained in the mouth part of the face. The error is 0.9372 where 0.0046 and 0.0017 are obtained as train and validation losses respectively.

References

[1] Facial keypoints detection. (n.d.). Retrieved February 04, 2021, from https://www.kaggle.com/c/facial-keypoints-detection/data

[2] Nishad, G. (2019, March 24). Facial keypoint Detection: Detect relevant features of face in a go using CNN & your own dataset... Retrieved February 03, 2021, from https://garimanishad.medium.com/facial-keypoint-detection-detect-relevant-features-of-face-in-a-go-using-cnn-your-own-dataset-e09cf359c2bc

[3] Wu, S. (2020, June 14). What are the best metrics to evaluate your regression model? Retrieved February 04, 2021, from https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b

[4] Brownlee, J. (2019, April 22). A Gentle Introduction to Pooling Layers for Convolutional Neural Networks. Machine Learning Mastery. https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/

[5] Mishra, A. (2020, May 28). Metrics to Evaluate your Machine Learning Algorithm. Retrieved January 30, 2021, from https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234