**Gesture Recognition Experiment writeup**

**Theory: -**

**Model Architecture**

There are 3 different deep learning architectures which can be used to build model for gesture recognition. Let’s see those architectures below.

**CNN + RNN**

This is the standard architecture for processing videos. In this architecture, video frames are passed through a CNN layer which extracts features from the images and then these feature vectors are fed to an RNN network to simulate sequence behavior of the video. Output of RNN is regular SoftMax function.

1. We can use transfer learning in 2D CNN layer instead of training own network.

2. LSTM or GRU can be used in RNN

**3D Convolution Network or Conv3D**

3D convolutions are a natural extension to the 2D convolutions. Just like in 2D conv, we have to move the filter in two directions (x and y), in 3D conv, the filter is moved in three directions (x, y and z). In this case, the input to a 3D conv is a video (which is a sequence of 30 RGB images). If we assume that the shape of each image is 100x100x3, for example, the video becomes a 4-D tensor of shape 100x100x3x30 which can be written as (100x100x30)x3 where 3 is the number of channels. Hence, deriving the analogy from 2-D convolutions where a 2-D kernel/filter (a square filter) is represented as (fxf)xc where f is filter size and c is the number of channels, a 3-D kernel/filter (a 'cubic' filter) is represented as (fxfxf)xc (here c = 3 since the input images have three channels). This cubic filter will now '3D-convolve' on each of the three channels of the (100x100x30) tensor.

**Transfer Learning**

We can use some pre-trained models and try to use their knowledge to classify our cases correctly. These pre-trained models which are trained on millions of images may prove vital in solving our problem efficiently. For transfer learning.

• We can do our experiments with the following pre-trained model using MobileNet

• We need to experiment several such models in the 2D CNN layer.

• We can experiment with GRU units as well as LSTM units.

**Case study detail: -**

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| 1 | Conv3D | Throws Generator warning  **Accuracy: 0.794 Validation accuracy: 0.710** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **Raw image input, No cropping, No normalisation, No agumentation, No flipped images, No edge detection**  **Decision: -** Due to Generator warning we are ignoring this model. |
| **2** | **Conv3D** | **Accuracy: 0.656 Validation accuracy: 0.610** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **agumentation, No flipped images, No cropping, No normalisation, No edge detection**  **Decision: -** Training and Validation accuracy looks closely, this can be considered if no any other model with better accuracy available. |
| **3** | **Conv3D** | **Accuracy: 0.796 Validation accuracy: 0.750** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **agumentation, flipped images, No normalisation, No cropping, No edge detection**  **Decision: -** Training and Validation accuracy looks closely, this can be considered if no any other model with better accuracy available. **Best model so far.** |
| **4** | **Conv3D** | **Accuracy: 0.151 Validation accuracy: 0.220** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **agumentation, flipped images, normalisation, No cropping, No edge detection**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |
| **5** | **Conv3D** | **Accuracy: 0.168 Validation accuracy: 0.190** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **agumentation, flipped images, normalisation, cropping, No edge detection**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |
| **6** | **Conv3D** | **Accuracy: 0.206 Validation accuracy: 0.180** | batch size = 20 epochs = 20  Total params: 16,612,069  Trainable params: 16,612,069  Non-trainable params: 0  **Variables: -**  **agumentation, flipped images, normalisation, cropping, edge detection**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |
| **7** | **Conv3D** | **Accuracy: 0.348 Validation accuracy: 0.420** | batch size = 20 epochs = 20  Total params: 863,877  Trainable params: 863,877  Non-trainable params: 1,024  **Variables: -**  **All except normalization**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |
| **8** | **Conv3D** | **Accuracy: 0.292 Validation accuracy: 0.310** | batch size = 20 epochs = 20  Total params: 929,461  Trainable params: 928,437  Non-trainable params: 0  **Variables: -**  **No agumentation,No flipped images,No normalisation,No cropping,No edge detection**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |
| **9** | **Conv3D** | **Accuracy: 0.372 Validation accuracy: 0.240** | batch size = 20 epochs = 20  Total params: 864,101  Trainable params: 863,989  Non-trainable params: 112  **Variables: -**  **No agumentation,No flipped images,No normalisation,No cropping,No edge detection**  **Decision: -** Training and Validation accuracy looks closely, but their accuracy looks very low, we got model 3 as better model. |

**Observation:**

**Model 3** is having training and validation accuracy as **79.6%** and **75.0%**which shows model is Good model and able to learn the behaviour.

**Conclusion:**After all above model experiment, we are going to choose **Model 3** - Resize to 120\*120,  agumentation, flipped images, No normalisation, No cropping, No edge detection as good model which performed well.*Epoch 00020: val\_loss improved from 0.75279 to 0.66837, saving model to model\_init\_2021-11-1515\_44\_17.698701/model-00020-0.66264-0.74837-0.66837-0.75000.h5*Number of Epoch = 20  
Batch Size = 20  
Number of parameters = 16,612,069

Best Model file: **model-00020-0.66264-0.74837-0.66837-0.75000.h5**

**Prepared by :-  
Mohammad Shahid Rashid  
Dharmendra Budha**