**3D Convolutional Neural Network Models for Gesture Recognition**

This project aimed to recognize gestures using 3D CNN models and explored multiple model variations before employing transfer learning using MobileNet with an RNN layer.

Below is the comprehensive table based on the training and validation results from the different models and hyperparameters used:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Type** | **Epochs** | **Frames per Video** | **Height x Width** | **Training Accuracy** | **Validation Accuracy** | **Decision + Explanation** |
| Model-1 | Conv3D (Filters: 16, 32, 64, 128 | 25 | 30 | 120x120 | 91.25% | 87.00% | Model achieved rapid improvement in early epochs, followed by a decrease in performance towards later epochs. |
| Model-2 | Conv3D (Filters: 16, 32, 64, 128 | 20 | 20 | 120x120 | 88.8% | 77.00% | Performance was stable; slightly lower accuracy than Model 1. |
| Model-3 | Conv3D (Filters: 16, 32, 64, 128 | 20 | 30 | 160x160 | 89.70% | 83.00% | Larger input size resulted in increased computation; slight improvement in loss and accuracy. |
| Model-4 | Conv3D (Filters: 16, 32, 64, 128 | 20 | 20 | 160x160 | 78.88% | 70.00% | Lower validation accuracy, but gradual improvement observed with more epochs. |
| Model-5 | Conv3D (Filters: 16, 32, 64, 128 | 15 | 30 | 120x120 | 69.83% | 56.00% | Reduced epochs, and performance was less optimal compared to other models. |
| Model-6 | Conv3D (Filters: 8, 16, 32, 64) | 25 | 30 | 120x120 | 91.25% | 71.00% | Lower filter sizes with larger input frames achieved balanced performance. |
| Model-7 | Transfer Learning (MobileNet + GRU) | 25 | 20 | 120x120 | 99.70% | 82.00% | Transfer learning approach with GRU outperformed initial models in validation loss and accuracy. |
| Model-8 | Transfer Learning (MobileNet + LSTM) | 25 | 20 | 120x120 | 98.60% | 88.00% | LSTM variant of transfer learning performed better than GRU, achieving high accuracy and low validation loss |
| Model-9 | Fine-Tuned (MobileNet + GRU) | 30 | 20 | 120x120 | 99.10% | 94.00% | Achieved the highest accuracy and lowest validation loss, demonstrating the effectiveness of transfer learning with tuned dropout and GRU cells. |

**Summary of Observations**

* **Initial Models**: Models with basic 3D Convolution layers and varying input frame sizes showed stable performance, but struggled with overfitting in some cases.
* **Transfer Learning**: Integrating MobileNet with RNN layers significantly improved both validation accuracy and loss. GRU and LSTM variants showed differences in convergence speed and performance.
* **Final Model**: Fine-tuned MobileNet with GRU cells and adjusted dropout rates achieved the best results, indicating a successful balance between model complexity and generalization.