

Deep Learning Course Project - Gesture Recognition

In this comprehensive experiment, I aimed to explore various neural network architectures and hyperparameters to determine the most effective model for the task. My goal was to enhance video frame feature extraction and achieve high accuracy in video analysis. All models utilized the same image indexes and image size, ensuring consistency throughout the experiments.

Experiment Setup

Image Indexes and Size

- `image_index` : [0, 2, 4, 8, 10, 12, 14, 16, 18, 20, 22, 25, 29]
- `image_height` and `image_width` : 120x120 pixels

Experiment Details

Here's a summary of each experiment, highlighting the model, accuracy, validation accuracy, and the decision/explanation for each:

Experiment Number	Model	Accuracy	Validation Accuracy	Decision + Explanation
1	Conv3D	0.38	0.42	- Used 3 CNN layers, batch size 100, 15 epochs. - Poor accuracy; consider reducing batch size.
2	Conv3D	0.7143	0.7500	- Maintained Experiment 1 architecture, but reduced batch size to 32. - Accuracy improved but still subpar; explore model complexity.
3	Conv3D	0.60	0.50	- Introduced an additional CNN3D layer to enhance model complexity. - Model overfit data, not recommended.
4	CNN2D + GRU	0.74	0.25	- Utilized 4 CNN2D layers with GRU. - Overfitting issues; investigate model complexity.
5	CNN2D + GRU	0.7563	0.7500	- Removed a CNN2D layer and increased dropout. - Overfitting still present but considered usable.

Experiment Number	Model	Accuracy	Validation Accuracy	Decision + Explanation
6 (Best Model)	MobileNetV2 + GRU	0.79	0.73	- Employed MobileNetV2 + GRU with batch size 50. - Best performing model; consider batch size reduction for further exploration.
7	MobileNetV2 + GRU	0.84	0.62	- Maintained Experiment 6 architecture, but reduced batch size to 32. - Reduced batch size did not improve; model overfitting.

Conclusion

In summary, after extensive experimentation with various models and hyperparameters, we arrived at the following conclusions:

- For the Conv3D Model, Experiment 2 (Model 2) performed the best, achieving an accuracy of 71.43% and a validation accuracy of 75.00% at epoch 14/15.
- In the CNN2D + GRU Model, Experiment 5 also emerged as the best, with an accuracy of 75.63% and a validation accuracy of 75.00% at epoch 9/15.
- However, the top-performing model overall was the MobileNetV2 + GRU Model in Experiment 6, achieving an accuracy of 79.00% and a validation accuracy of 73.00% at epoch 12/15. We recommend further exploration by reducing the batch size to potentially improve performance.

These findings provide valuable insights into the selection of the most suitable model architecture for our video analysis task, with a focus on optimizing model complexity and batch size for improved accuracy.