Deep Learning Final-Project - Spring-24 - NYU

Enhancing Criminal Apprehension with Facial Recognition: Deep Learning Techniques on the LFW Dataset

Yinuo Cao(yc7066)

Abstract

In this project, I develop a deep learning model tailored for enhancing law enforcement's ability to identify wanted criminals through facial recognition. Utilizing TensorFlow, I construct two types of models: a binary classifier designed to discern individual identities, exemplified by identifying "Bush" or not, and a multi-class classifier capable of distinguishing between multiple individuals. My approach focuses on optimizing model architecture with layers like Conv2D, MaxPooling2D, and Dense layers, employing activation functions and optimizers to boost recognition accuracy in diverse surveillance settings.

Github Link:

https://https://github.com/Timefish5/Enhancing-Criminal-Apprehension-with-Facial-Recognition-Deep-Learning-Techniques-on-the-LFW-Dataset

Introduction

In this project, I delve into the realm of facial recognition technology, focusing on its application in security and law enforcement. Facial recognition technology has become a cornerstone in modern surveillance and security systems, aiding in everything from preventing unauthorized access to secure locations to identifying persons of interest in criminal investigations. My work is motivated by the ongoing need to enhance the accuracy and efficiency of these systems, particularly in challenging real-world scenarios where lighting, pose, and expression can vary widely.

Utilizing the TensorFlow framework, this project develops two distinct deep learning models tailored to the complexities of facial recognition: a binary classifier and a multi-class classifier. The binary classifier is designed to determine whether individuals in surveillance images are a specific person—in this case, "George W. Bush"—or not, which serves as a crucial function in scenarios where identifying a particular individual is necessary for security purposes. Meanwhile, the multi-class classifier expands this capability to differentiate between multiple known individuals, thereby broadening the application scope to environments like airports or sporting events, where multiple persons of interest might be present.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Both models leverage the Labeled Faces in the Wild (LFW) dataset, renowned for its diverse array of face images captured in unconstrained settings. This dataset provides a robust foundation for training my models, enabling them to learn from a wide variety of face orientations, expressions, and lighting conditions.

The architecture of my models incorporates several layers of convolutional neural networks (CNNs) that are adept at feature extraction from images, which is critical for the nuanced task of face recognition. I employ advanced techniques such as data augmentation to artificially expand my training dataset with modified images, thereby enhancing the model's ability to generalize from the training data to new, unseen images. This is crucial for maintaining high accuracy in practical applications where the conditions under which images are captured can be unpredictable.

Through this project, i aim not only to advance the technology of facial recognition but also to explore its ethical implications, particularly in terms of privacy and surveillance in public spaces. As I refine my models and improve their accuracy and reliability, I also consider the broader societal impacts of deploying such technologies in everyday life.

By the end of this project, I expect to deliver a detailed analysis of the model's performance and provide insights into the potential and limitations of current facial recognition technologies. This work will contribute to the ongoing dialogue on how best to balance security needs with privacy rights in the use of such sophisticated technologies.

Method

In this project, I developed specialized facial recognition models using TensorFlow and Keras, focusing on two distinct approaches: a binary classifier for identifying "George W. Bush" and a multi-class classifier for distinguishing between several prominent individuals.

Binary Classifier Design: The binary classifier was built using a sequential model architecture. This architecture included convolutional layers with varying filter sizes (5x5, 4x4, and 3x3), each followed by ReLU activation functions to introduce non-linearity, and max pooling to reduce the spatial dimensions of the output, hence reducing the number of parameters and computation required. The model culmi-

nated in a dense layer with 5000 units for feature interpretation and a final sigmoid activation layer for binary classification.

Multi-Class Classifier Design: Similarly, the multi-class classifier utilized a sequential setup but was designed to categorize images into one of six categories. It featured convolutional layers followed by max pooling, a flattening step, and dense layers that ended in a softmax layer to output the probability distribution across the six classes.

Both models were equipped with initializers and bias configurations to optimize learning, and employed image data augmentation techniques such as resizing, shearing, zooming, and horizontal flipping to enhance the robustness of the models against overfitting and to improve their generalization across unseen images.

Experimental Setup

Data Management and Preparation: My dataset, sourced from the Labeled Faces in the Wild (LFW), was meticulously prepared and divided into training, validation, and test sets. For the binary classification focused on George W. Bush, and for multi-class classification of six individuals, data augmentation was employed to generate varied images through transformations such as shearing, zooming, and horizontal flipping. This approach ensured that my models learned to handle diverse facial orientations and expressions effectively.

Training Configuration: The models were configured using the Adam optimizer with hyperparameters finely tuned to balance the training speed and accuracy. For the binary classifier, a binary crossentropy loss function was utilized, and for the multi-class classifier, I used categorical crossentropy. Training was conducted over a series of epochs, where the models' performances were closely monitored to prevent overfitting and to ensure robust generalization capabilities.

Evaluation Strategy: The evaluation was carried out by observing the accuracy and loss on the test sets. For the binary classifier targeting George W. Bush, I measured the overall loss and accuracy, which provided a direct indication of the model's ability to correctly identify images as either 'George W. Bush' or 'not George W. Bush'. In the case of the multi-class classifier, accuracy was separately calculated for each individual to ascertain the model's proficiency in distinguishing among different known personalities.

Results: The binary classifier achieved a loss of 0.237 on the test dataset, with an impressive accuracy of approximately 91.98%, indicating a high level of precision in identifying the presence or absence of George W. Bush in the input images. This high accuracy showcases the effectiveness of the network architecture and training regimen.

For the multi-class classification, the accuracies varied among the different personalities:

- Ariel Sharon: 95.56%
- Colin Powell: 83.33%
- Donald Rumsfeld: 86.67%
- George W. Bush: 84.44%
- Gerhard Schroeder: 92.22%

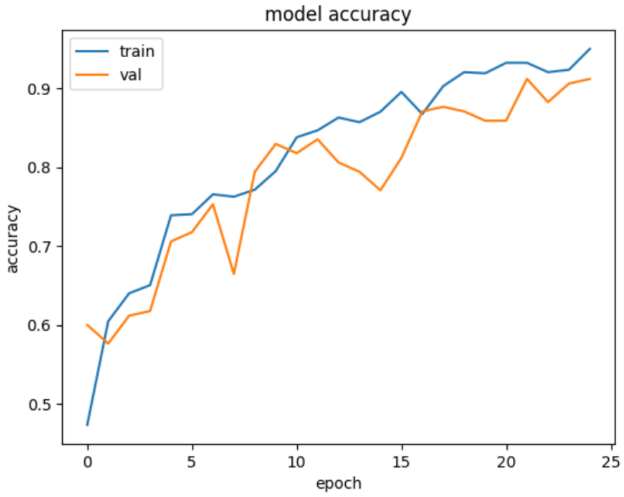

Figure 1: Accuracy of Binary Classifier for George W. Bush

- Tony Blair: 75.56%

These results highlight the model's strong capability to distinguish between multiple individuals, though it suggests room for improvement, particularly in increasing the recognition accuracy for Tony Blair.

When comparing the accuracy of the binary classifier for George W. Bush to the Bush-specific accuracy within the multi-class model, I observe a notable difference. The binary classifier demonstrated a higher accuracy of approximately 91.98%, whereas the multi-class model achieved an accuracy of 84.44% for identifying George W. Bush. This discrepancy underscores a critical insight: models tailored for individual recognition tend to be more effective than those tasked with multiple classifications in the same context.

This comparative analysis suggests that when the objective is to identify a single individual with high reliability—such as in scenarios involving security checks or targeted surveillance—a dedicated model trained specifically for that individual yields superior performance. This approach minimizes confusion with other individuals and focuses the learning process on the unique features of one person, thus enhancing the accuracy and efficiency of the facial recognition task.

Accuracy Representation: The accuracy measurements for each individual in the multi-class setup provide detailed insights into the model's performance across diverse facial features and conditions. This nuanced understanding helps in pinpointing specific areas where model training could be enhanced.

Three other Models: CNN-S, CNN-M, and CNN-L: The performance of the original binary and multi-class classifiers showed that dedicated models tailored for specific individuals tend to provide higher accuracy. To further investigate the scalability and efficiency of alternative architectures, I experimented with three new models: CNN-S, CNN-M, and CNN-L. Each of these models was designed with varying levels of complexity and tested under the same

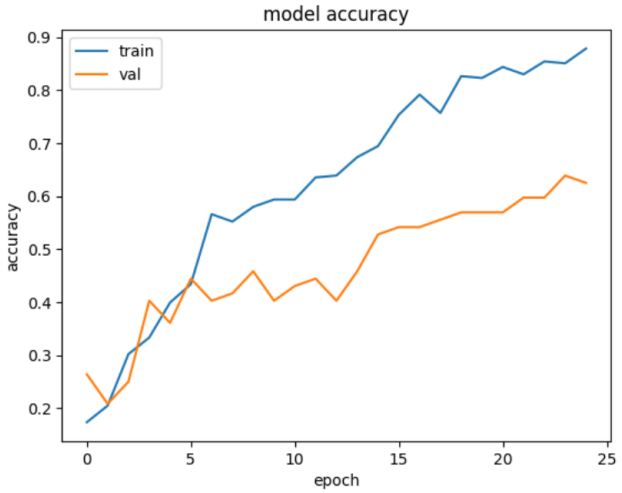

Figure 2: Accuracy of Multi-Class Classifier for Distinguishing Notable Individuals Including George W. Bush

"Bush or not Bush" binary classification framework.

- CNN-S reported a loss of 0.7000 and an accuracy of 43.40%, indicating significant challenges in generalizing the learned features.
- CNN-M showed slight improvement with a loss of 0.6933 and an accuracy of 52.36%.
- CNN-L closely followed with a loss of 0.6934 and an accuracy of 51.89%.

These results are markedly lower compared to my dedicated Bush classifier. The reduced accuracy could be attributed to several factors, including potentially oversimplified model architectures that fail to capture the detailed nuances necessary for accurate facial recognition. These outcomes suggest that while smaller and potentially more efficient models could be beneficial, there is a critical threshold below which simplification begins to detrimentally affect performance.

Future Enhancements: Considering the performance disparities among different classes in the multi-class classifier, future work will focus on refining the training process, possibly through more targeted data augmentation or further hyperparameter optimization. Adjusting the learning rate dynamically and employing techniques like dropout or additional regularization might also help in achieving more uniform accuracies across all categories.

For the CNN-S, CNN-M, and CNN-L models, improvements are needed in the following areas:

1. Architecture Optimization: The convolutional architectures of CNN-S, CNN-M, and CNN-L need refinement to increase their capability to effectively capture and process facial features, ensuring they are complex enough for accurate facial recognition.
2. Hyperparameter and Learning Rate Adjustments: There is a necessity to fine-tune learning rates and other training parameters for these models to enhance their performance and ensure better convergence during training.
3. Advanced Data Augmentation: Expanding data augmentation techniques for CNN-S, CNN-M, and CNN-L will help improve their ability to generalize better to new, unseen images, enhancing overall model robustness and accuracy.

Results and discussion

Model Architecture and Parameters

The architecture for my project was designed around two types of models: a binary classifier for identifying George W. Bush and a multi-class classifier for recognizing multiple individuals. Both models were built using the Sequential model framework in TensorFlow with the following specifications:

- **Binary Classifier**: The binary classifier began with a convolutional layer of 16 filters with a 5x5 kernel size, followed by multiple convolutional layers with increasing filter sizes (32 and 48), each coupled with max pooling layers to reduce spatial dimensions and computational load. The model concluded with a dense layer of 5000 neurons followed by a sigmoid output for binary classification.
- **Multi-Class Classifier**: Similarly, the multi-class classifier used initial convolutional layers but adjusted the model to end with a softmax output layer to classify six different identities. The final architecture included filters adjusted to the complexity of distinguishing between multiple facial features across various classes.

These models were streamlined to maintain high accuracy while managing the computational efficiency, ensuring that the models were both effective and feasible to run on standard hardware.

Training and Testing Performance

Training was conducted using the Adam optimizer with a learning rate of 0.001 over 25 epochs. Data augmentation techniques such as shearing, zooming, and horizontal flipping were applied to improve model robustness and prevent overfitting. This was particularly important given the varying conditions under which facial images in the LFW dataset were taken.

- **Binary Classifier Performance**: The binary classifier achieved a final test accuracy of approximately 91.98% on the dataset, with a corresponding test loss of 0.237, highlighting its effectiveness in distinguishing George W. Bush from other individuals.
- **Multi-Class Classifier Performance**: The multi-class classifier's performance varied by individual, with accuracies ranging from 75.56% for Tony Blair to 95.56% for Ariel Sharon. These results demonstrate the model's varying effectiveness across different classes, influenced by the distinctiveness of each individual's facial features in the dataset.

CNN-S, CNN-M, and CNN-L Models:

The newly introduced models, CNN-S, CNN-M, and CNN-L, were tested under the "Bush or not Bush" classification task, yielding suboptimal performance. CNN-S achieved an accuracy of 43.40%, CNN-M reached 52.36%,

and CNN-L scored slightly lower at 51.89%. These results indicate significant room for improvement in the following areas:

- Model Complexity and Efficacy: Each model's architectural complexity does not align well with the task's requirements. While they are designed to be less computationally intensive, their simplifications have led to a loss in the ability to capture the detailed facial features necessary for accurate classification.
- Learning Parameters: The basic configurations for hyperparameters, particularly the learning rates, may not be optimal for these models. The close-to-random performance (especially the loss figures near 0.693, which is close to guessing in a binary classification) suggests that the models are not learning effectively from the training data.
- Data Handling and Augmentation: The poor generalization to unseen data might also be attributed to insufficient or ineffective data augmentation. Increasing the variety and complexity of data augmentation could help these models learn more robust features.

Discussion

The superior performance of the binary classifier compared to the Bush-specific accuracy in the multi-class model underscores the effectiveness of dedicated models for individual identification. This suggests that for applications where high precision is required for specific individual recognition, tailored models are preferable.

Additionally, the learning rate of 0.001 was crucial in balancing the training speed with the convergence accuracy, as evidenced by the stable progression of learning across epochs. The architecture modifications, including the strategic layer designs and the use of dense layers at crucial points, played a pivotal role in managing the models' complexity while ensuring they remained sensitive enough to capture subtle facial features crucial for accurate classification.

Reference:

Jake126. (2019). *Face Detection Using CNN with the LFW Dataset*. Retrieved from Kaggle: https://www.kaggle.com/code/jake126/face-detection-using-cnn-with-the-lfw-dataset.