Sai Harsha Nagabothu

Dinakar Burgala Nithin

Dr. Feng Luo

Deep Learning- CPSC 8100

**Deep learning model to detect images with cyberbully actions**

**ABSTRACT**

The model developed focuses on identifying a person’s action or behavior from a single image. As image-based human action recognition has become an active research topic in computer vision. Unlike the regular action recognition where video clips or image sequences are used, a still image contains no temporal information for action characterization. Thus the prevailing spatiotemporal features for video-based action analysis are not appropriate for still image-based action recognition. It is more challenging to perform still image-based action recognition than the video-based one. Finally , We present our views and model approach to the mentioned challenge.

**ADVANTAGES**

ResNet helps in solving when deeper networks starts converging, a degradation problem has been exposed: with the network depth increasing, accuracy gets saturated and then degrades rapidly by Instead of learning a direct mapping of x ->y with a function H(x) (A few stacked non-linear layers). Let us define the residual function using F(x) = H(x) — x, which can be reframed into H(x) = F(x)+x, where F(x) and x represents the stacked non-linear layers and the identity function(input=output) respectively. The implemented attention model is pre-trained Resnet152, it accelerate the speed of training of the deep networks. Instead of widen the network, increasing depth of the network results in less extra parameters. Obtaining higher accuracy in network performance especially in Image Classification. This model also allows to reduce the effect of vanishing gradient problem.

**ARCHITECURE**

ResNet-152 introduced the concept of residual learning in which the subtraction of feature is learned from the input of that layer by using shortcut connections (directly connecting input of (n)th layer to some (n+x)th layer, which is shown as curved arrow). It has proven that the residual learning can improve the performance of model training, especially when the model has deep network with more than 20 layers, and also revolve the problem of degrading accuracy in deep networks.

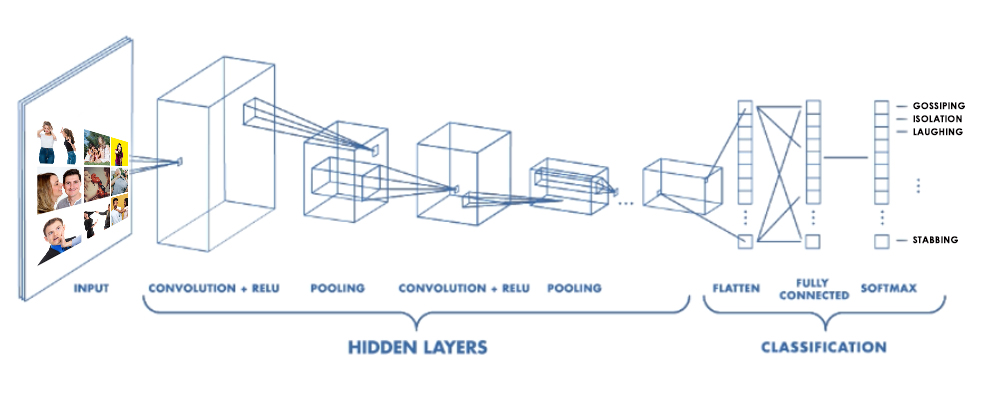
****

Figure 1. CNN – Architecture

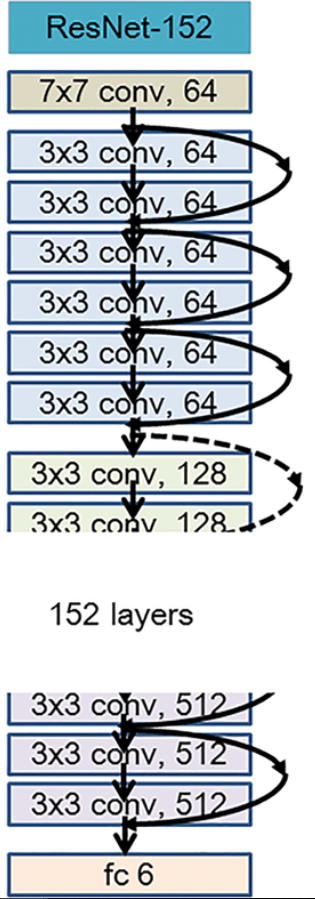


Figure 2. ResNet-152 Architecture

We construct 152-layer ResNet by using more 3-layer blocks. Remarkably, although the depth is significantly increased, the 152-layer ResNet (11.3 billion FLOPs) still has lower complexity than VGG-16/19 nets.

|  |  |  |
| --- | --- | --- |
| Layer name | Output size | 152-layer |
| Conv1 | 112×112 | 7×7 |
| Conv2.x | 56×56 | 1×1, 64  3×3, 64 × 3  1×1, 256 |
| Conv3.x | 28×28 | 1×1, 128  3×3, 128 × 8  1×1, 512 |
| Conv4.x | 14×14 | 1×1, 256  3×3, 256 × 36  1×1, 1024 |
| Conv5.x | 7×7 | 1×1, 512  3×3, 512 × 3  1×1, 2048 |

Table 1. 3-layer Block

**REFERENCES**

[1] <https://arxiv.org/pdf/1711.01467.pdf>

[2] [https://arxiv.org/pdf/1603.05027.pdf](https://arxiv.org/pdf/1711.01467.pdf)

[3] <https://arxiv.org/pdf/1512.03385.pdf>

[4] <https://github.com/facebook/fb.resnet.torch/tree/master/pretrained>

[5] <https://pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html>

[6] <https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8>