**Image Detection – Simpson's characters**

**Assignment 5 - Analytics and Algorithms (Mid Sprint3)**

As part of our core project development scope, the team decided to attempt understanding and developing algorithms for 6 Neural Net Architectures. This would allow us to compare and contrast the accuracy, performance and other metrics associated with these various architectures. These are the 6:

1. Convolutional Neural Network (CNN)
2. Hybrid CNN-ELM
3. Mask R-CNN
4. Faster R-CNN with different Hyper parameter
5. YOLO (You only look once)
6. Single shot Multi-box detection (SSD).

Before diving into these architectures, there are some basic concepts about Deep Learning that we had to get familiar with.

Activation functions are responsible for transforming the weighted inputs to a neuron and determining if the resulting output is ‘activated’ or not. The simplest type of an activation function is Step function (Perceptron) which can input and output binary values only. A step above is a Sigmoid, which can output any value between 0 and 1. Tanh based activation functions are similar to Sigmoids, but faster and can return values between -1 and 1(scaled sigmoids). Rectified Linear Unit (ReLU) returns 0 if the if the output is negative else returns the output.

Cost functions are used to quantify difference between output received from a neuron versus output expected to be received. Quadratic cost function is similar to the Mean Squared Error metric used in Linear Regression models. The other type of cost function is Cross-Entropy. This log-based function enables faster learning when difference between received and expected values are high.

To learn from a cost function and make changes to rectify the error, a gradient descent approach can be used to reduce or eliminate the magnitude of error. Learning rates that determine the size of step to correct the error, Batch sizes that sample input data to feed one run of a network, second order calculations that can use acceleration or momentum of previous steps to adjust size of next step are all Gradient descent related features that can be used to tune a model.

To reduce overfitting a model, techniques like L1/L2 based normalization, Dropout (where a subset of neurons are dropped), artificially expanding data can be used.

To initialize weights, bias and other outputs, Glorot-normal and Glorot-uniform values can be used.

And various types of layers can be used to build a neural network, like for e.g. Dense layer, where all neurons in one layer are fully connected to neurons in the next layer. Softmax layer that outputs a class probability score based on generated weights, Max Pooling that can be used to reduce the size of an image by grouping pixels, Flatten that can modify a 2D or 3D array to a single dimensional structure and Convolutional layer that can learn shapes irrespective of location.

## Description of the various algorithms:

The Hybrid R-CNN is used to do the image classification and the Mask R-CNN is used to segment those images. The Yolo and the SSD are both do the statistics visualization.

CNN

CNN is a multilayer perceptron specially designed for recognizing image contents. This network structure is highly invariant to translation scale slant or covariant deformation. It is the basic algorithm for other algorithms that we needed use in our research and in our project. We used it to do the image classification. CNN though the input image and classifies all dataset into a specifies category the computer treats the input image as an array of pixels. According to the image resolution, we can see h \* w \* d (h = height, w = width, d = size) (Prabhu, 2018).

For the deep learning CNN model, each image will go through a series of convolution layers with kernel, pooling and full connected layer (FC). We can add many convolutional layers and flatten the output and sent to a FC Layer. At last though an activation function to output the class and all classifies images.

R-CNN

R-CNN combines region proposals with CNNS, it can use high-capacity convolutional neural networks to bottom up region in order to locate and image segmentation (Girshick, Donahue, Darrell, & Malik, 2014). R-CNN can also be understood as regions with CNN features.

# References

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation.* Berkeley: UC Berkeley.

Prabhu. (2018, 3 4). *A Medium Corporation[US]*. Retrieved from Understanding of Convolutional Neural Network (CNN) — Deep Learning: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148