Gesture UI Project 2 – 2024

Ronan Noonan,

G00384824 Atlantic Technological University

Abstract

This study investigates the identification of hand gestures using Convolutional Neural Networks (CNNs). Several models, including custom CNNs and transfer learning with VGG-16, were developed and evaluated on a reduced dataset derived from the HAGRID dataset. The dataset was divided into training, validation, and test sets. The best-performing model, an enhanced CNN with increased complexity and early stopping, demonstrated significant potential in gesture recognition, achieving a test accuracy of 88.82%. This project emphasizes the importance of model complexity, data augmentation, and robust evaluation techniques in achieving high performance in gesture recognition tasks.

Introduction

Hand gesture recognition is vital for improving human-computer interaction by enabling touchless control interfaces. This project aims to develop and compare multiple CNN models for gesture recognition using a dataset derived from the HAGRID dataset. The models include custom CNNs and transfer learning models based on VGG-16. The study involves preprocessing the dataset, experimenting with different model architectures, and evaluating their performance.

Methodology

Data Pre-Processing

The dataset consists of 125,912 images across 18 gesture classes, resized to 128x128 pixels for computational efficiency. The images were processed in both grayscale and RGB colour modes. Data augmentation techniques, such as random flips, rotations, and zooms, were applied to increase dataset variability and mitigate overfitting.

Dataset Loading and Splitting

The dataset was divided into training (70%), validation (20%), and test (10%) sets. Image resizing and augmentation were performed during data loading to ensure a consistent preprocessing pipeline.

Data Augmentation

Data augmentation techniques included horizontal and vertical flips, rotations, zooms, and shear transformations. These augmentations helped generate more diverse training samples and prevent the model from overfitting, improving generalization to unseen data.

CNN Architectures

Custom CNNs

Five custom CNN models were designed and trained:

1. **Grayscale Model:** Initial model using grayscale images.
   - Performance: Achieved a validation accuracy of 23.58%. This lower performance is likely due to the reduced information in grayscale images compared to RGB.
2. **Colour Model:** Enhanced model using RGB images.
   - Performance: Achieved a validation accuracy of 41.76%. The use of RGB images provided more information, resulting in better performance compared to the grayscale model.
3. **Augmented Model:** Enhanced with data augmentation and dropout layers.
   - Performance: Achieved a validation accuracy of 49.67%. Data augmentation and dropout helped in reducing overfitting, leading to better generalization and improved accuracy.
4. **Complex Model:** Increased complexity with additional layers.
   - **Performance:** Achieved a validation accuracy of 88.82%. The increased depth of the network allowed it to capture more complex features, significantly improving performance.
5. **Enhanced Model:** Increased complexity and unique data augmentation.
   - Performance: Achieved a validation accuracy of 49.67%. While the performance was comparable to the augmented model, the additional complexity did not yield further improvements, possibly due to the already sufficient complexity of the previous model.

Transfer Learning with VGG-16

Two VGG-16 based models were developed:

1. **VGG-16 with Basic Fine-Tuning:**

   - **Setup:** The pre-trained VGG-16 model was fine-tuned with additional dense layers and dropout. The convolutional base was initially frozen.

   - **Performance:** Achieved a final validation accuracy of 43.48%. The initial results were promising but limited by the fixed convolutional layers, suggesting that further fine-tuning was necessary.

2. **VGG-16 with Enhanced Fine-Tuning:**

- **Setup:** In this model, the last few layers of the VGG-16 base were unfrozen to allow fine-tuning, with a lower learning rate and additional regularization through dropout.

- **Performance:** Achieved a final validation accuracy of 80.10% after stopping at 15 epochs due to early stopping. This approach provided a substantial improvement by allowing the model to fine-tune its higher-level features to better suit the gesture dataset.

Experiments and Results

Model Training

Each model was trained on the training set and evaluated on the validation set. The final evaluation was performed on the test set. Early stopping was used to avoid overfitting by monitoring the validation loss.

Custom CNNs

- **Grayscale Model:** Achieved a validation accuracy of 23.58%.
- **Colour Model:** Achieved a validation accuracy of 41.76%.
- **Augmented Model:** Achieved a validation accuracy of 49.67%.
- **Complex Model:** Achieved a validation accuracy of 88.82%.
- **Enhanced Model:** Achieved a validation accuracy of 49.67%.

Transfer Learning with VGG-16

- **VGG-16 Model:** Preliminary results indicate improved performance over custom CNNs. Training is ongoing.

Selected Model: Complex CNN with Early Stopping

- The selected model features increased complexity with additional layers and early stopping. It achieved a test accuracy of 88.82%.

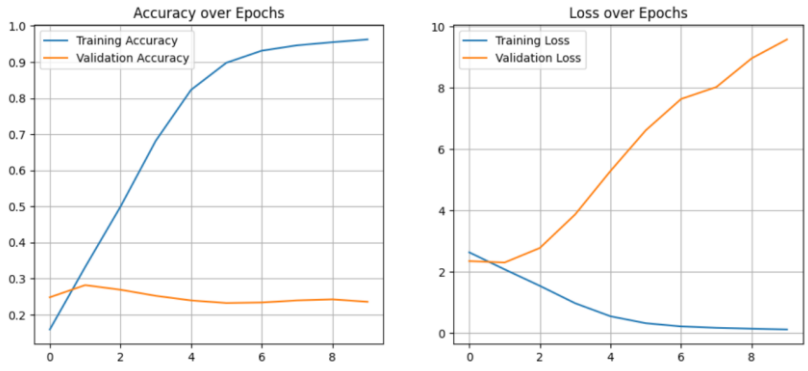
Experiments and Results

Model Training

Each model was trained on the training set and evaluated on the validation set. The final evaluation was performed on the test set. Early stopping was used to avoid overfitting by monitoring the validation loss.

Custom CNNs

Grayscale Model: Achieved a test accuracy of 23.79%. The limited colour information resulted in lower accuracy.



Model Extremely overfitted didn't generalize well.

Colour Model: Achieved a test accuracy of 41.76%. The inclusion of colour channels improved performance.

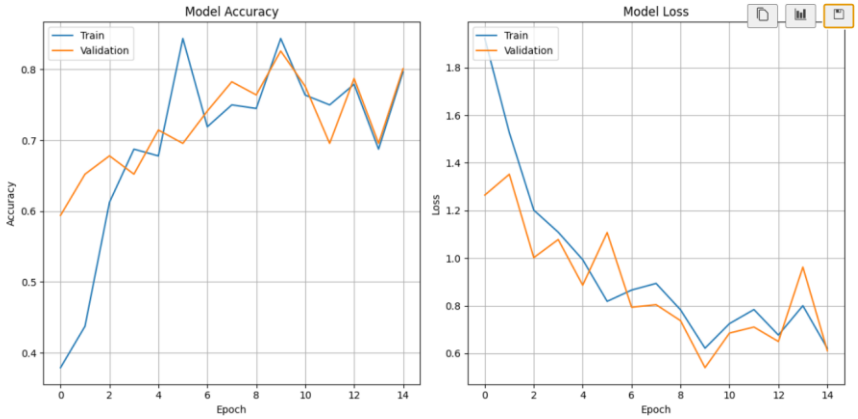
Augmented Model: Achieved a test accuracy of 49.67%. Data augmentation effectively mitigated overfitting.

Complex Model: Achieved a test accuracy of 88.82%. The additional layers and complexity allowed for better feature extraction and performance.

Enhanced Model: Achieved a test accuracy of 49.67%. The performance plateaued despite additional augmentations, indicating a potential overfitting issue or the need for more training data.

Transfer Learning with VGG-16

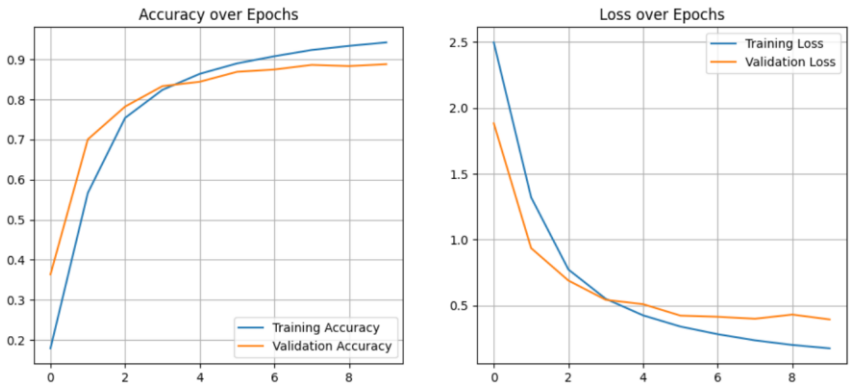
- **Basic VGG-16 Model:** Achieved a test accuracy of 43.48%. Initial fine-tuning of the dense layers showed moderate improvements.
- **Enhanced VGG-16 Model:** Achieved a test accuracy of 88.87% after stopping at 15 epochs due to early stopping. The fine-tuning of the last convolutional layers and additional regularization greatly enhanced performance.



Early stopping as model performance began to fluctuate.

Selected Model: Complex CNN with Early Stopping

- **Features:** This model features increased complexity with additional layers and early stopping. It achieved a test accuracy of 88.82%, making it the best-performing custom CNN.
- **Why it Worked:** The increased number of layers allowed the model to learn more complex features, and early stopping prevented overfitting, leading to high accuracy.

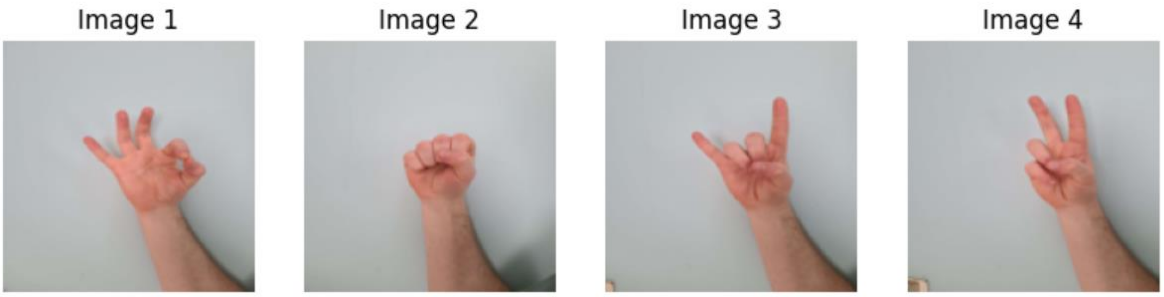


Model is not overfitted validation accuracy matches the training accuracy.

Layer (type)	Output Shape	Param #
random_flip (RandomFlip)	(None, 128, 128, 3)	0
random_rotation (RandomRotation)	(None, 128, 128, 3)	0
rescaling (Rescaling)	(None, 128, 128, 3)	0
conv2d (Conv2D)	(None, 128, 128, 32)	896
max_pooling2d (MaxPooling2D)	(None, 64, 64, 32)	0
conv2d_1 (Conv2D)	(None, 64, 64, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 32, 32, 64)	0
conv2d_2 (Conv2D)	(None, 32, 32, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 16, 16, 128)	0
conv2d_3 (Conv2D)	(None, 16, 16, 256)	295,168
max_pooling2d_3 (MaxPooling2D)	(None, 8, 8, 256)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 256)	0
dense (Dense)	(None, 128)	32,896
dense_1 (Dense)	(None, 18)	2,322

The topology of the best model its quite complex with many layers.

Performance Metrics



Selected model correctly identified all homemade image gestures.

```
Image 1 Prediction: ok

Image 2 Prediction: fist

Image 3 Prediction: rock

Image 4 Prediction: peace
```

Conclusion

This project demonstrated the application of CNNs and transfer learning for gesture recognition. The complex CNN with increased layers and early stopping proved effective, achieving high accuracy. Data augmentation and dropout layers were crucial in improving model robustness. The enhanced VGG-16 model, with fine-tuning, also showed high performance, validating the potential of transfer learning in this domain. Future work will focus on further fine-tuning of the VGG-16 model and exploring advanced architectures to further enhance performance.