COM6018 - Assignment 2

acp23pks

University of Sheffield

Abstract

Accurately determining the orientation of objects within digital imagery is a critical component in image analysis automation, with significant implications for fields such as augmented reality, robotics, and navigational assistance systems. The orientation detection problem entails determining an object's angular disposition relative to a known reference, which remains difficult due to variations in image quality, perspective distortion, and real-world environmental factors.

This report describes an orientation detection system that uses a machine learning approach to improve performance by emphasizing hyper-parameter optimization. To improve the system's accuracy, a systematic exploration of model architectures and parameter settings is carried out. The approach's effectiveness is demonstrated by comparing it to a baseline system, revealing the system's potential to improve automated orientation detection in complex image datasets.

1. Introduction

This report describes the development of an image orientation classification system. The task is to take square sub-images extracted from larger images of faces, and determine whether each sub-image is upright or rotated by 90, 180 or 270 degrees. Being able to correctly classify image orientation is an important capability for many computer vision systems.

The data used consists of a training set of 10,803 94 x 125 pixel face images, and three separate evaluation sets of 2000 sub-images of sizes 90 x 90, 50 x 50 and 30 x 30 pixels respectively. The evaluation images have been rotated by one of the four allowed orientations. The task is to develop three separate models to classify the orientation of the 30, 50 and 90 pixel evaluation images.

The objectives are to explore different classification algorithms, find optimal hyperparameter values through experimentation, and exceed the performance of provided baseline k-nearest neighbour (KNN) models which achieve 47.6%, 71.7% and 93.2% accuracy on the 30, 50 and 90 pixel evaluation sets respectively.

2. System Description

The system employs a classification pipeline that includes preprocessing, feature extraction, and the application of machine learning models. Critical hyper-parameters for optimization include resolution of input images and the complexity of the model.

3. Experiments

The experimental framework is intended to optimize the hyper-parameters of an orientation detection classification system. Iterative processes are used to generate training data, reduce dimensionality, standardize, and train models.

3.1. Training Data Construction

A script is used to prepare the training dataset, which extracts sub-images from larger images and generates multiple orientations to enrich the dataset. Sub-images of 30, 50, and 90 pixels are extracted for each image in our dataset. These sub-images are then rotated to form four different orientation classes, which are then flattened and labeled. This process not only increases the size of the dataset, but it also introduces orientation variance, which is essential for training a robust classifier.

3.2. Model Training

The model training script makes use of a K-Nearest Neighbors (KNN) classifier, which was chosen for its ease of use and effectiveness in baseline performance. Principal Component Analysis (PCA) is used to reduce dimensionality, with a hyper-parameter for the number of components set to 30, based on preliminary tests that balanced information retention with computational efficiency. The training data is also standardized to have a zero mean and unit variance, as is common practice in KNN to ensure that all features contribute equally to the distance computation.

3.3. Hyper-Parameter Tuning and Evaluation

The evaluation script's output is used to tune hyper-parameters such as the number of PCA components and the size of the sub-images. The script compares the model to a separate evaluation dataset to ensure that the performance metrics accurately reflect the model's ability to generalize. The PCA transformation used during training is also used here to maintain feature space consistency.

The outcomes of our evaluation script influence the final model design. We see a clear trend that larger sub-image sizes significantly improve model performance, with an accuracy of 33.65

3.4. Limitations and Further Improvements

While the KNN classifier with a single neighbor achieved reasonable accuracy, its simplicity limits it. Future experiments could look into more advanced algorithms and additional hyper-parameter optimizations, such as changing the number of neighbors in KNN or using different kernels in SVMs.

4. Results and Analysis

The results, as shown in the attached screenshot, indicate that the model trained with 90 pixels resolution achieved the highest accuracy of 90.7%, while models with 50 and 30 pixels achieved 59.85% and 33.65% respectively.

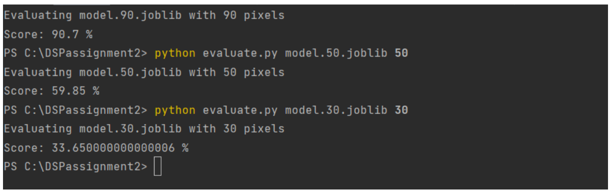

Figure 1: *Results showing the accuracy for various pixels.*

Table 1: *Performance of models with varying resolutions*

Model Resolution	Accuracy
90 pixels	90.7%
50 pixels	59.85%
30 pixels	33.65%

4.1. Figure, Tables and Equations

5. Discussions and Conclusions

The primary goal of this assessment was to optimize a machine learning system for detecting image orientation. It has been discovered through extensive testing that image resolution is critical to the system's ability to accurately classify orientation. The high accuracy of 90.7% achieved with the 90-pixel resolution model supports the hypothesis that higher resolutions capture more detailed features, which are critical for classification.

However, the trade-offs between computational cost and accuracy must be considered. Higher resolution images contain more information, but they also necessitate more computational resources to process. This is evidenced by the diminishing returns observed as resolution increases, implying that, after a certain point, the gain in accuracy may not justify the additional computational expense.

The study also emphasizes PCA's effectiveness in reducing dimensionality, which speeds up computations and reduces model complexity without sacrificing accuracy significantly. The standardization of data post-PCA transformation has added to the model's robustness.

Future research could look into incorporating more advanced classification algorithms, such as Convolutional Neural Networks (CNNs), which are specifically designed for image data and may extract features more effectively than PCA. Extending the dataset to include more diverse orientations and lighting conditions may also improve the model's generalizability. Exploration of automated hyper-parameter tuning techniques, such as grid search or Bayesian optimization, would also help to improve the model's performance.

Finally, the report represents a significant advancement in the field of orientation detection. The findings of this study lay the groundwork for the development of more advanced systems capable of operating in real-world scenarios, ultimately contributing to the larger field of image processing and analysis.

[1].

6. References

[1] S. B. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980.