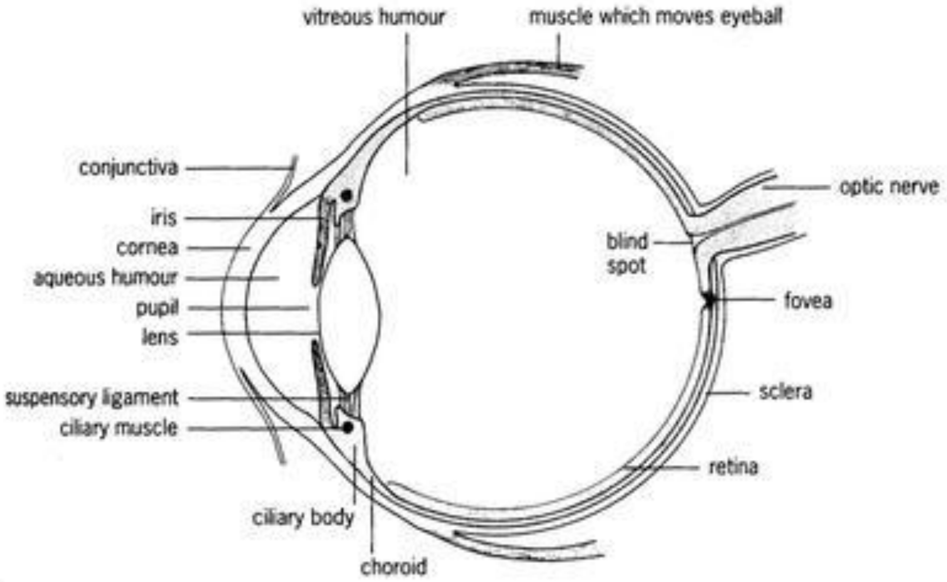
Extending Bio-inspired Human Visual System algorithms

By: Rajshekhar, Nameera
Submitted to : Prof. Dajani
University of Ottawa

Structure of the Eye

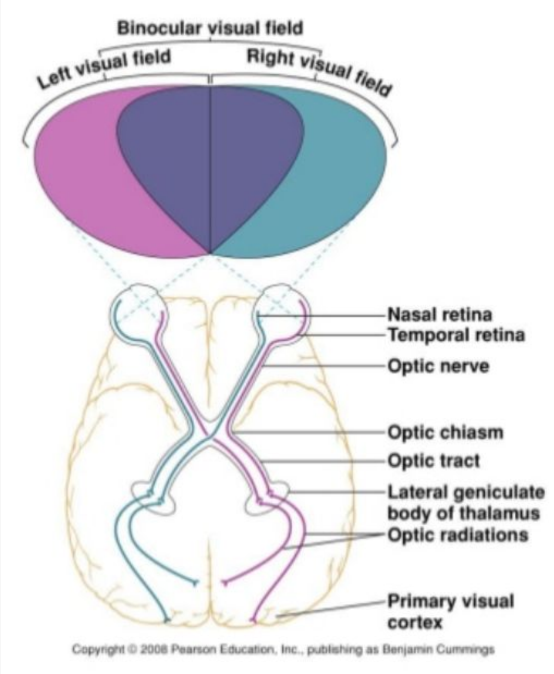


Visual Pathway

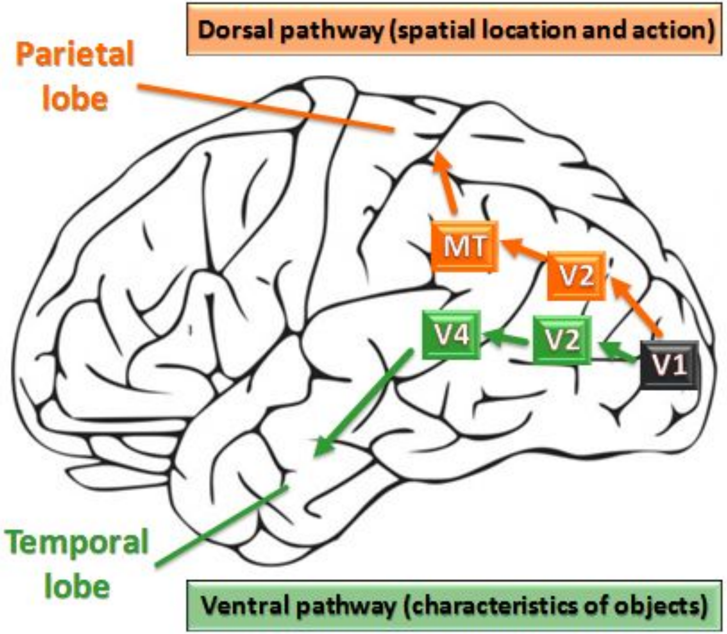
Divided into Left visual field and Right visual field

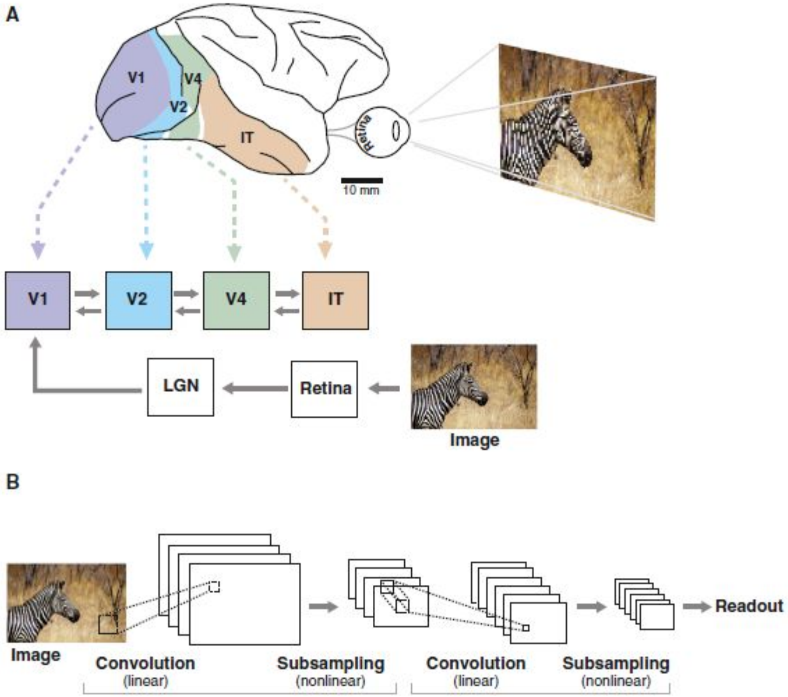
Right visual field- seen by left temporal retina and nasal right retina

Left visual field- seen by nasal left retina and temporal right retina



Visual Cortex





Current Biology

GPU vs Brain

●Computational elements in manmade silicon computers run at fast clock speed- up to billion cycles per second. Visual signals would take tens of ms to traverse from one side of the brain to another.

●GPU might have thousands of cores working on data in parallel. But the brain has billions of neurons operating simultaneously.

Limitations of Deep Learning

●Humans can accurately learn complex visual object categories from fleeting numbers of examples. In contrast, current deep learning approaches require vast quantities of data to work.

● Most systems trained on one data set perform better on that data set than on another that contains the same categories of objects, suggesting some degree of bias in these benchmark dataset.

●One can add carefully crafted ‘noise’ to images and cause them to be arbitrarily misclassified by a current deep learning system. While the original image and the altered image are classified as completely different objects by the deep learning system, they are effectively indistinguishable by humans.

Processing of Time Varying Images

- Another area for growth in computer vision
- Responses of inferotemporal neurons to particular stimuli can nearly be suppressed if the animal learned to expect their appearance in a particular ordered sequence.
- On machine learning side, there have been beginnings in interest for temporal learning.

Our Environment

Our Reality Perception

Our approach

1

Object Recognition and classification task on images: performed by HVS compared to Biologically inspired Convolutional Neural Networks with various layers [See and improve accuracy with different training data and slight change in the deep learning network architecture]

2

Improve and explore various models to construe efficient Biologically inspired computation algorithms

3

Spatial and temporal information of subtle changes in environment: Compare HVS output to Pyramid decomposition approaches to amplify subtle motions and changes in real life scenarios [See accuracy with different signal processing approaches]

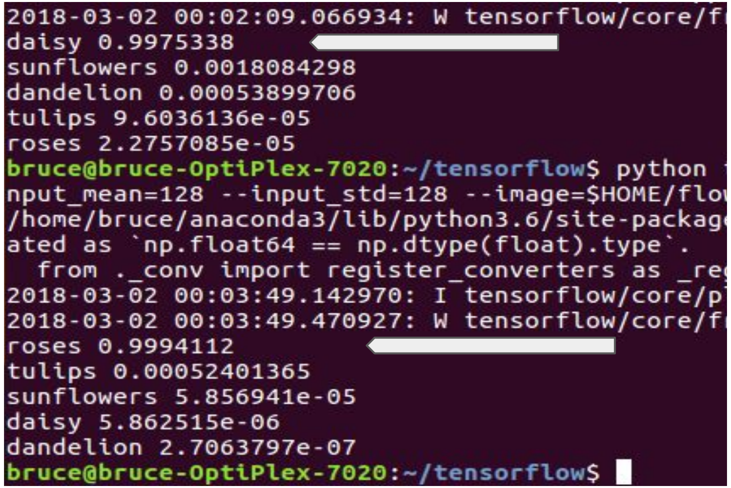
An Object Recognition Exercise [for humans first, then machines]

The Flower Classification Problem

Daisy

Roses

How a machine responds ?



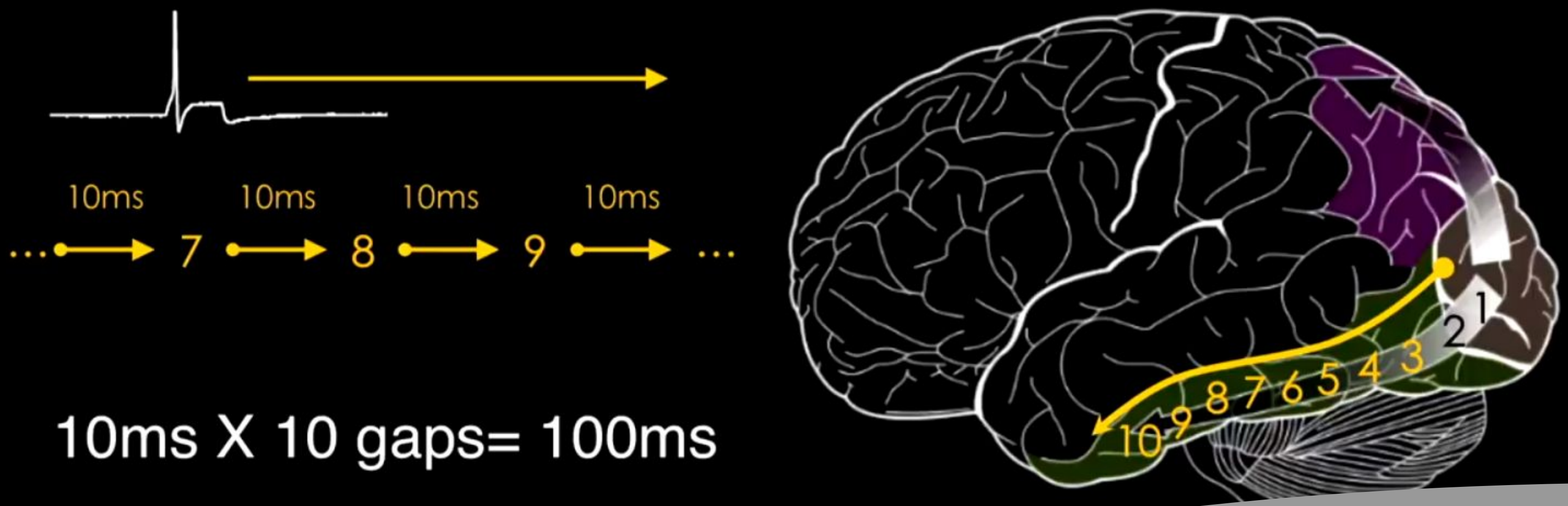


Figure reference : TED.com

How a human responds : Information flow
Model of HVS

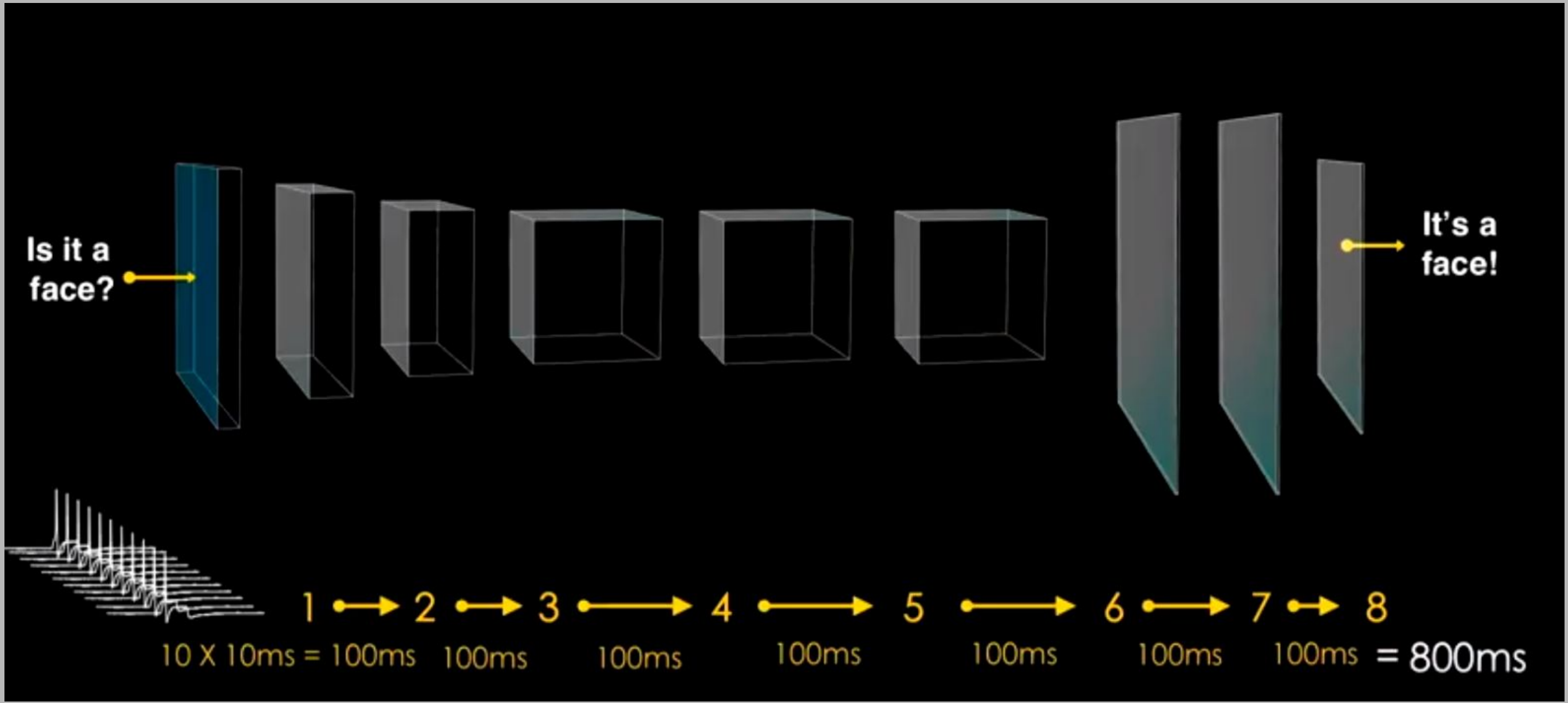


Figure reference : TED.com

How a Neural Network model responds

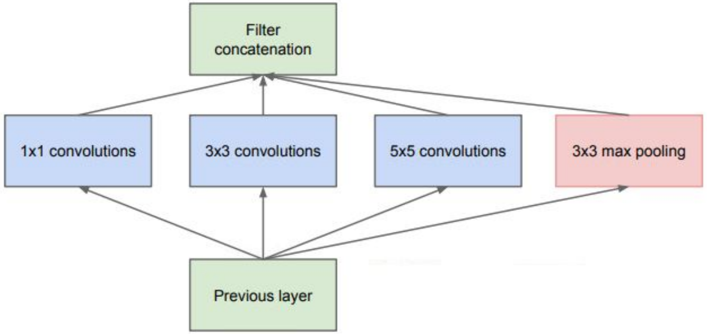
Training and deploying a Neural Network

The Inception Module

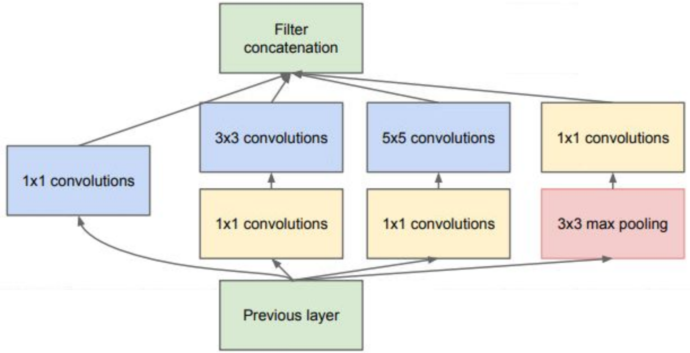
A Simpler Dimensional Machine based Approach for the Flower Problem

The Inception Model : Simplified version

- Stacked convolutional layers followed by max-pooling, fully connected layer
- Concept of drop-out introduced to avoid overfitting
- Inspired by neuroscience, Gabor filters of different sizes to handle multiple scales of features
- All layers of Inception learned, contrary to fixed 2-layer deep model
- 1x1 convolution filters help in dimensionality reduction, thus reduce complexity
- Allows more width and depth in network with less computational penalty

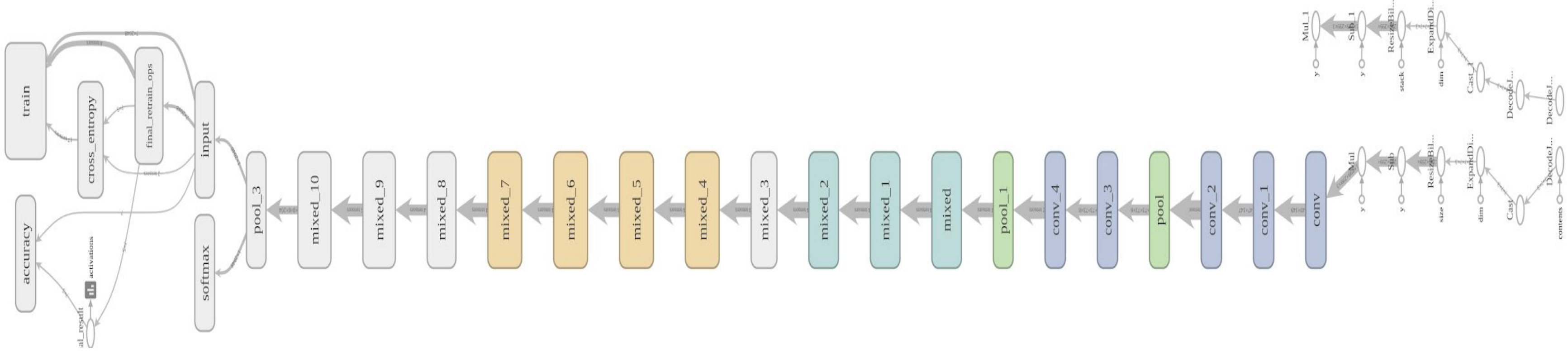


(a) Inception module, naïve version

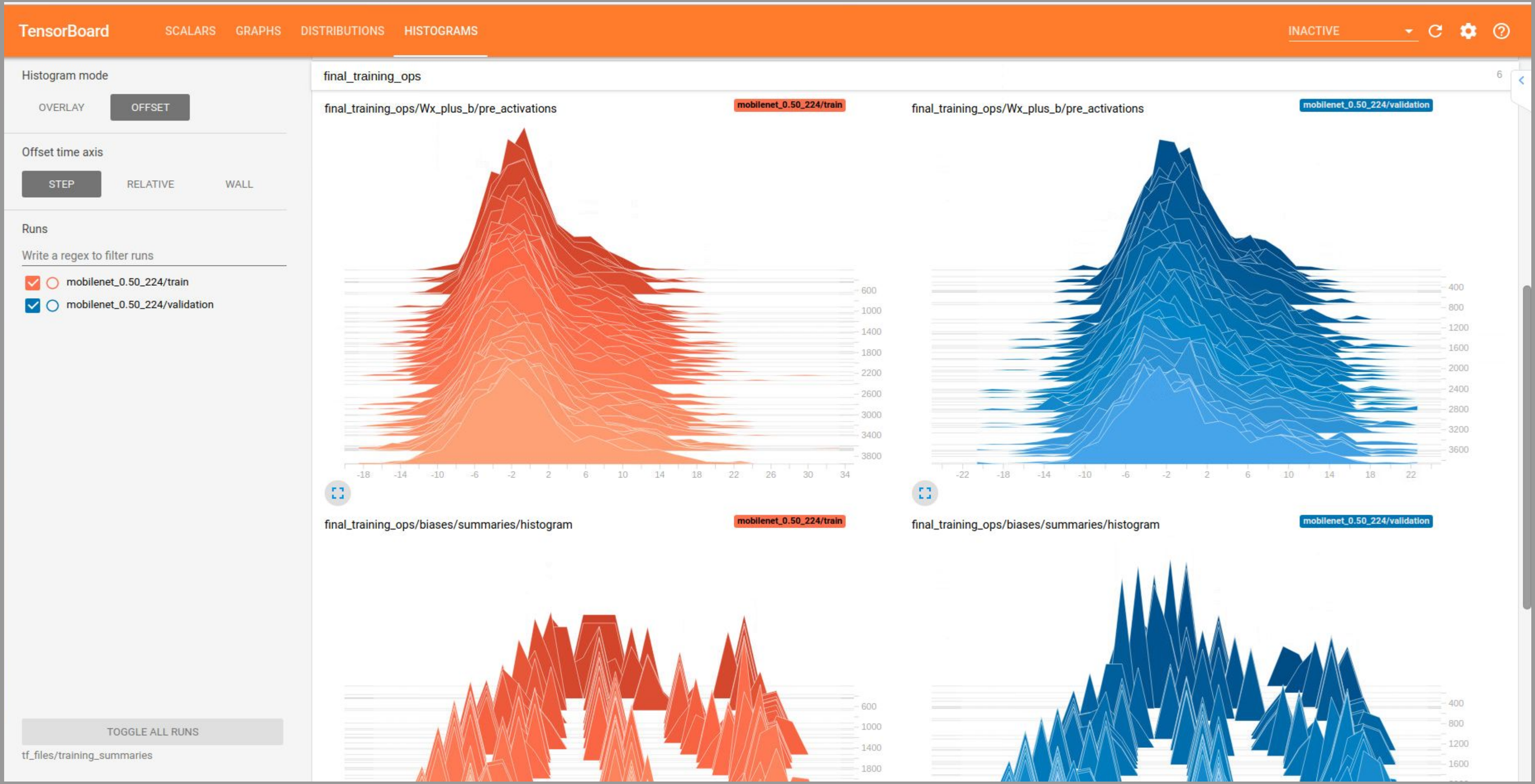


(b) Inception module with dimensionality reduction

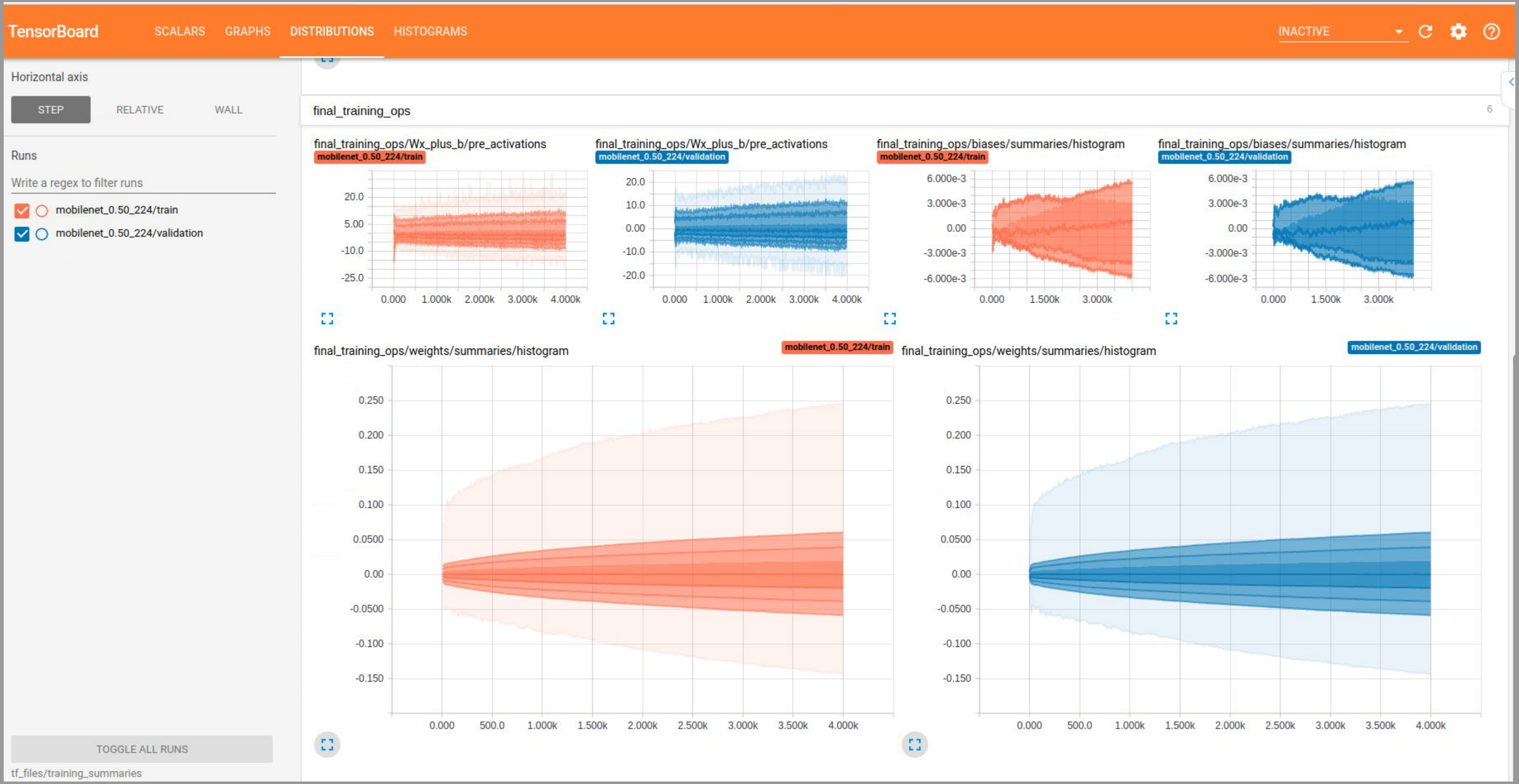
Inception V3 training Model architecture



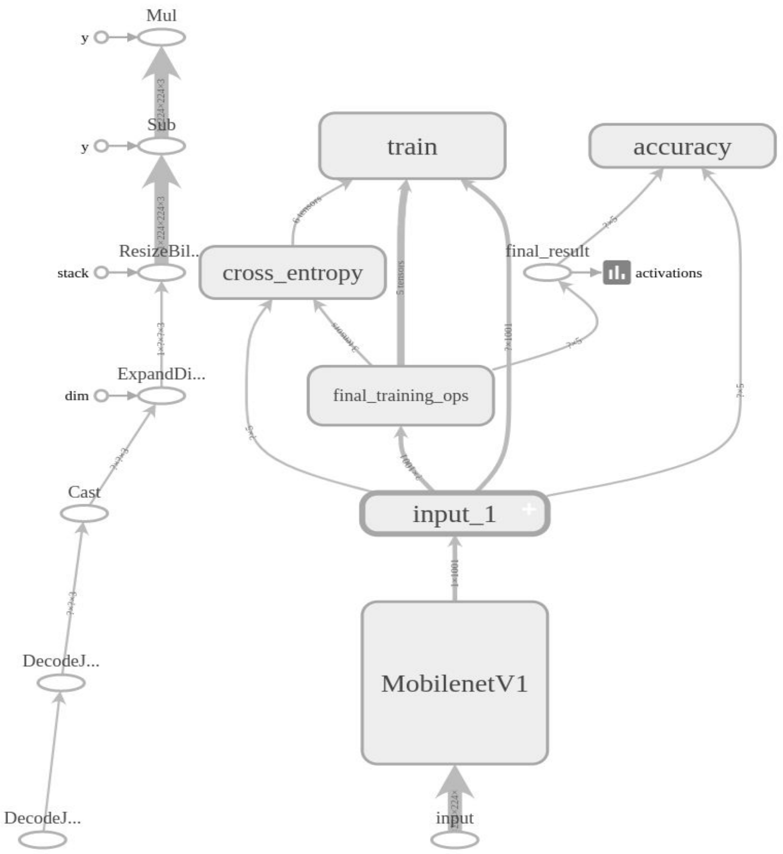
Pre-activation and biases graphs



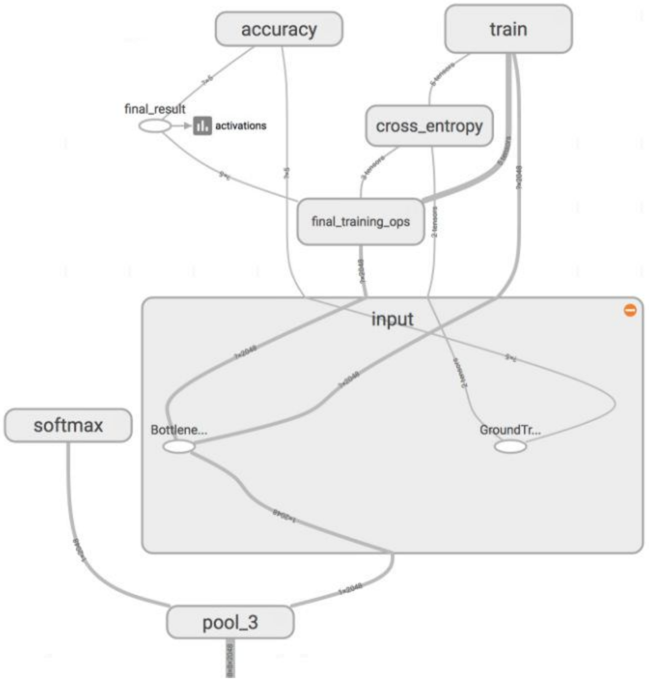
Cross validation graphs [red-training , blue-validation]



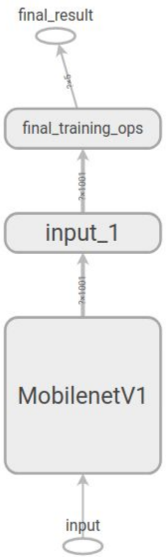
Data-Flow graph of our model : MobilenetV1 training using multiple pooling, cross validation and activation



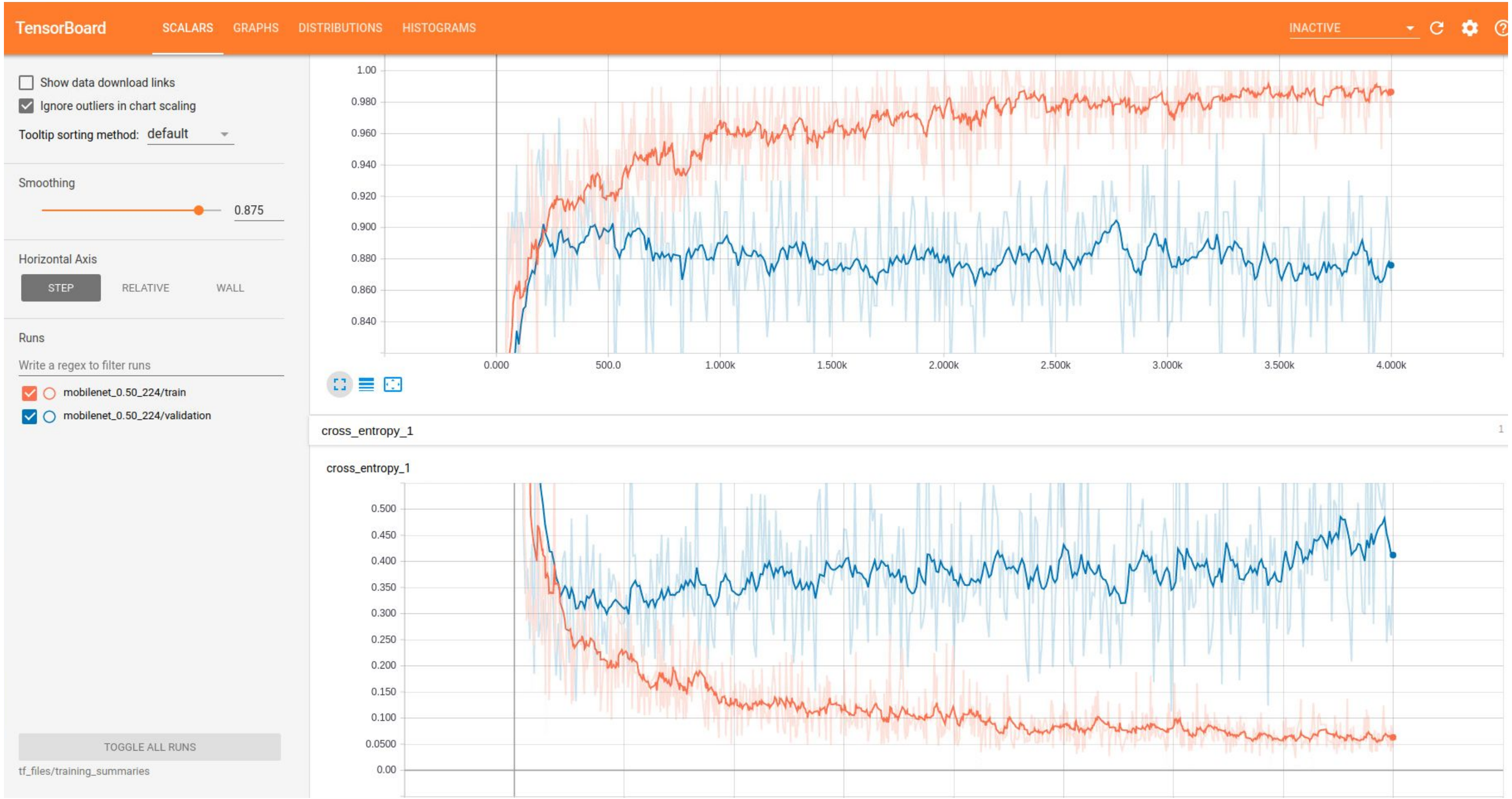
Training the final output layer



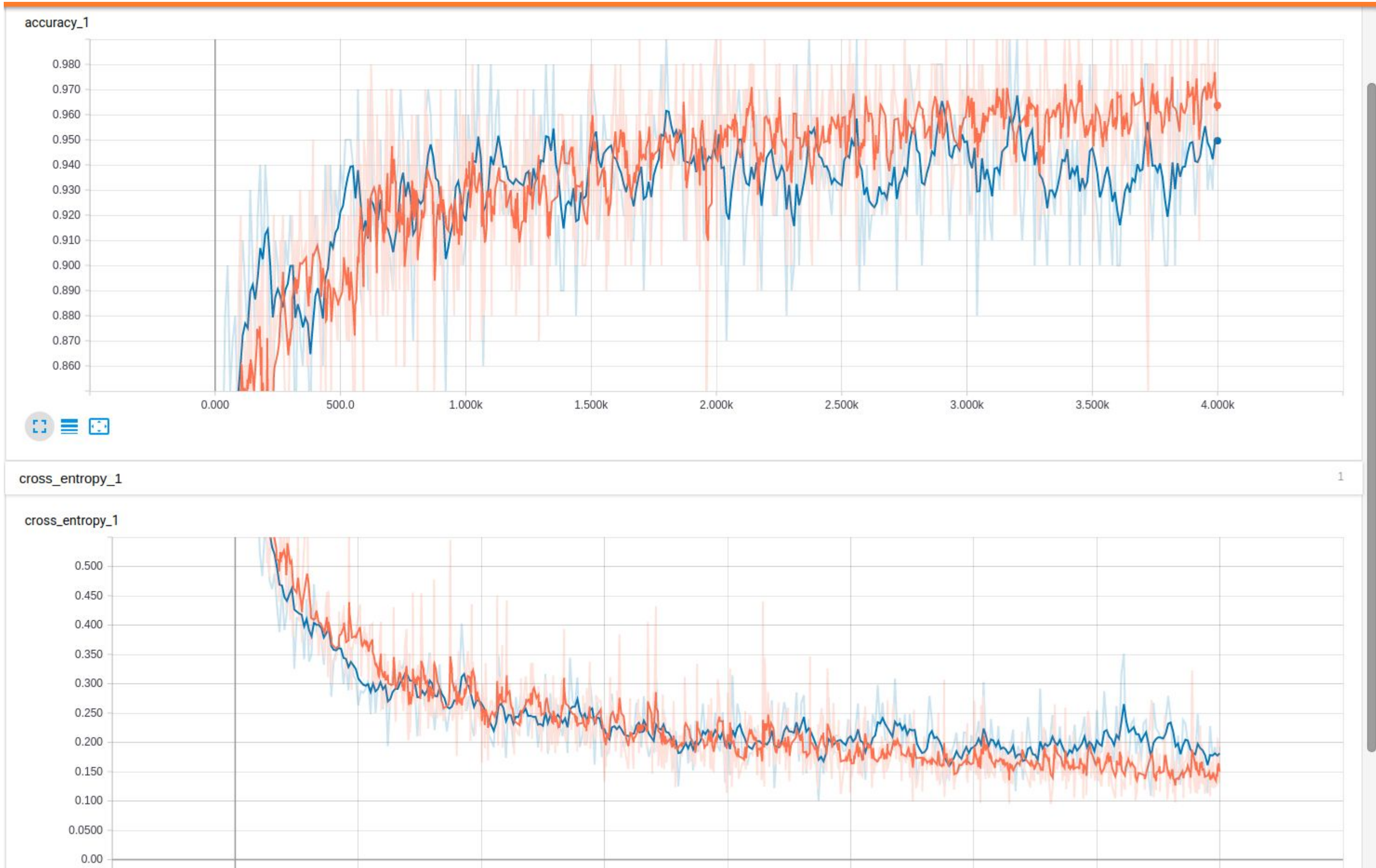
For the mobile platform : An optimized version of reduced complexity



Training epochs on MobileNet model (GooGLe net)

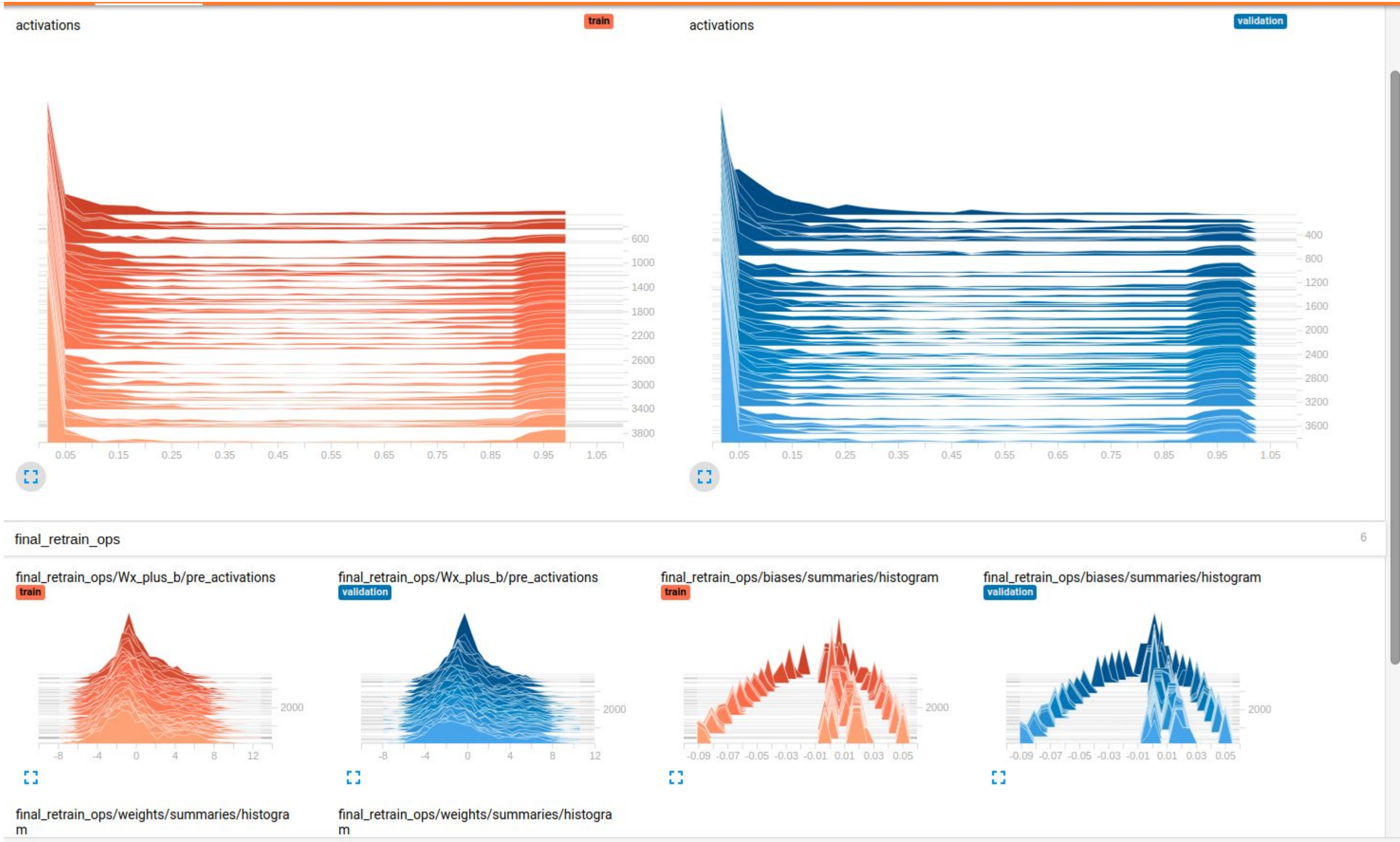


Training epochs on Inception V3 model

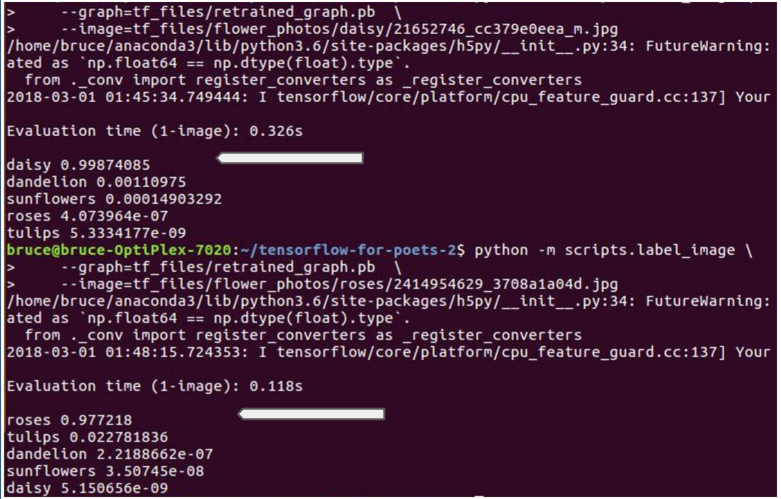


Cross validation graphs [red-training , blue-validation] : Inception V3

Pre-activation and bias graphs : Inception V3

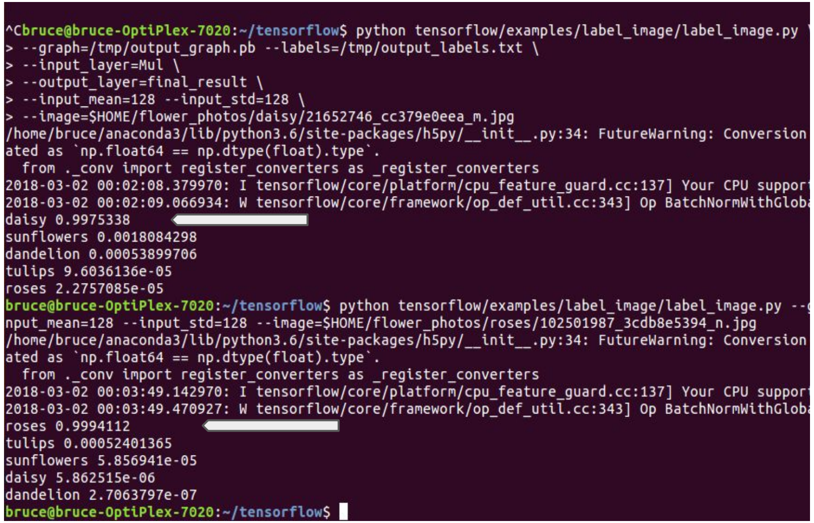


Results of classification of Flowers : MobileNet



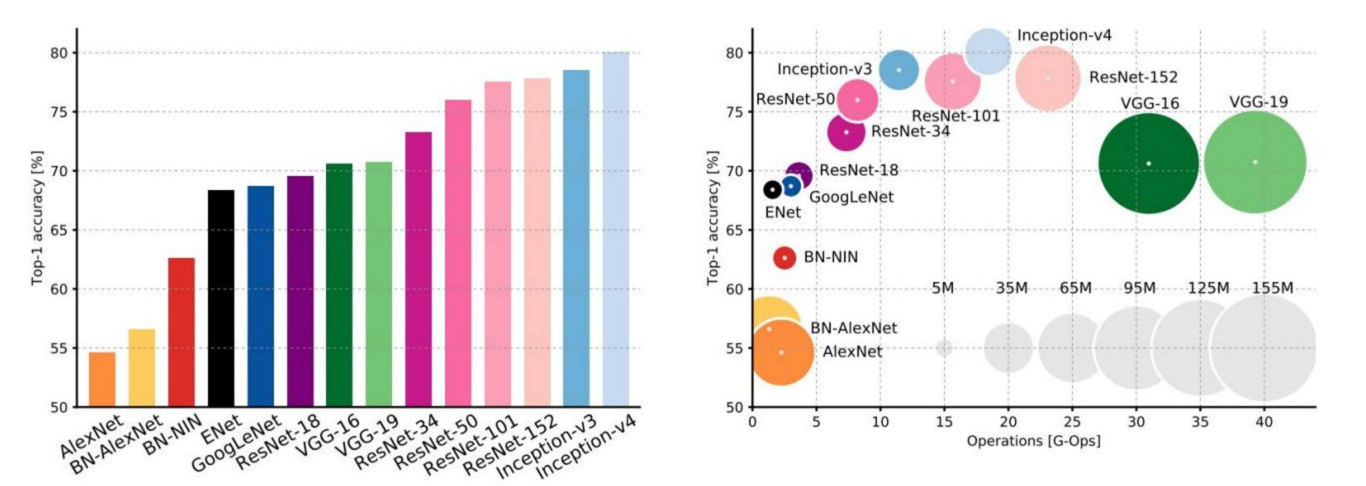
```
>     --graph=tf_files/retrained_graph.pb  \
>     --image=tf_files/flower_photos/daisy/21652746_cc379e0eea_m.jpg
/home/bruce/anaconda3/lib/python3.6/site-packages/h5py/__init__.py:34: FutureWarning:
ated as `np.float64 == np.dtype(float).type`.
  from ._conv import register_converters as _register_converters
2018-03-01 01:45:34.749444: I tensorflow/core/platform/cpu_feature_guard.cc:137] Your

Evaluation time (1-image): 0.326s

daisy 0.99874085
dandelion 0.00110975
sunflowers 0.00014903292
roses 4.073964e-07
tulips 5.3334177e-09
bruce@bruce-OptiPlex-7020:~/tensorflow-for-poets-2$ python -m scripts.label_image \
>     --graph=tf_files/retrained_graph.pb  \
>     --image=tf_files/flower_photos/roses/2414954629_3708a1a04d.jpg
/home/bruce/anaconda3/lib/python3.6/site-packages/h5py/__init__.py:34: FutureWarning:
ated as `np.float64 == np.dtype(float).type`.
  from ._conv import register_converters as _register_converters
2018-03-01 01:48:15.724353: I tensorflow/core/platform/cpu_feature_guard.cc:137] Your

Evaluation time (1-image): 0.118s

roses 0.977218
tulips 0.022781836
dandelion 2.2188662e-07
sunflowers 3.50745e-08
daisy 5.150656e-09
```

Results of classification of Flowers : Inception V3



```
^Cbruce@bruce-OptiPlex-7020:~/tensorflow$ python tensorflow/examples/label_image/label_image.py \
> --graph=/tmp/output_graph.pb --labels=/tmp/output_labels.txt \
> --input_layer=Mul \
> --output_layer=final_result \
> --input_mean=128 --input_std=128 \
> --image=$HOME/flower_photos/daisy/21652746_cc379e0eea_m.jpg
/home/bruce/anaconda3/lib/python3.6/site-packages/h5py/__init__.py:34: FutureWarning: Conversion
ated as `np.float64 == np.dtype(float).type`.
  from ._conv import register_converters as _register_converters
2018-03-02 00:02:08.379970: I tensorflow/core/platform/cpu_feature_guard.cc:137] Your CPU suppor
2018-03-02 00:02:09.066934: W tensorflow/core/framework/op_def_util.cc:343] Op BatchNormWithGloba
daisy 0.9975338
sunflowers 0.0018084298
dandelion 0.00053899706
tulips 9.6036136e-05
roses 2.2757085e-05
bruce@bruce-OptiPlex-7020:~/tensorflow$ python tensorflow/examples/label_image/label_image.py --g
nput_mean=128 --input_std=128 --image=$HOME/flower_photos/roses/102501987_3cdb8e5394_n.jpg
/home/bruce/anaconda3/lib/python3.6/site-packages/h5py/__init__.py:34: FutureWarning: Conversion
ated as `np.float64 == np.dtype(float).type`.
  from ._conv import register_converters as _register_converters
2018-03-02 00:03:49.142970: I tensorflow/core/platform/cpu_feature_guard.cc:137] Your CPU suppor
2018-03-02 00:03:49.470927: W tensorflow/core/framework/op_def_util.cc:343] Op BatchNormWithGloba
roses 0.9994112
tulips 0.00052401365
sunflowers 5.856941e-05
daisy 5.862515e-06
dandelion 2.7063797e-07
bruce@bruce-OptiPlex-7020:~/tensorflow$
```

Comparing performance of today's Deep Neural Networks



An Analysis of Deep Neural Network Models for Practical Applications, 2017.

Figure reference http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture9.pdf:

So, can we put these deep learning models into a human brain and expect it to function like the visual cortex layers for object classification ?

- No, not just yet !

- To apply it on a brain like platform, major limitation is that of complexity at information transfer pathway & energy usage [human brains work at ~20 Watts]

- Each layer increases complexity by a quadritic function. If we add more paraters or neurons or hidden layers but their weights close to zero, less increase in accuracy at cost of increased complexity

- Bigger size of network means increase in complexity, hence more prone to overfitting

Neuroscience with Neural Networks : Comparison

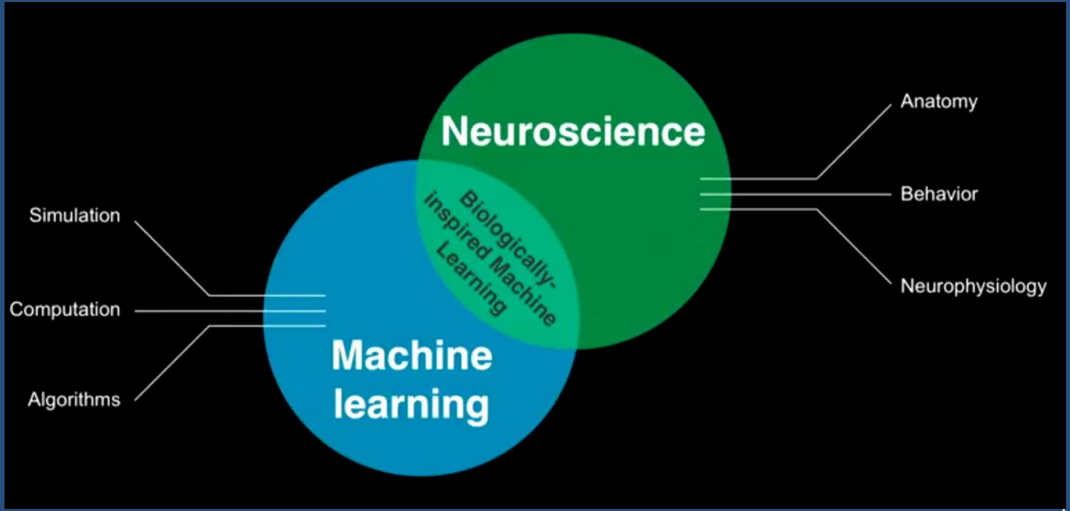


Figure reference : TED.com

Biologically Feasible computing

Observation

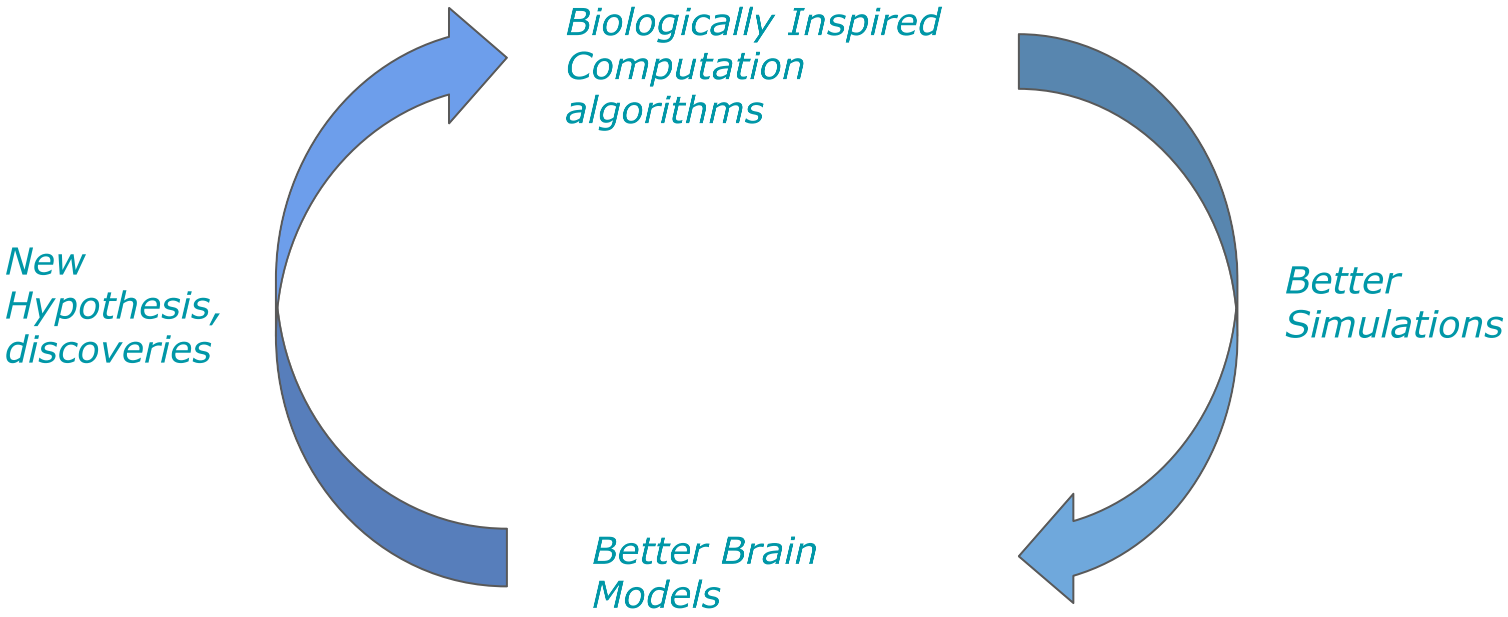
Algorithm

(How good Human visual system is)

(What can be done using resources available)

Biologically Feasible computing
Examples : BRAIN, IBM, Intel

Biologically Feasible Feedback loop



Scaling down the model on a brain like processor and energy intensive platform

To overcome the limitation of bottlenecks at information transfer pathway & minimize energy usage, we tried to constrain whole deep learning system's algorithm and compress it's weights while minimally dropping the model's accuracy.

Scaling down the model on a brain like processor and energy intensive platform : Deployment

Optimize the model architecture into a simpler version. Drop in accuracy <1%. (Refer earlier slide for figure)

Can use a fused model : Deep learning with Probabilistic model such as Nisarg Shah's Markov Decision Process to decrease order of symbolic steps , hence reduce propagation time w.r.t visual layers (Refer to earlier figure)

The retrained model was still 100 MB in size. This large size of data might be a limiting factor for any algorithm being processed in the brain. Normal zip compression algorithm : Only 8% compression obtained

Without any changes to the structure of the network, all floating point weights were quantized with the constants in place. Successfully reached 70% compression ratio with a drop in accuracy of only 1-2%.

Further improvements : Neuroscience with Neural Networks

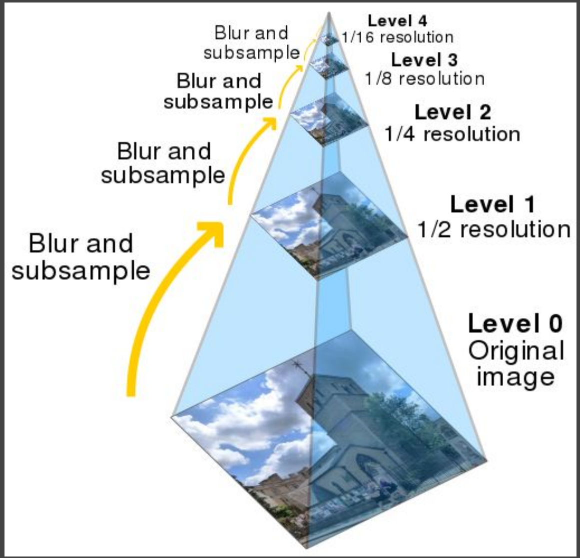
R-CNN and LSTM networks : how they take context of the whole image into account for recognising the objects. Similarities can be drawn with Bayesian model of the HVS.

Inferotemporal cortex neurons take into account sequences in object recognition tasks .

Multimodal systems : Take into account the integrating part of different disparate systems and the ventral visual cortical responses happen with feedback loops along with motor and other sensory systems

Zero Shot learning : Correct response in the first exposure itself **[Is the flower beautiful ?]** : **Instinct, context**

Implementation of the Pyramid model for amplifying sequential and temporal features



Extract temporal and spatial features by decomposing image into Laplacian pyramidal layers

Enhancing or amplifying those features in the Laplacian layers

Stitching back the video to show the magnified image or video for comparison with the normal Human Visual System : this signifies the direction of the extension of HVS model using a biologically inspired model at backend

Figure : reference : wikipedia

Human eye vs Biologically inspired vision : working with sequential data

Another Exercise ?

[again for humans first, then machines]

Do you 'see' the difference in the contiguous frames ?

Video Reference : MIT CSAIL lab

Our algorithm can ‘see’ the difference in the contiguous frames

Can you ‘see’ it now ?

An enhanced, biologically inspired algorithm to see something that was not in realm of reality perception of human visual system

Video Reference : MIT CSAIL lab

Final notes to take-away amid all this chaos

So we have to constraint the current machine learning and deep learning systems to process less info, more effectively *(Like in the compression of mobilenet model)*

Make decisions based on instinct, context and zero-shot learning .

Security in DNN(noise to misclassify images) : GANs

This ‘general artificial intelligence’ has to learn and surpass the biological algorithm running in our brain all the time.

That is how AI systems will help us to extend and augment our abilities, and not only emulate them. *(Like in the pyramid model)*

References

- http://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf
- http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture9.pdf
- https://codelabs.developers.google.com/codelabs/tensorflow-for-poets/index.html#3
- https://codelabs.developers.google.com/codelabs/tensorflow-for-poets-2/#1
- https://www.tensorflow.org/tutorials/image_retraining
- http://neuralnetworksanddeeplearning.com/index.html
- http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/
- https://youtu.be/iPdKMs9cEAs (Machine learning + neuroscience = biologically feasible computing | Benjamin Migliori | TEDxSanDiego)