Design and Evaluation of a SIFT Facial Landmark Detection System for Eyes and Lips Colour Modification

Abstract

This report evaluates a facial recognition model designed to accurately detect and modify keypoints on eyes and lips. Employing advanced preprocessing and feature extraction techniques, including Scale-Invariant Feature Transform (SIFT), the study aims to assess the model's performance across a variety of facial images.

Introduction

Face alignment is a crucial component in the field of computer vision, serving as a foundational technology in numerous applications such as facial recognition, augmented reality animation and video games “Face alignment is a computer vision technique used for identifying the geometric structure of human faces in digital images.” (Gu & Kanade, 1970) The process involves identifying specific landmarks on a face, and using these points to establish a standard facial structure that can be analysed and manipulated algorithmically.

In addition to the primary task of face alignment, this project addresses a secondary, yet significant task—modifying the colour of the lips and eyes in an image. Using the help of the already defined keypoints from the face alignment task, the system alters the colour of those features. This project aims to examine a technique in computer vision and machine learning as well as explore the creative aspects of visual design in digital environments.

Methodology

1.1 Data Preprocessing

To effectively implement a model that can detect keypoints from images it is crucial that the input data is preprocessed. Data preprocessing is a fundamental task in computer vision, and it is used to improve the models performance by enhancing data quality and consistency

The following preprocessing steps were taken:

1. Grayscale Conversion
    a. The images were converted to grayscale to reduce computational complexity since it narrows the data to a single intensity value per pixel rather than three
    b. It does so by averaging the RGB values
2. Noise Augmentation
    a. “Noise augmentation is a form of data augmentation used in machine learning to improve the accuracy of the model.” (AWS, Data Augmentation)
    b. The approach used applies noise augmentation to an image by adding random noise, which is generated from a Gaussian distribution

    c. This technique helps prevent overfitting during the training of the model by introducing variability in the training data
3. Gaussian Blur
    a. Gaussian blurring is used to smooth the image by averaging the pixels under a kernel area, reducing high-frequency noise and details (OpenCV)
    b. Smoothing allows for less spatial information in the image, making it beneficial for preprocessing in image face alignment tasks
4. Histogram Equalization
    a. Histogram equalisation improves the global contrast of the image, making the intensities more distributed (Sudhakar, 2021)
    b. Enhanced contrast can lead to better visibility of facial features, making it easier for the model to perform accurately

These preprocessing techniques improve images by highlighting important features and reducing variations like colour and lighting. This helps the face alignment model perform better and more efficiently.

1.2 Feature Extraction Method

For the feature extraction task the approach taken is using Scale-Invariant Feature Transform (SIFT) descriptors. There are many other techniques that could be used to solve this problem such as using a Histogram of Oriented Gradients(HOG), Convolutional Neural Networks (CNN), etc. However SIFT descriptors are chosen because of their ability to offer a concise description of image features. These descriptors are invariant to image scaling and rotation. These qualities make SIFT descriptors appropriate for analysing faces in a large data set.

To make the task easier the features are extracted using SIFT descriptors using predefined key points. These predefined keypoints are gathered based on the image's shape. This allows for more focus on specific regions of the image, particularly the eyes and the lip, reducing computational cost, thus improving efficiency.

1.3 Model Training

The model that was used for training and test is a linear regression model. Linear regression is a straightforward yet powerful method for modelling the relationship between a dependent variable and one or more independent variables. In this context, the independent variables are the image features extracted via SIFT descriptors, and the dependent variables are the coordinates of facial landmarks.

The training process begins by defining keypoints and extracting corresponding SIFT descriptors for each image. The facial landmark points are flattened to create a vector of coordinates for each image, suitable for regression analysis. To evaluate the model’s performance a training-validation split is implemented. This ensures it generalises well to unseen data. Specifically, 15% of the data is held out as a validation set, the rest is used for training.

Results and Evaluation

1.1 Model Performance

To perform a quantitative analysis of the model, we use the validation set. The model was trained using the training data, but its performance metrics are tested on the validation set.

Table 1: Performance Metrics Result on Preprocessed Images

Mean Squared Error	60.065849741551
Mean Absolute Error	5.676225235229432
Euclidean Distance	[15.45258523 13.48501208 13.29947735 ... 17.49559067 17.6126343 19.23885984]
Average Euclidean Distance	8.927642488350148
Maximum Euclidean Distance	75.5193941601277
Minimum Euclidean Distance	0.0534000764296166

As can be seen from the table, the mean squared error of the model was approximately 60.1, indicating that the average of the squared differences between predicted and actual values is relatively high, indicating that the model is not optimal and could perform better. As for the mean absolute error the model scored 5.68, which indicates that on average the model's predictions are about 5.68 units away from their actual values. This is an average value, suggesting that the model performs well on some cases, and fails on others.

When looking at the maximum and minimum euclidean distance, in the best case scenario the lowest error that the model makes is 0.053 indicating that the model is accurate. However the maximum euclidean distance is 75.5. This is the largest error between a predicted point and an actual point. This indicates that the worst case in the predictions can be very inaccurate.

The histogram below represents the distribution of Euclidean distances between predicted and actual keypoints from the model. The histogram shows a positively skewed distribution. The peak of the histogram is close to the lower end, indicating that most of the Euclidean distances are small, suggesting that for most predictions, the model is relatively accurate. On the other hand, the long tail extending to the right indicates that there are a significant number of predictions where the model performs poorly.

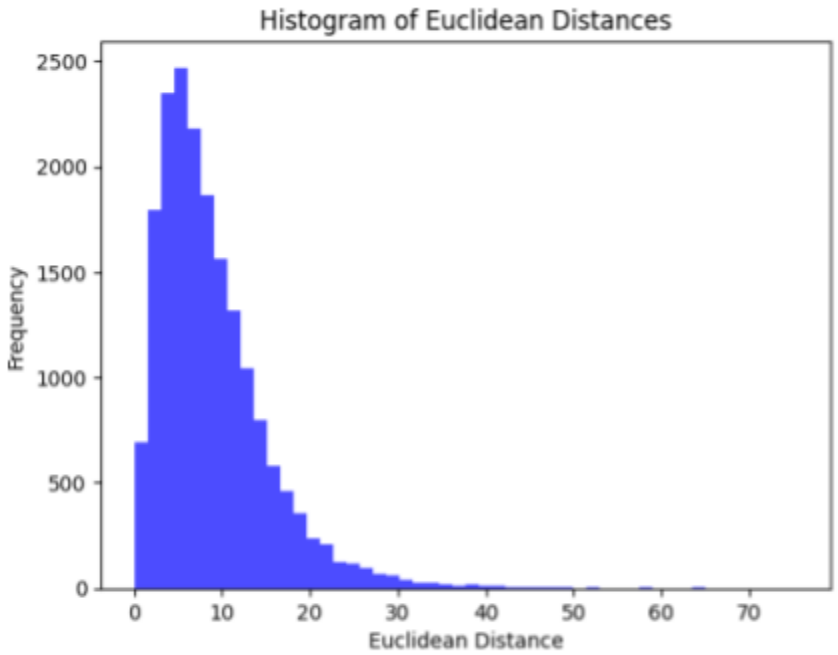

Figure 1: Histogram of Euclidean Distances

1.2 Qualitative Analysis

To test how accurately the model detects eyes/lips keypoints we will also look at example cases.

Figure 2: Detected Keypoints on Example dataset

Figure 2 presents the detected keypoints on six different faces, illustrating the model's response to diverse facial expressions, orientations, and lighting conditions. Each image is annotated with keypoints that the model has identified using the SIFT detector.

For the most part of the images, the model is able to successfully detect the keypoints despite the differences in each picture. However a noticeable issue arises in the second image, where the model struggles with accuracy due to the significant tilt of the head. Keypoints around the eyes and lips are misplaced, highlighting a limitation in handling extreme angles or distortions in facial orientation.

The accuracy in keypoint detection across these varied examples suggests that the model is generally reliable in recognizing facial features. However, the slight imperfections observed, especially in cases where the face is tilted, suggest that there is a need for further improvements to enhance precision and robustness of the model.

1.3 Eyes/Lips Colour Modification Results

Finally, we will test the models ability to change the colour of the lips and eyes.

Figure 3: Eyes/Lips Colour Modification on Example data

Overall, the model demonstrates the capability to detect and modify the colour of eyes and lips in images. However, its performance exhibits variability across different images. In five of the six images presented, the model successfully identifies and changes the eyes and lips' colours. These results suggest that under typical conditions, the model performs well.

A key observation is the model's performance under geometric transformations such as rotations or tilting of the faces. For example, the first image, which is slightly rotated, shows

successful detection and colour application, indicating that the model is capable of accurately identifying the eyes/lips features even though the image is titled. Contrastingly, the fifth image, which also displays a tilt, presents a failure in proper detection and colorization. This indicates a potential limitation in the model's current geometric invariance capabilities. The accurate applications in other images, including those with minor rotations, suggest that the model has a certain degree of robustness. However, the failures highlight the need for an enhanced feature detection that is more resilient to a wider range of facial orientations.

Conclusion

Overall, the analysis revealed that the model is capable of detecting keypoints for the eyes and the lips with moderate accuracy. However, it exhibited less precision in the colour modification of these features, indicating significant potential for enhancement.

One suggestion would be to implement a Canny Edge Detector to the model. Canny Edge Detection could enhance the model's ability to identify clear boundaries around facial features, by looking for areas with strong gradients. By accurately detecting edges, the model can more precisely define areas for colour modification. Incorporating Canny Edge Detection is likely to reduce errors in keypoint localization.

Another suggestion would be to improve the model training process. The use of Linear regression, while straightforward, might be overly simplistic for the complex nature of facial recognition tasks. Transitioning to more advanced machine learning approaches such as Convolutional Neural Networks (CNNs) could significantly enhance the model's ability to learn from a diverse set of facial images. These techniques may enhance both colour modification and keypoint accuracy since they are more appropriate for capturing differences in human faces.

References

Gu, L. and Kanade, T. (1970) Face alignment, SpringerLink. Available at: https://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5_186#:~:text=Definition,human%20faces%20in%20digital%20images. (Accessed: 1 May 2024).

Goals (no date) *OpenCV*. Available at: https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html (Accessed: 1 May 2024).

Mean square error (MSE): Machine learning glossary: Encord (no date) *Encord*. Available at: https://encord.com/glossary/mean-square-error-mse/#:~:text=In%20the%20fields%20of%20regression,target%20values%20within%20a%20dataset (Accessed: 1 May 2024).

Sudhakar, S. (2021) *Histogram equalization*, *Medium*. Available at: https://towardsdatascience.com/histogram-equalization-5d1013626e64#:~:text=Histogram%20Equalization%20is%20a%20computer,intensity%20range%20of%20the%20image (Accessed: 1 May 2024).

(No date) *What is data augmentation? - data augmentation techniques explained - AWS*. Available at: https://aws.amazon.com/what-is/data-augmentation/ (Accessed: 1 May 2024).