

Documentation for Action Recognition Model

1. Model Architecture

The model is based on Long-term Recurrent Convolutional Networks (LRCN) for action recognition. This architecture is designed to combine the advantages of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs):

1. Convolutional Layers:

- Extract spatial features from the input video frames.
- Layers like `Conv2D` and `MaxPooling2D` are used to reduce spatial dimensions while preserving essential information.
- Activation function: `ReLU` for non-linearity.
- Example Configuration:
 - Filters: `[16, 32, 64]`
 - Kernel Size: `(3x3)`
 - Padding: `same` for retaining dimensions.

2. Recurrent Layers:

- Model temporal dependencies in the video sequence.
- Layers like `LSTM` are used to process the sequence of features extracted by the CNN.
- Example Configuration:
 - Units: `128`
 - Activation: `tanh` with `sigmoid` gates.

3. Fully Connected Layers:

- Combine the temporal features and provide final predictions.
- Dropout is used for regularization.

4. Output Layer:

- Softmax activation for multi-class classification.
- Number of units corresponds to the number of action classes.

2. Optimization Techniques

1. Loss Function:

- Categorical Crossentropy is used as it measures the performance of a classification model with probabilities.

2. Optimizer:

○ Adam Optimizer:

- Combines the benefits of RMSProp and Momentum.
- Adaptive learning rate for faster convergence.
- Default Parameters:
 - Learning Rate: `0.001`
 - Beta1: `0.9`
 - Beta2: `0.999`

3. **Regularization:**

- Dropout Layers: Applied with a dropout rate of 0.5 in fully connected layers to prevent overfitting.
- L2 Regularization: Applied to the weights of certain layers.

4. **Learning Rate Scheduler:**

- Reduces the learning rate dynamically based on validation loss improvement.
- Technique: ReduceLROnPlateau with a patience of 3.

5. **Batch Normalization:**

- Normalizes the activations to stabilize training and improve performance.
-

3. **Evaluation Metrics**

1. **Accuracy:** Measures how often the predictions are correct.
 2. **Precision and Recall:** For evaluating performance on individual action classes.
 3. **F1 Score:** A harmonic mean of precision and recall for imbalanced datasets.
-

4. **Key Features**

- End-to-end learning for video-based action recognition.
- Handles both spatial and temporal features effectively.
- Modular design for easy tuning and customization.