Facial Orientation Detection: A Comprehensive Report on MLP Classifier Development and Evaluation

acp22abj

University of Sheffield

Abstract

This report elaborates on developing a classifier for identifying orientations (upright, rotated 90 degrees) in 30, 50, and 90-pixel facial image regions from a dataset of 10,803 images. It covers construction, data pre-processing, model selection, and evaluation methods. Discussions emphasize efficacy, implications, and include a summary of task specifics and dataset details for clarity.

1. Introduction

The focus of this report lies in decoding orientation within square patches extracted from facial images. The task involves precisely identifying orientations—upright, rotated right, left, or completely inverted—across 90, 50, or 30-pixel regions. This pursuit targets the complex problem of facial region orientation detection, critical in understanding facial features. Its significance spans computer vision and biometric domains, underscoring the necessity for accurate orientation recognition. The approach involves crafting a specialized classifier adept at interpreting orientation across varied-sized facial image regions. This exploration encompasses nuanced data pre-processing, intricate model selection, and finely tuned training strategies. Beyond technical aspects, it seeks to underscore the classifier's relevance and usability, emphasizing the need for precise orientation classifiers in image analysis.

2. System Description

The system devised for orienting facial image regions presents a meticulous step-by-step pipeline, meticulously detailed to facilitate replication and in-depth comprehension:

2.1. Data Preparation:

The process commences by extracting facial images and their corresponding person IDs from the 'train.full.joblib' file. Subsequently, sub-images of varying sizes (30x30, 50x50, and 90x90 pixels) are generated from these facial images, employing a patch extractor with a cap of two patches per image.

2.2. Patch Generation and Labeling:

Each derived sub-image undergoes rotations at 0, 90, 180, and 270 degrees, thereby acquiring orientation labels for subsequent training.

2.3. Feature Engineering:

The rotated sub-images are transformed into flattened feature vectors, resulting in the creation of distinct datasets denoted as *all_features_30*, *all_features_50*, and *all_features_90*.

2.4. Model Construction:

Meticulously crafted pipelines, tailored for each sub-image size, encapsulate crucial stages. Beginning with Standard Scaling, pivotal for feature scaling using StandardScaler within every pipeline. The subsequent step employs PCA, employing dimensionality reduction aided by the hyperparameter *n_components*. Finally, the integration of the MLPClassifier, a multi-layer perceptron classifier, involves fine-tuning hidden layer sizes and maximum iterations. Post-construction, these pipelines undergo comprehensive training using the entire dataset to develop specialized models for different sub-image sizes.

2.5. Model Evaluation:

For evaluation purposes, test data is loaded from 'eval1.joblib,' and distinct test sets corresponding to each sub-image size (30, 50, and 90) are extracted from this data.

2.6. Hyperparameter Tuning:

The report meticulously documents crucial hyperparameters like *n_components* in PCA and neural network architecture parameters, such as hidden layer sizes and maximum iterations in MLPClassifier, that were finetuned during experiments.

3. Experiments

3.1. Selection of Best Training Data

The selection of optimal training data was crucial for system resilience. By comparing a dataset of 10,803 images with a 2,000-image subset, we conducted a swift grid search, enabling streamlined iterations. Additionally, we rotated images to create varied orientations for training data. This approach facilitated a seamless transition of the trained model onto the larger dataset, ensuring efficacy and reliability during evaluation. This rotation process involved iterating through complete images, generating rotated sub-images, and assigning corresponding labels for each orientation, a methodology replicated in our evaluation data preparation.

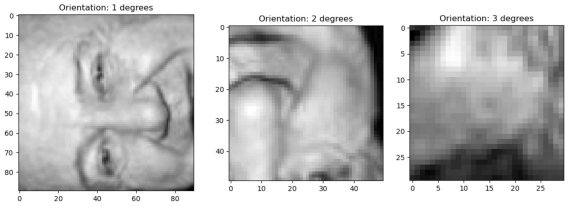

Figure 1: *Sampled face images for 90, 50 and 30 respectively*

Patch Size: 30		
PCA Components	**Hidden Layers**	**Score/Accuracy**
80	(300, 300)	0.6535
80	(330, 330)	0.651
100	(330, 330)	0.6505
Patch Size: 50		
PCA Components	**Hidden Layers**	**Score/Accuracy**
100	(300, 300)	0.8405
100	(340, 340)	0.851
300	(300, 300)	0.834
Patch Size: 90		
PCA Components	**Hidden Layers**	**Score/Accuracy**
70	(340, 340)	0.9865
80	(340, 340)	0.9835
90	(340, 340)	0.983

Table 1: *Top 3 MLP Model Iterations for Different Patch Sizes*

3.2. Hyperparameter Optimization

The optimization process for each model involved tailoring hyperparameters to maximize performance. For the 30-pixel model, 80 PCA components and (300, 300) hidden layers in the MLP were chosen. The 50-pixel model utilized 100 PCA components and (340, 340) hidden layers, while the 90-pixel model employed 70 PCA components and (340, 340) hidden layers. These configurations balanced dimensionality reduction with complex pattern capture, aligning the MLP's hidden layers to encapsulate intricate nonlinear structures akin to kernel PCA for precise orientation detection.[1]

3.3. Managing Model Size Constraints

Ensuring control over the MLP architecture size was pivotal. PCA played a key role in feature reduction, while regulating hidden layer dimensions maintained a delicate balance between complexity and performance. Imposing a maximum iteration limit (max iter=500) ensured convergence, effectively constraining the MLP model within predefined size boundaries. Following exhaustive evaluation, the MLP classifier emerged as the optimal choice, backed by comprehensive assessments of accuracy metrics and efficiency in precise orientation detection across varied scenarios. Meticulous exploration across diverse models solidified the MLP classifier's position for accurate determination of facial region orientation.1

4. Results and Analysis

4.1. Model Evaluation Results

The performance evaluation of the models built for 30, 50, and 90-pixel sub-image sizes revealed distinct accuracy scores and notable trends across different image sizes.

Sub-image Size	Score (%)
30	65.75
50	84.15
90	98.4

4.2. Confusion Matrices

4.2.1. Sub-image Size: 30 pixels

The model achieved an accuracy score of 65.75%. The confusion matrix indicates varying levels of misclassifications, particularly notable in differentiating certain orientations.

$$\begin{bmatrix} 299 & 50 & 105 & 46 \\ 28 & 352 & 36 & 84 \\ 94 & 31 & 339 & 36 \\ 36 & 92 & 47 & 325 \end{bmatrix}$$

Notably, the model struggles to differentiate between orientations, especially mistaking a few orientations for others. For instance, the model often confuses rotations by 90 degrees left with rotations by 90 degrees right.

4.2.2. Sub-image Size: 50 pixels

The 50-pixel sub-image model exhibited a significantly higher accuracy of 84.15%. Analysis of the confusion matrix indicates improved performance compared to the 30-pixel model, depicting fewer instances of misclassification.

$$\begin{bmatrix} 425 & 11 & 49 & 15 \\ 16 & 437 & 9 & 38 \\ 61 & 15 & 411 & 13 \\ 14 & 68 & 8 & 410 \end{bmatrix}$$

Some confusion persists, primarily between rotations by 90 degrees left and right, although to a lesser extent than in the 30-pixel model.

4.2.3. Sub-image Size: 90 pixels

The 90-pixel sub-image model showcased exceptional performance, scoring 98.4%. Its confusion matrix illustrates significantly fewer misclassifications across all orientations, indicating high precision.

$$\begin{bmatrix} 489 & 1 & 5 & 5 \\ 4 & 491 & 1 & 4 \\ 3 & 0 & 494 & 3 \\ 2 & 4 & 0 & 494 \end{bmatrix}$$

The model excels in accurately distinguishing between different orientations, exhibiting minimal confusion across all classes.

4.3. Analysis

The evaluation emphasizes the effect of sub-image size on accuracy. Larger sizes showed better performance than smaller ones, impacting orientation detection precision. This highlights their significance in accurate facial orientation determination.

5. Discussion and Conclusions

Assessing across sub-image sizes revealed notable disparities: the 90-pixel model excelled at 98.4%, while the 30-pixel counterpart struggled at 65.75%. This underscores pixel size's pivotal role. The MLP's performance relied on fine-tuning PCA components for feature representation and adjusting hidden layers for complex pattern extraction in orientation detection. Further enhancements could involve transfer learning or synthetic sample augmentation techniques.

6. References

[1] M. U. G. K. S. Q. W. Gulraiz Khan, Aiman Siddiqi and S. Samyan, "Geometric positions and optical flow basedemotion detection using mlp and reduceddimensions," *IET Image ProcessingVolume 13, Issue 4*, vol. 13, no. 4, pp. 634–643, Mar. 2019.