

Deep Learning Case Study: Face Recognition Development

November 7, 2018

Chantana Chantrapornchai

HPCNC Laboratory, Dept. of Computer Engineering

Kasetsart University





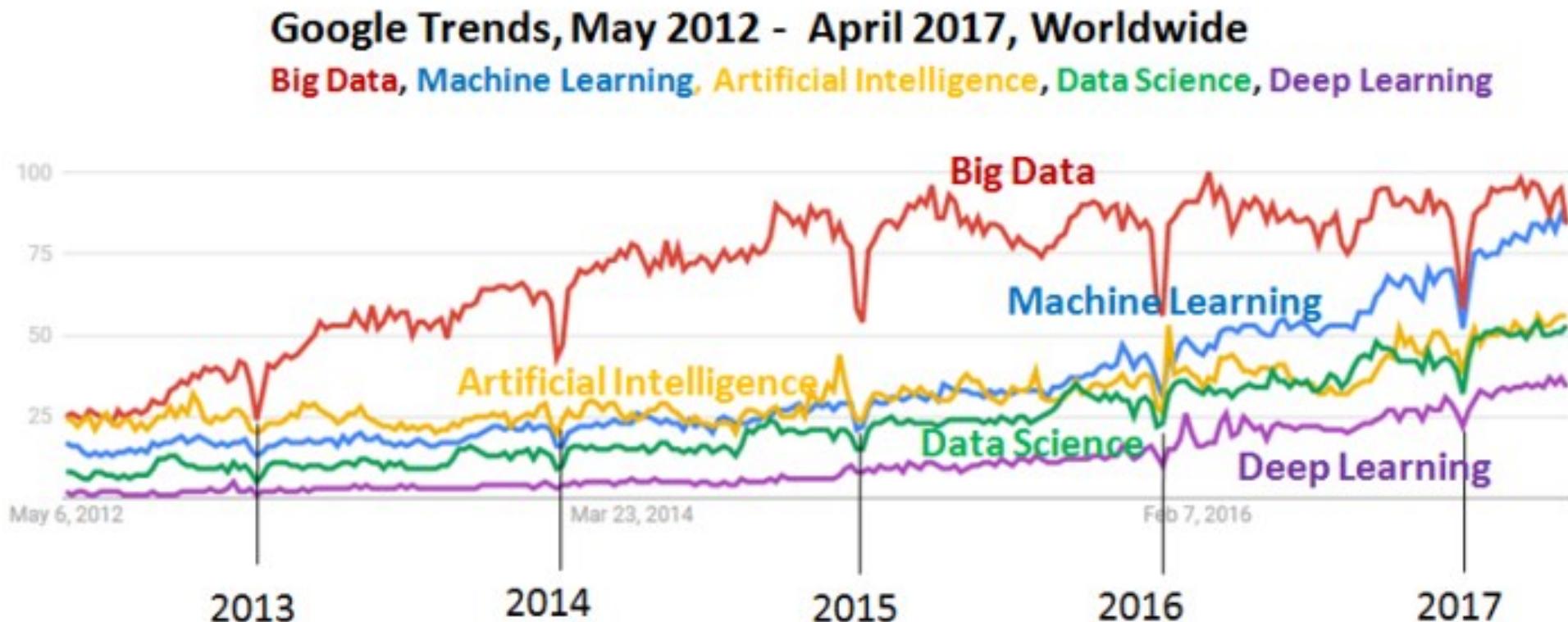
https://github.com/cchantra/face_training

Outline

- Introduction to deep learning
- Top applications of deep learning
- Case study: Hands on Face recognition

Google Trends

1 Machine Learning is rising like a true champion, 2) Deep Learning is new and rising fast.



<https://www.kdnuggets.com/2017/05/machine-learning-overtaking-big-data.html>

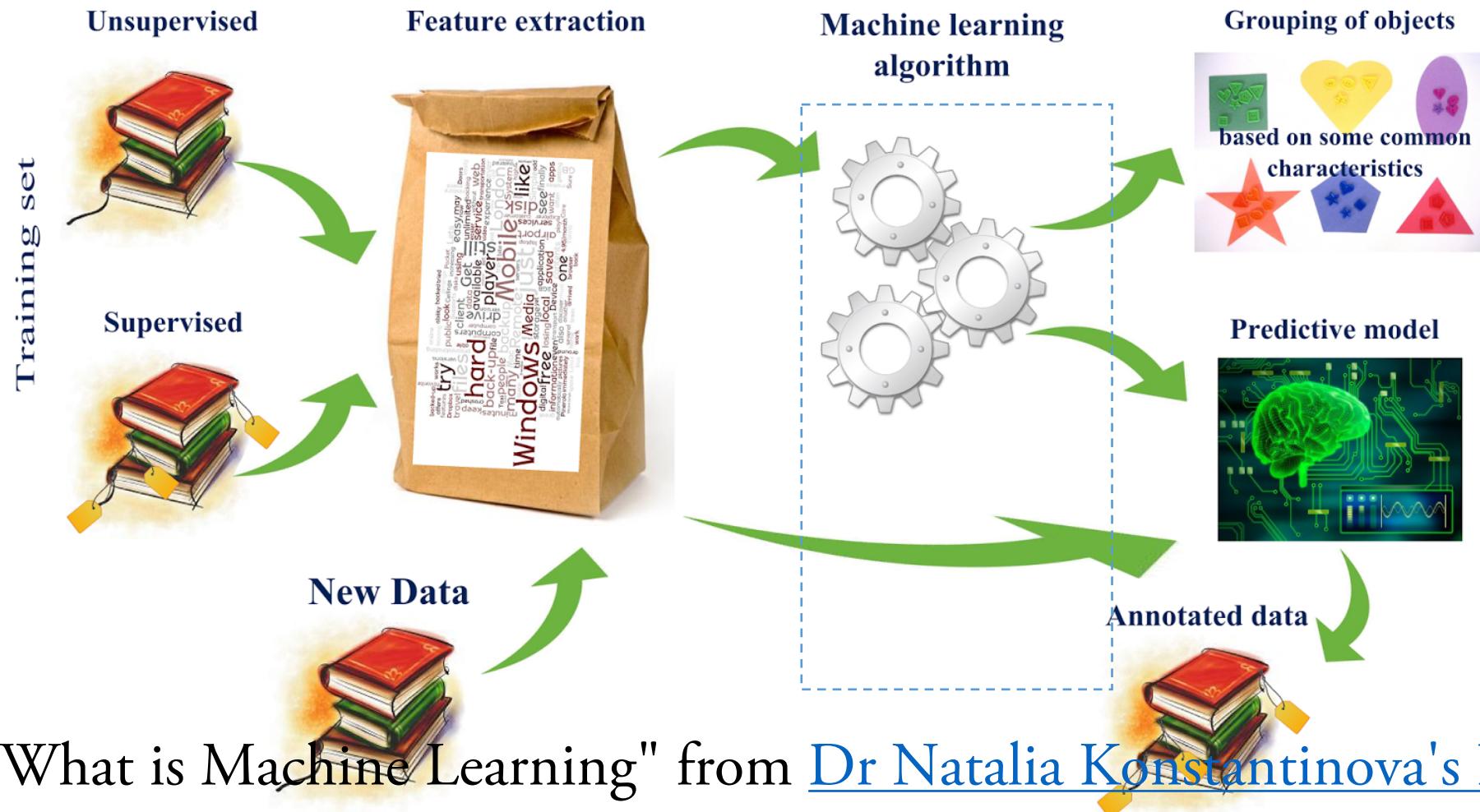
Rising from Pattern Recognition

- Getting a computer program to do something “smart” like recognize the character “3”
- Optical Character Recognition grew out of this community (OCR)
- It is fair to call “Pattern Recognition” as the “Smart” Signal Processing of the 70s, 80s, and early 90s.
- Decision trees, heuristics, quadratic discriminant analysis, etc. all came out of this era. Pattern Recognition became something CS people did like, and not EE folks. – e.g. template matching

Machine learning

- A powerful way to build pattern recognition algorithms is to **replace an expert (who probably knows way too much about pixels)** with data (which can be mined from cheap laborers).
- Collect lots of **face images** and **non-face images**, choose an algorithm, and wait for the computations to finish.

Machine learning workflow



"What is Machine Learning" from [Dr Natalia Konstantinova's Blog](#). The most important part of this diagram are the "Gears" which suggests that crunching/working/computing is an important step in the ML pipeline.

Application Areas

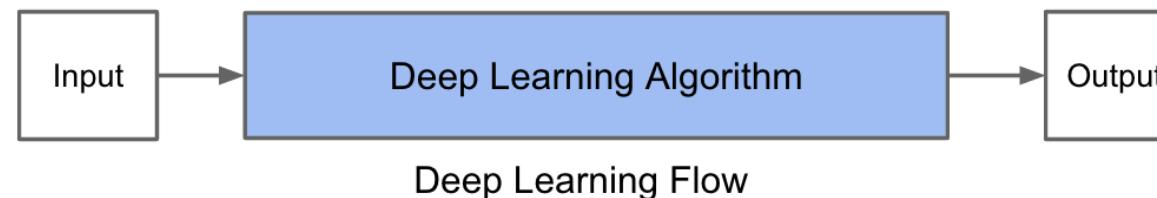
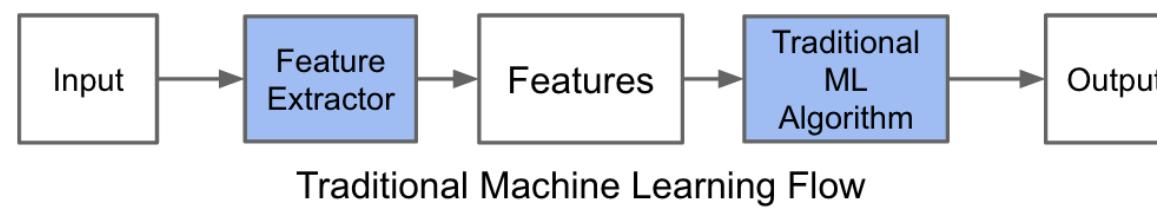
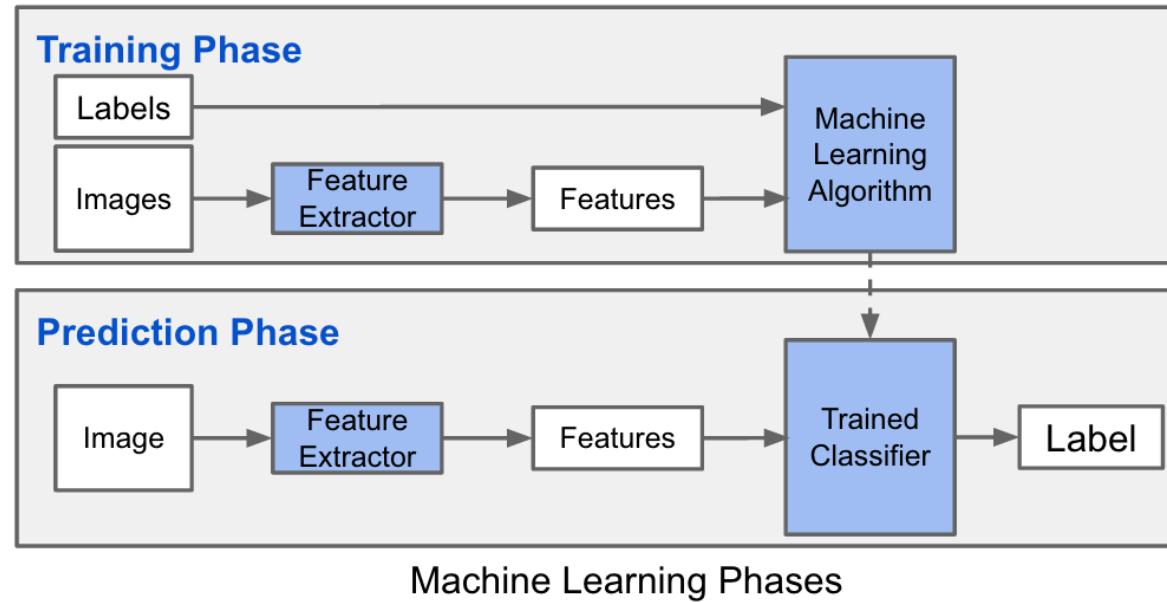
- Researchers apply Machine Learning to Robotics (reinforcement learning, manipulation, motion planning, grasping, vision), to genome data, to remote sensing data as well as to predict financial markets.
- One of the best application areas for machine learning for many years was computer vision.

Until recently neural networks were all but shunned by the AI research community. The problem is the most basic computation of neural networks were very computationally intensive, it just wasn't a practical approach.

But it wasn't until GPUs were deployed in the effort that the promise was realized.

“Andrew Ng’s breakthrough was to take these neural networks, and essentially make them huge, increase the layers and the neurons, and then run massive amounts of data through the system to train it. In Ng’s case it was images from 10 million YouTube videos. Ng put the “deep” in deep learning, which describes all the layers in these neural networks.”

--Google’s AlphaGo

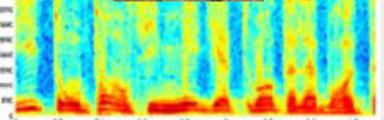
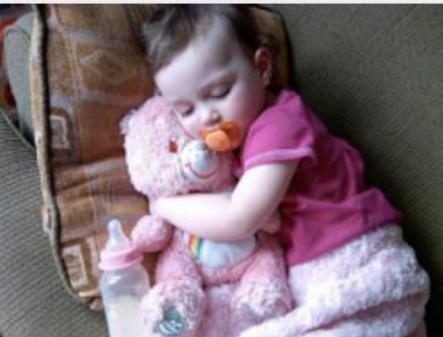


"What is Machine Learning" from [Dr Natalia Konstantinova's Blog](#). The most important part of this diagram are the "Gears" which suggests that crunching/working/computing is an important step in the ML pipeline.

Deep Learning

- Deep Learning has become the most popular approach to developing Artificial Intelligence (AI) –machines that **perceive and understand the world**
- The very first focus is currently on specific **perceptual tasks**, and there **are many successes**.
- GPUs have become the parallel platform for **for deep learning in research and production**.

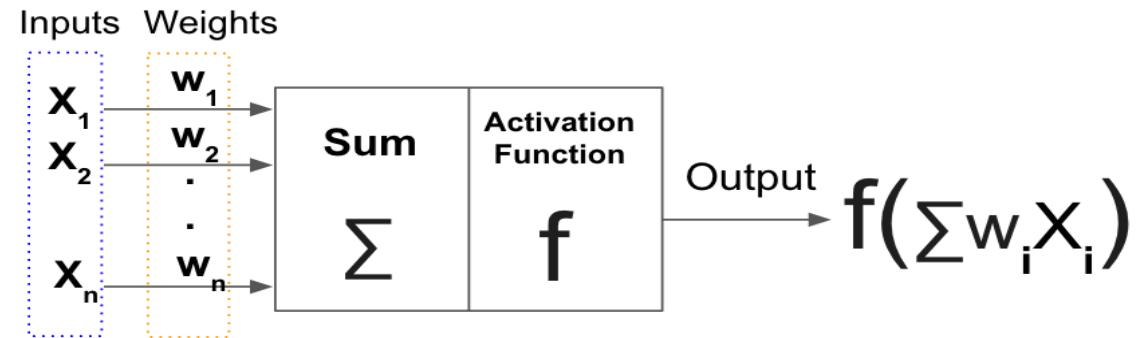
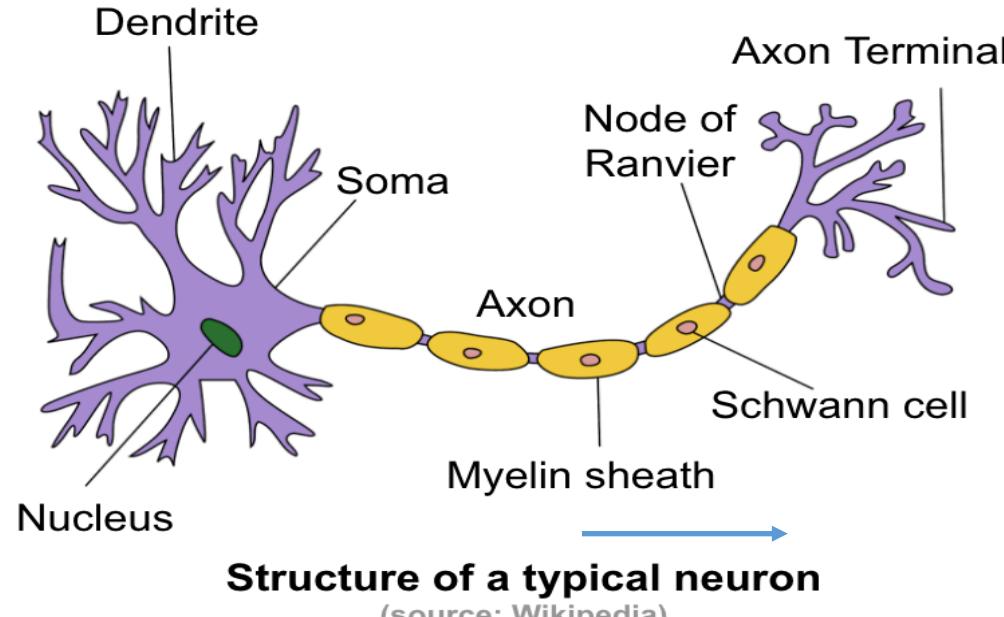
Examples

Input	Output
Pixels: 	“lion”
Audio: 	“see at tuhl res taur aun ts”
<query, doc>	P(click on doc)
“Hello, how are you?”	“Bonjour, comment allez-vous?”
Pixels: 	“A close up of a small child holding a stuffed animal”

Some Review of ANNs (Artificial Neural Networks)

Note: The tutorial slides are from many places.

- It is inspired by biological neural networks: Biological Neurons
- It processes and transmit information to other neurons by emitting electrical signals. Each neuron receives input signals from its dendrites and produces output signals along its axon. The axon branches out and connects via synapses to dendrites of other neurons.
- Artificial neurons are originated by biological neurons, and formulate the model in a computational form
 - has a finite number of inputs with weights associated to them,
 - an activation function (also called transfer function)
 - the output of the neuron is the result of the activation function applied to the weighted sum of inputs.
 - Artificial neurons are connected with each others to form artificial neural networks.



Structure of artificial neuron

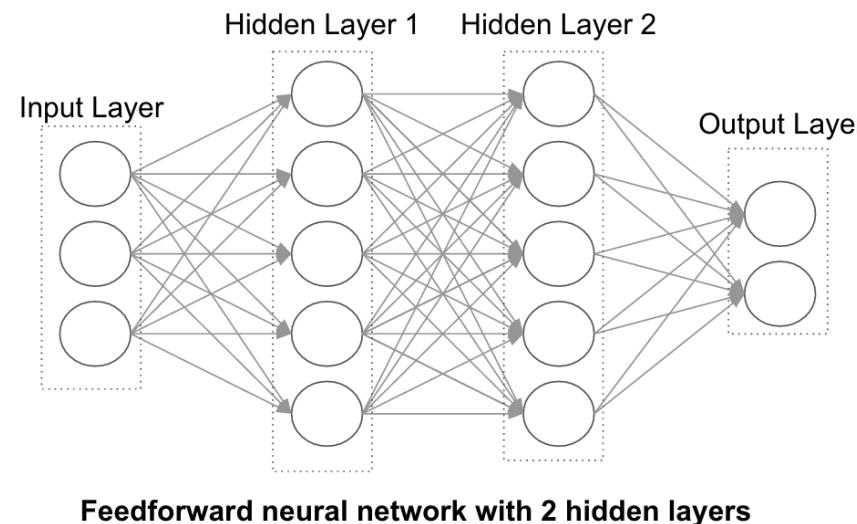
$$\text{e.g. } f(x) = \text{Max}(0, x)$$

Biological Neurons are the core components of the human brain. A neuron consists of a cell body, dendrites, and an axon.

Human brain contains about 10,000 computational power than the computer's

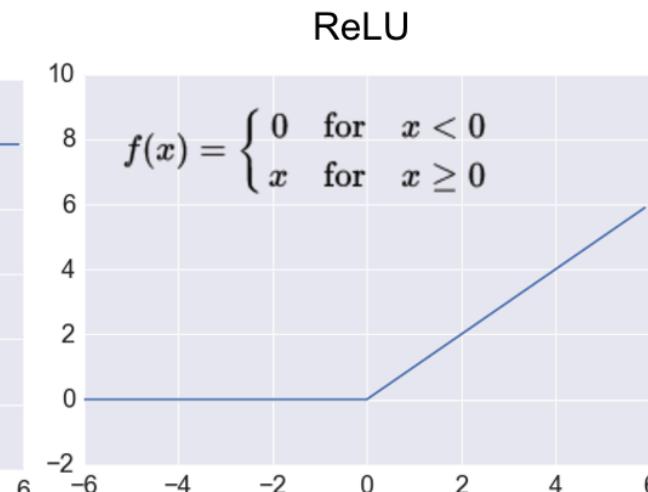
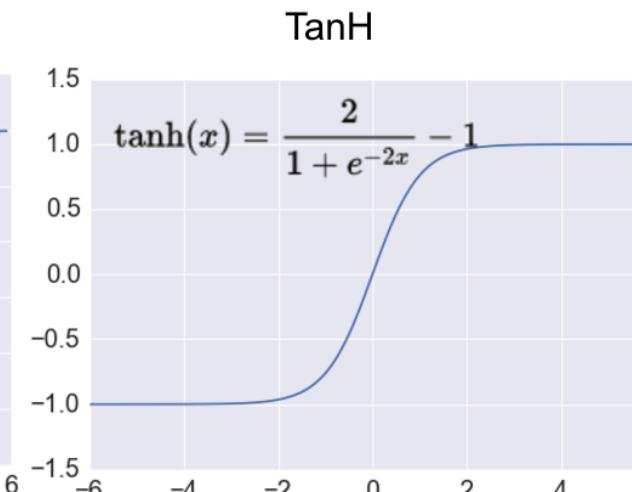
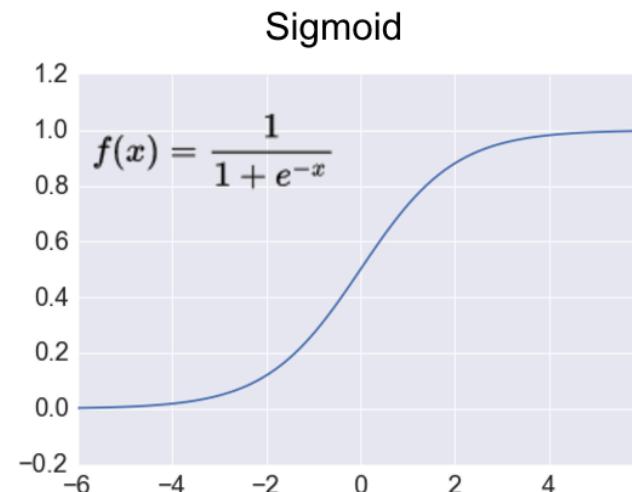
Feedforward Neural Networks

- These networks have **3 types of layers: Input layer, hidden layer and output layer.** In these networks, data moves from the input layer through the hidden nodes (if any) and to the output nodes.
- "Fully-connected" means that **each node is connected to all the nodes in the next layer.**

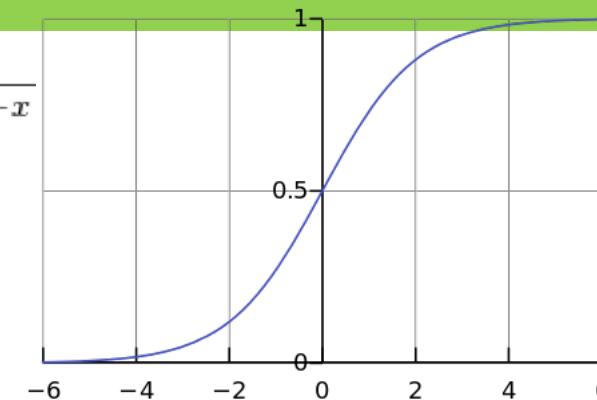


Activation Functions

- Activation functions transform the weighted sum of inputs that goes into the artificial neurons.
- The most popular activation functions are Sigmoid, Tanh and ReLU (rectified linear unit). ReLU is the most popular activation function in deep neural networks.



$$f(x) = \frac{1}{1 + e^{-x}}$$



-0.06
2.7

-2.5
-8.6

0.002

1.4

$f(x)$

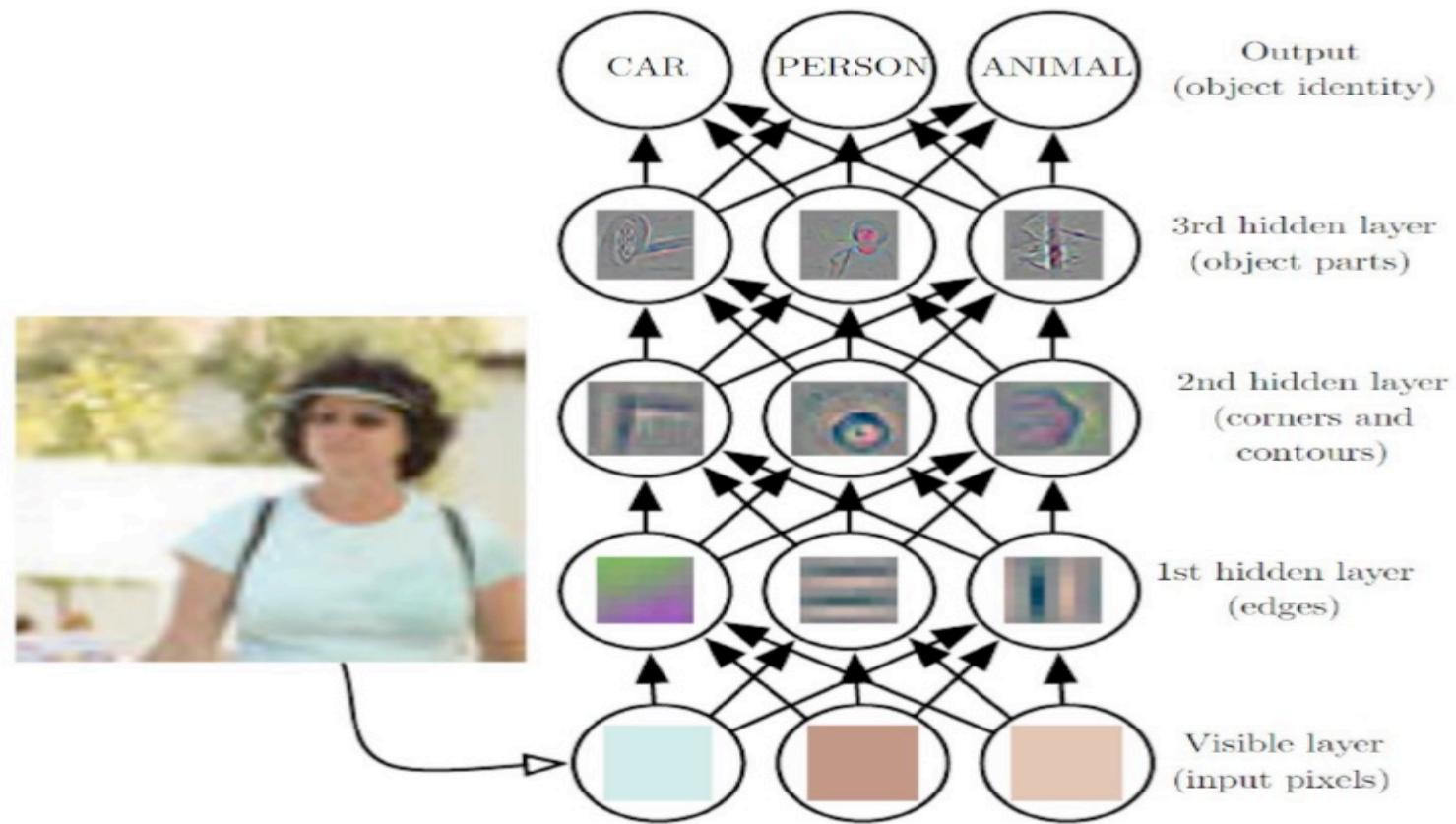
$$x = -0.06 \times 2.7 + 2.5 \times 8.6 + 1.4 \times 0.002 = 21.34$$

Training Artificial Neural Networks

- The goal of the training phase is to learn the network's weights.
- 2 elements to train an artificial neural network:
 - Training data: In the case of image classification, the training data is composed of images and the corresponding labels.
 - Loss function: A function that measures the inaccuracy of predictions.
- We train the ANN using an algorithm called backpropagation together with gradient descent (or one of its derivatives).

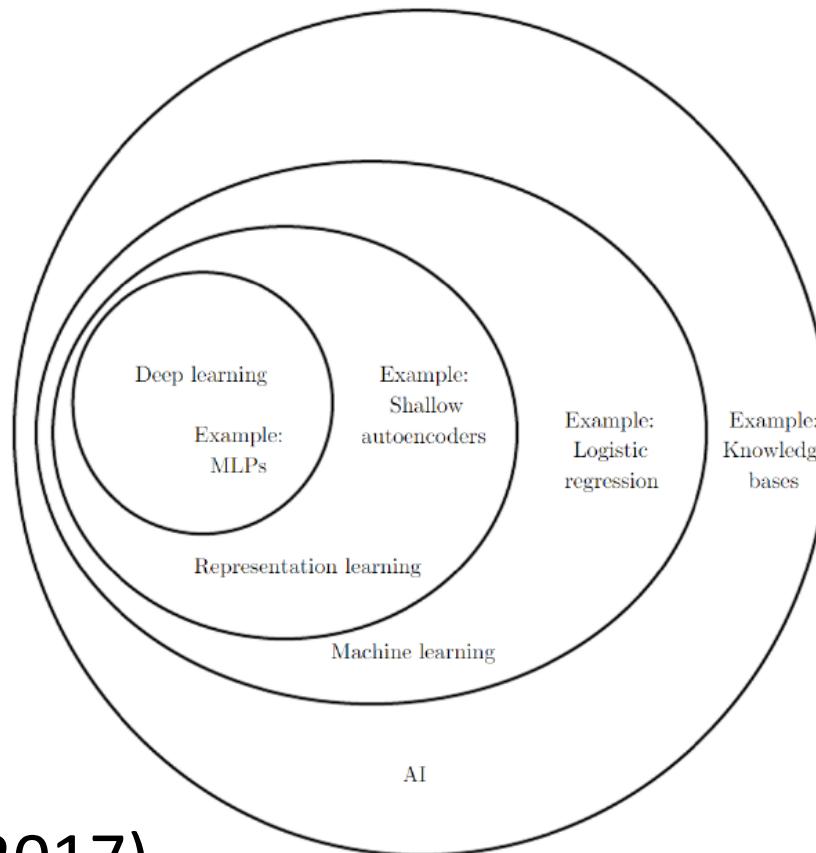
<https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

Deep learning is a representation learning

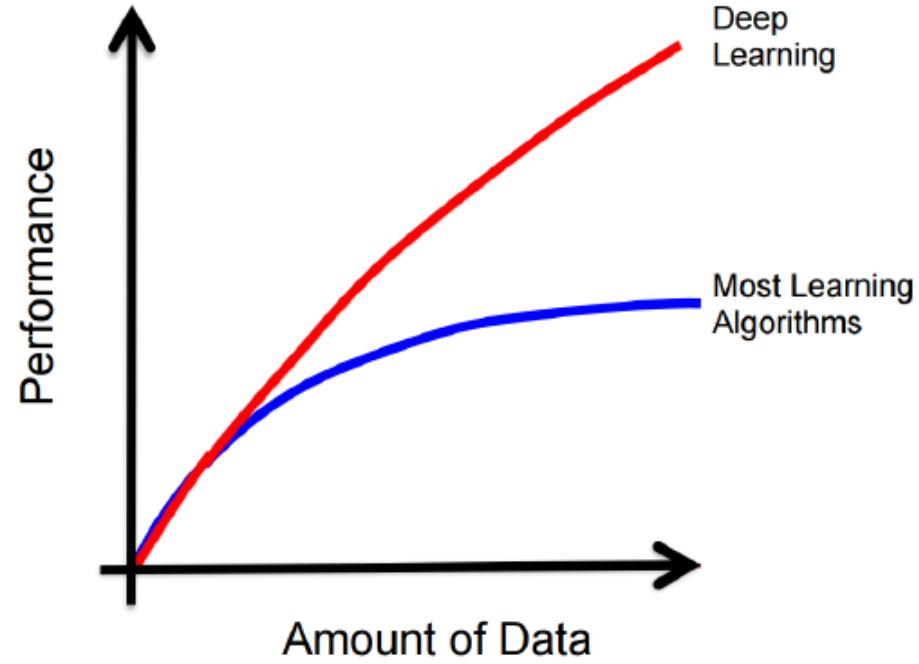


Goodfellow et al. "Deep learning." (2017).

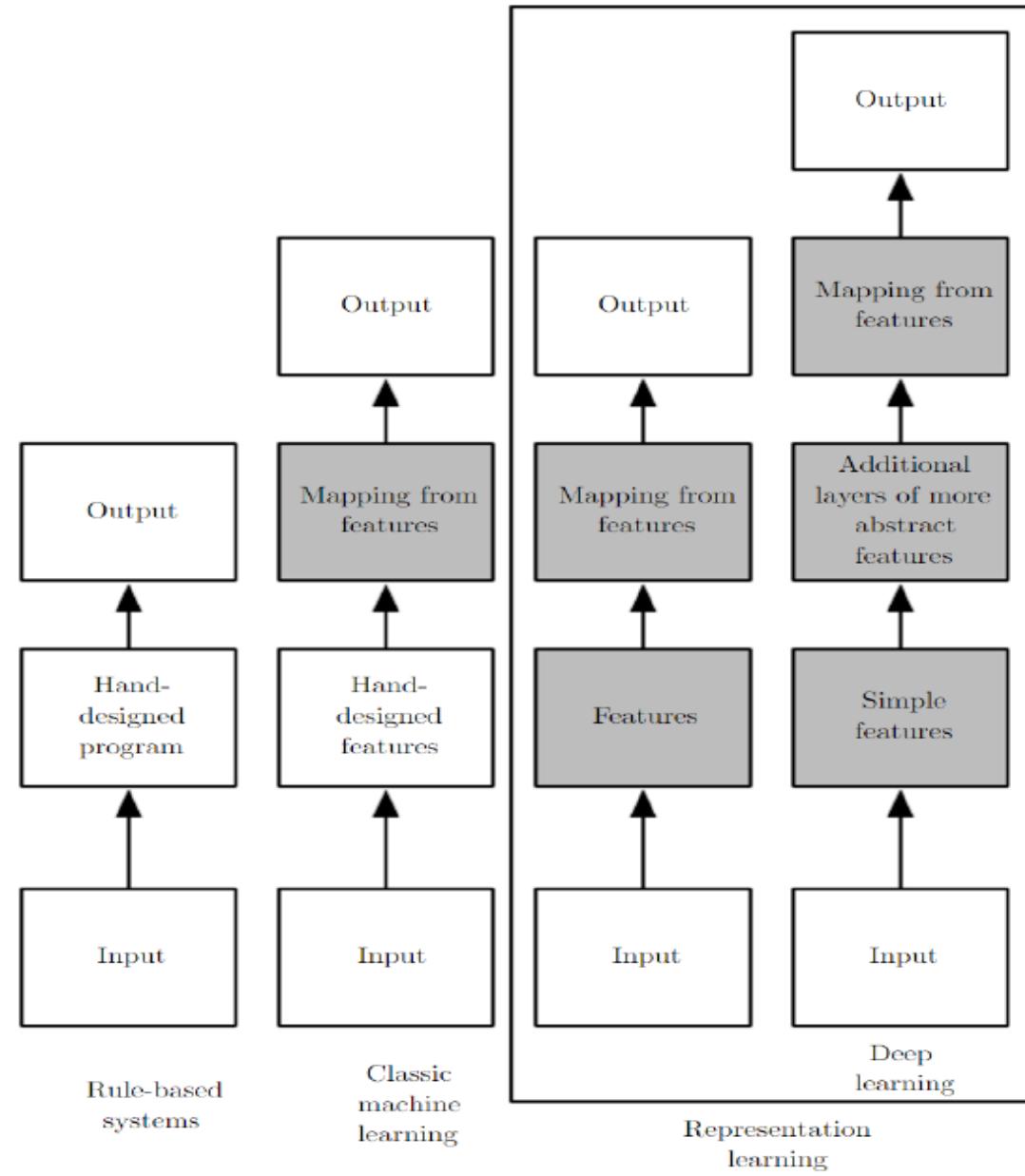
Deep learning is a representation learning



Ian Good Fellow et.al. (2017)
Deep learning.



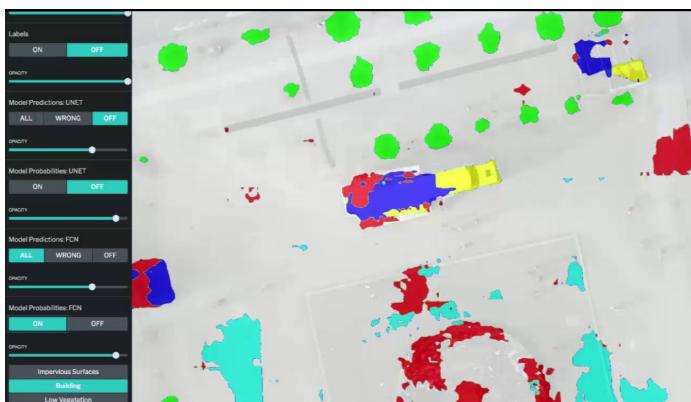
The more data...



Applications

Agriculture/ Land use

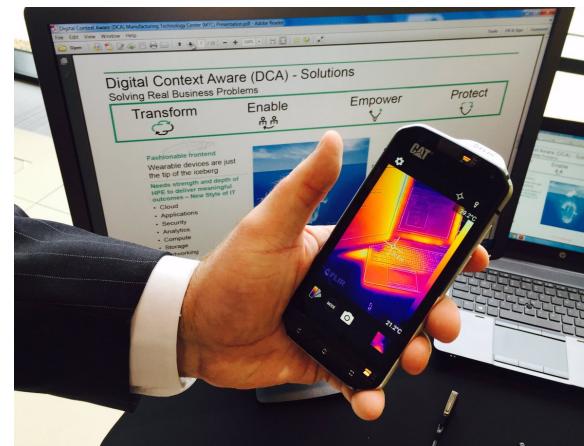
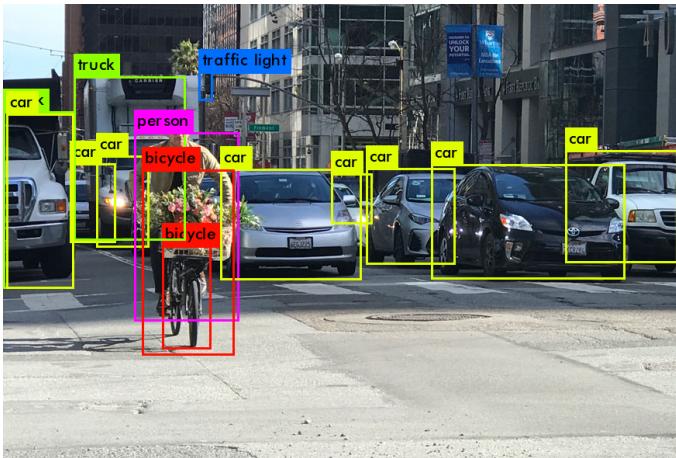
- Weed/Plant classification
- Productivity prediction
- Fruit inspection: NIR images
- Crop disease



Manufacturing

Automatic detection

- Detecting overheating (Equipment)
- QC of Finished products
- Glass defect detection



Video Classification/Semantics

Majorly use CNN: Describe event/ semantic

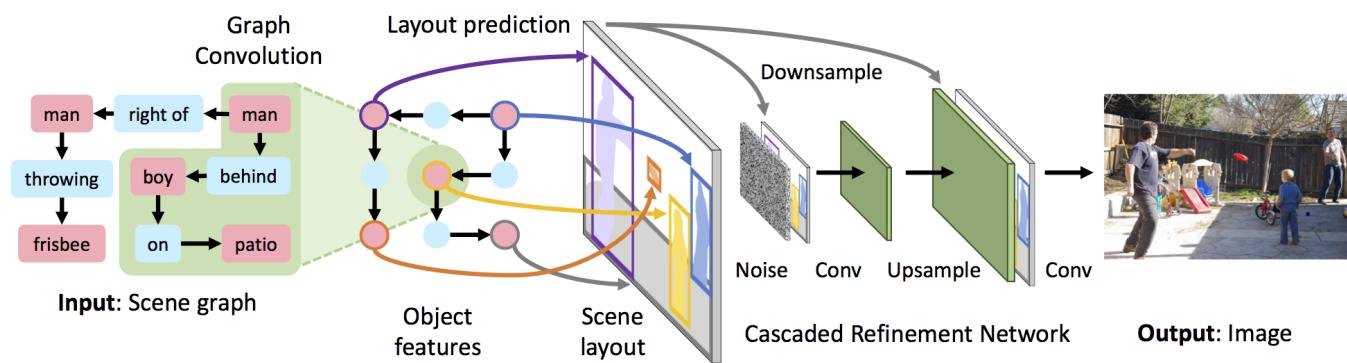
- Have **temporal information**: learning a global description of the video's temporal evolution is important.

- Approaches:

- Feature pooling
 - Process each frame using CNN and combine using pooling layers.
- Long Short Term Memory (LSTM)/GRU
 - Use memory cells to store, modify, access internal states. Operate on each frame using CNN but integrate information overtime.

- 3D-CNN

- ...



Application areas

- Data security: Predict malware, suspicious attack.
- Personal security: security screenings at airports, stadium.
- Financial trading
- Healthcare: compute assisted diagnosis (images)
- Marketing personalization: eg. Facebook ads
- Fraud detection: suspicious transfer transaction.
- Recommendations/suggestion: TV program, Song
- NLP- Text summarize, Song/poem generation, sentence prediction
- Self driving cars: Sensors

<https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/#74ff4c2f94c9>

Classic Problems and models

- Image classification
 - Available pretrained net: AlexNet, SqueezeNet, GoogLeNet, ResNet
- Object bounding box
 - RCNN/Fast RCNN/ Faster RCNN/ YOLO/ FPN
 - Pixelwise: FCN/EddyNet
- Semantics in language
 - RNN, GRU, LSTM
- Video semantic
 - 3DCNN, LSTM
- Image generation
 - GAN



https://medium.com/@jonathan_hui/gan-whats-generative-adversarial-networks-and-its-application-f39ed278ef09

Face recognition

- Face recognition applications in military
- Integrate deep learning technology to industries /real-world applications.

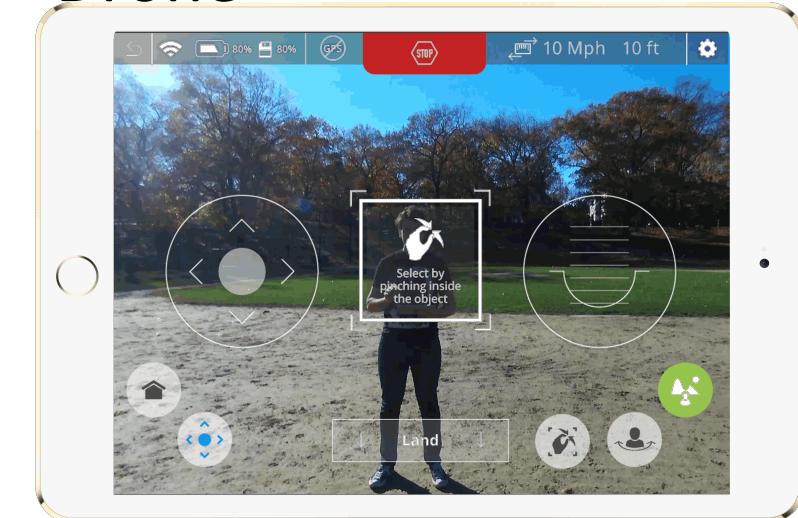


Criminal



UK, the world's most surveilled state, begins using automated face recognition to catch criminals

Drone



Retailers Increasingly Using Facial Recognition To Spot Thieves

Literature Review

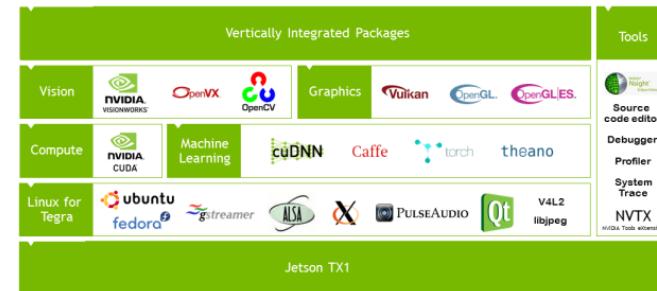
- Deep Face, FaceNet, Parkhi et.al (2015), Sarkar et.al. (2016).... so many
- Open source githubs

Methodology

(Web service?)

<http://www.iis.ee.ic.ac.uk/cxiong/database.html>

JETSON SDK



Deployment



Model Deployment



Linux Centos 7.0

The screenshot shows the DIGITS web interface running in Mozilla Firefox. It displays a dashboard with sections for "Datasets" and "Models". Under "Datasets", there are sections for "In progress" and "Completed". Under "Models", there are sections for "In progress" and "Completed". Both sections show a list of items with details like submission time, status, and delete options.

Dataset	Status	Details
dataset_imagenet@256x256	In progress	Submitted: 05-29-57 PM (63 seconds ago) Status: Running
dataset_mnist_10k@256x256	Completed	Submitted: 05-21-27 PM (12 minutes, 7 seconds ago) Status: Done after 2 minutes, 26 seconds
voc_cropper@256x256	Completed	Submitted: 05-18-43 PM (12 minutes, 7 seconds ago) Status: Done after 2 minutes, 26 seconds

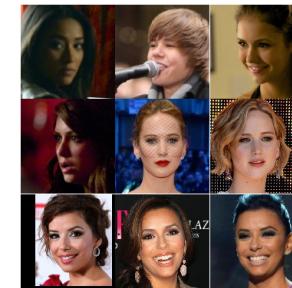
Model	Status	Details
model_mnist10k	In progress	Submitted: 05-30-48 PM (2 seconds ago) Status: Running
LeNet_model_voc_cropped@256x256	Completed	Submitted: 05-18-43 PM (12 minutes, 7 seconds ago) Status: Done after 2 minutes, 26 seconds

Public data sets

- LFW
- FaceScrub
- CASIA
- MegaFace
- MS-Celeb
- WIDER..



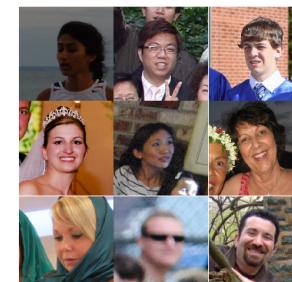
(a) LFW



(b) CASIA



(c) FaceScrub



(d) MegaFace

Available pretrained nets

- AlexNet
- CaffeNet
- GoogleNet
- SqueezeNet

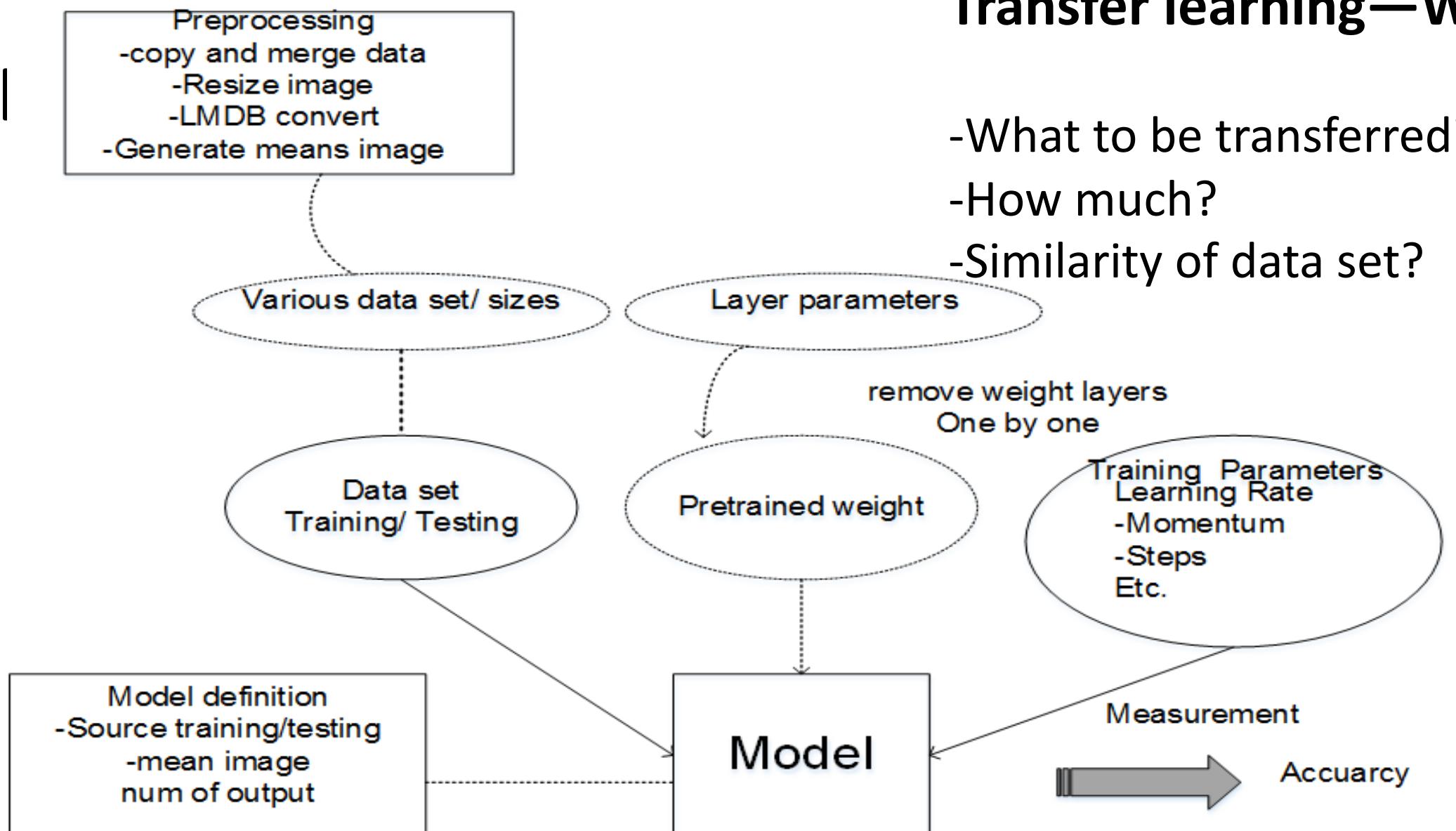
- ResNet
- VGG

Etc

Find more at model zoo.

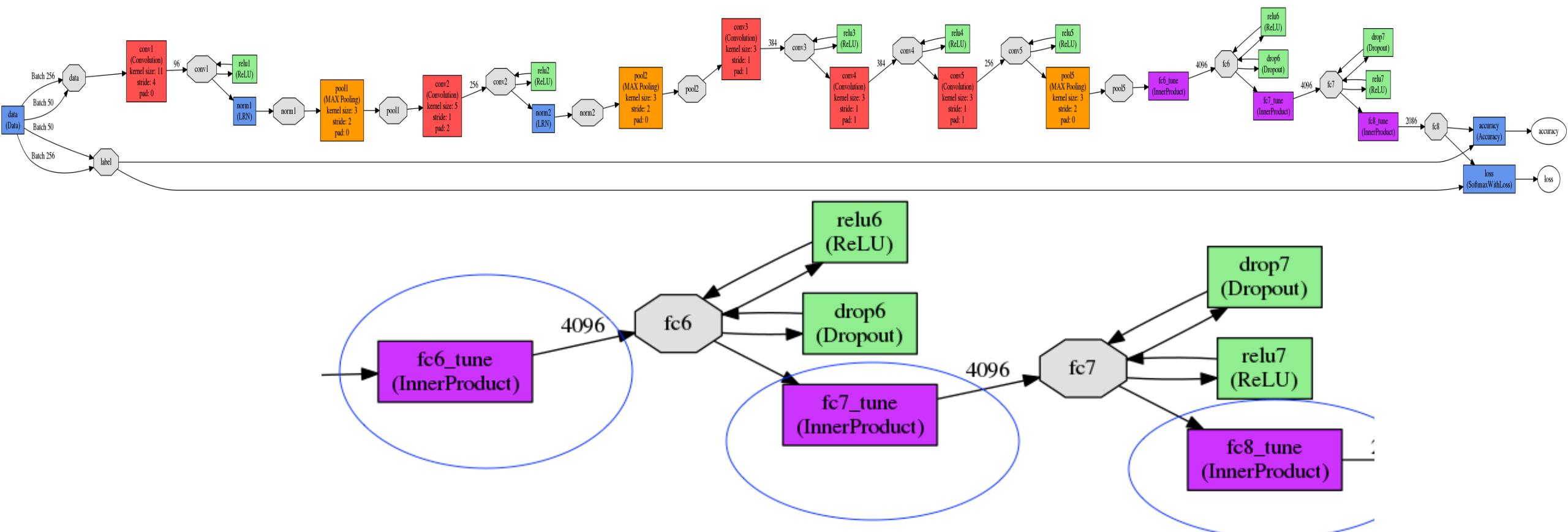
Type	AlexNet	GoogLe	Caffe	SqueezeNet	ResNet10	VGG16	RCNN
Conv	5	59	5	26	12	13	5
InnerProd	3	5	3	-	1	3	3
Pooling	3	16	3	3	2	5	3
ReLU	7	61	7	26	9	15	7
LRN	2	2	2	-	-	-	2
Dropout	2	3	2	1	-	2	2
Misc.					10 BN 10 Scale 4 Eltwise		

Transfer learning—What is?



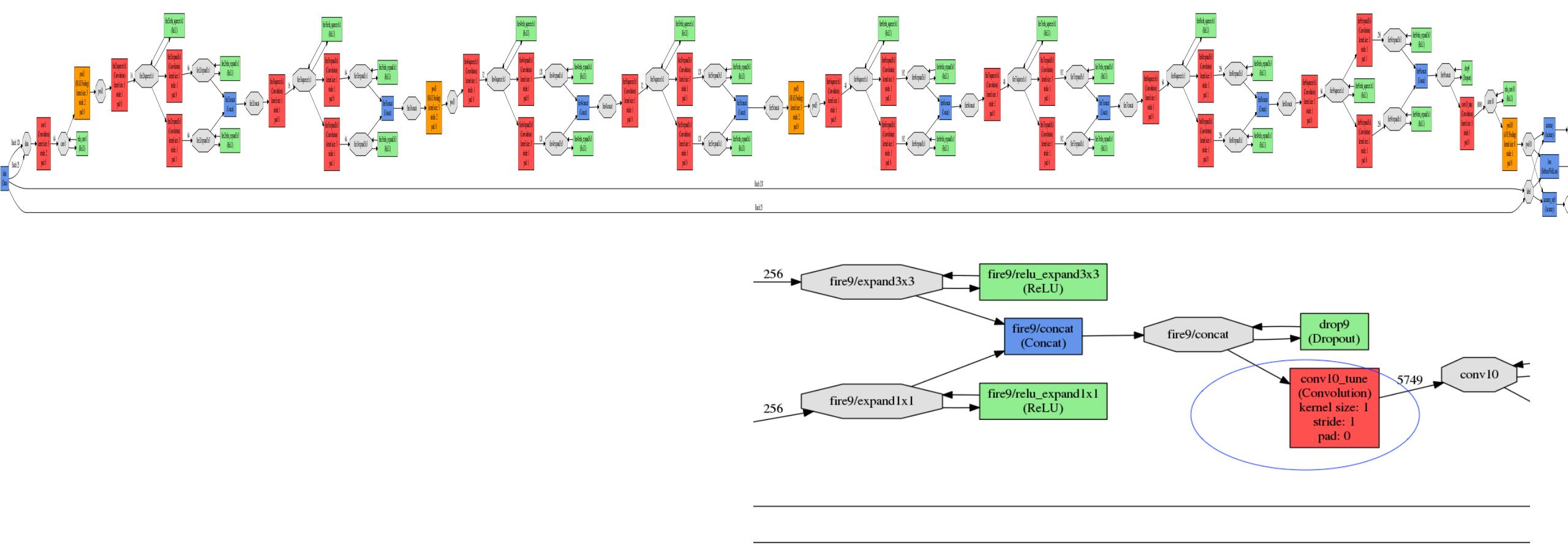
Fine tune network

- AlexNet



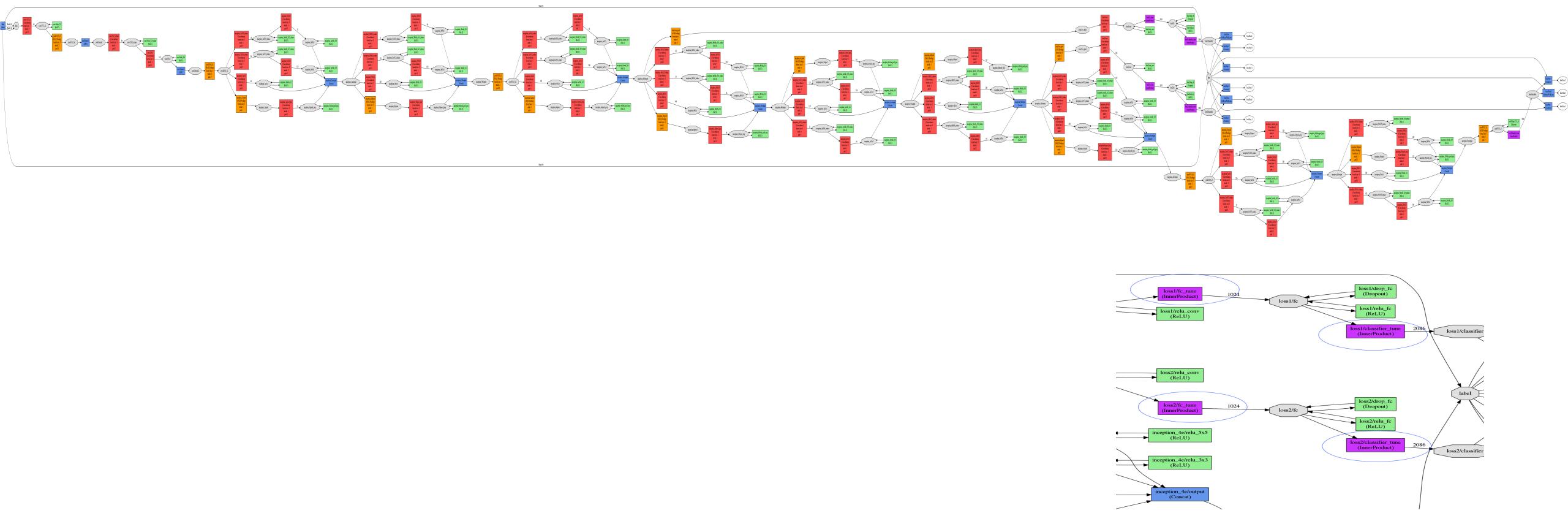
Fine tune network

- SqueezeNet



Fine tune network

- GoogLeNet

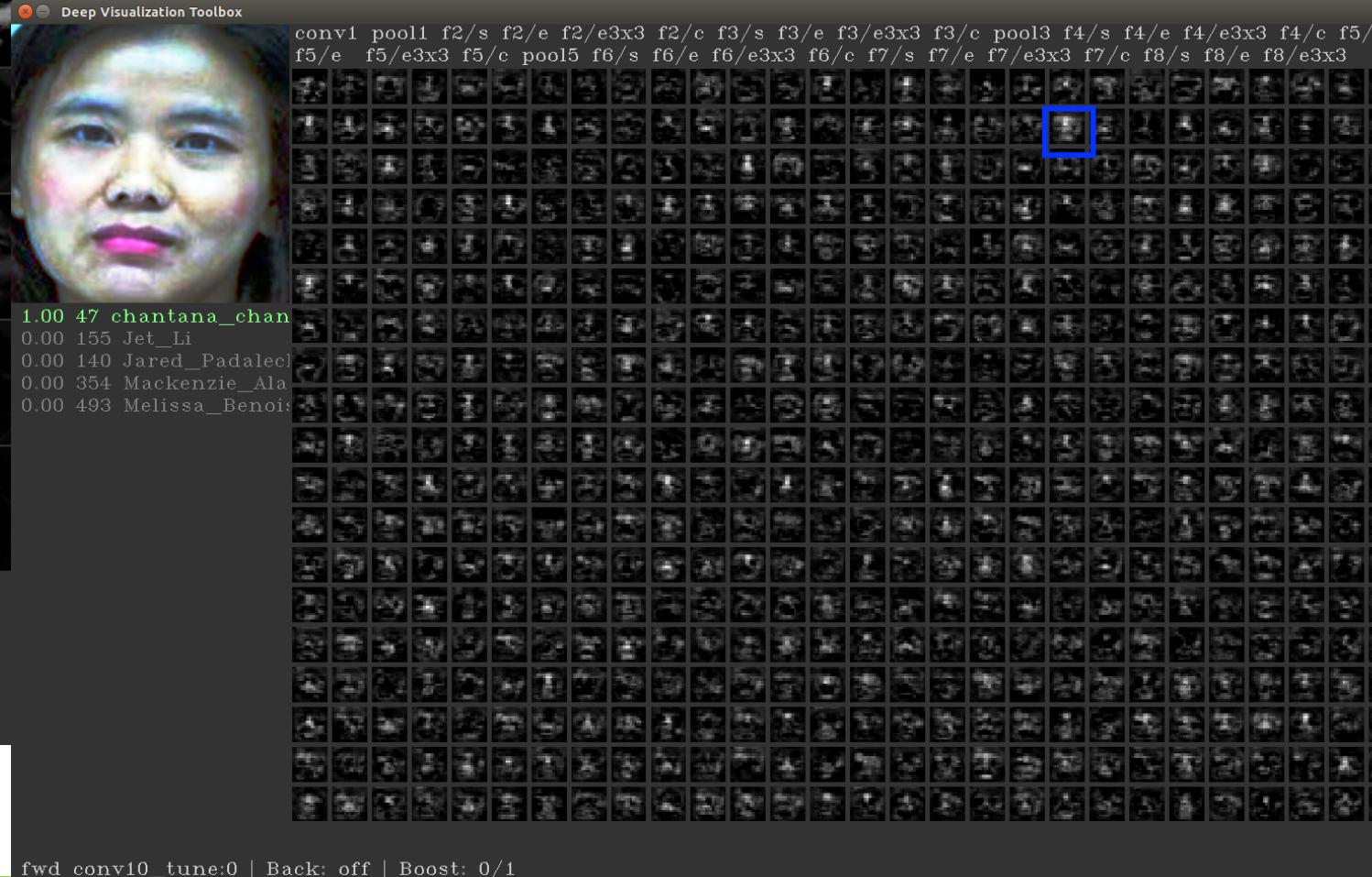


Visualize weights



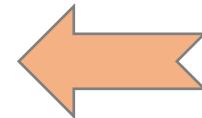
Deep visualization toolbox

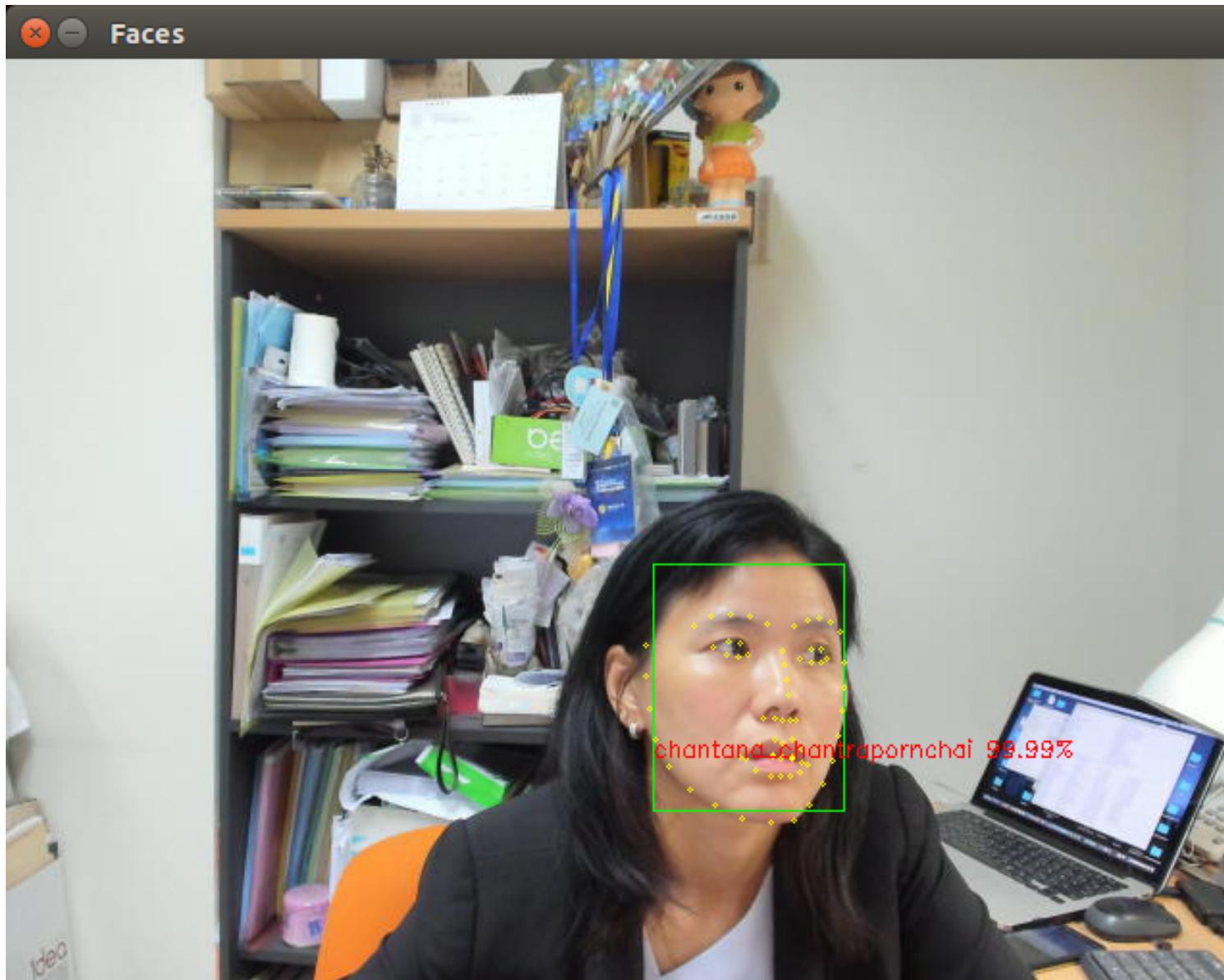
Visualize weights



Comparison to other models

- Openface (CMU) (2016)
<https://cmusatyalab.github.io/openface/>
- https://github.com/ageitgey/face_recognition/blob/master/examples/face_recognition_knn.py
- FaceNet (Triplet loss)
<https://github.com/davidsandberg/facenet/wiki/Triplet-loss-training>
- Faster RCNN with face recognition (2016)
<https://github.com/playerkk/face-py-faster-rcnn>

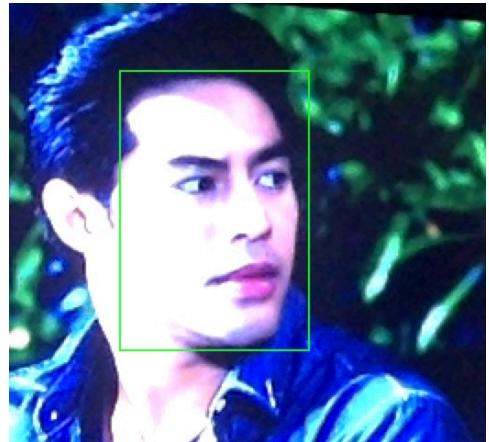




Software Process



Face detection



MTCNN

<https://github.com/DuinoDu/mtcnn>



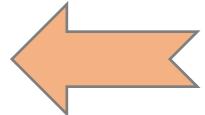
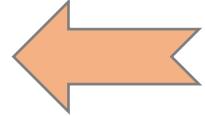
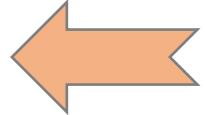
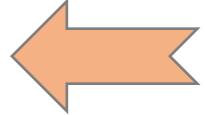
Recog. Model?
Image means

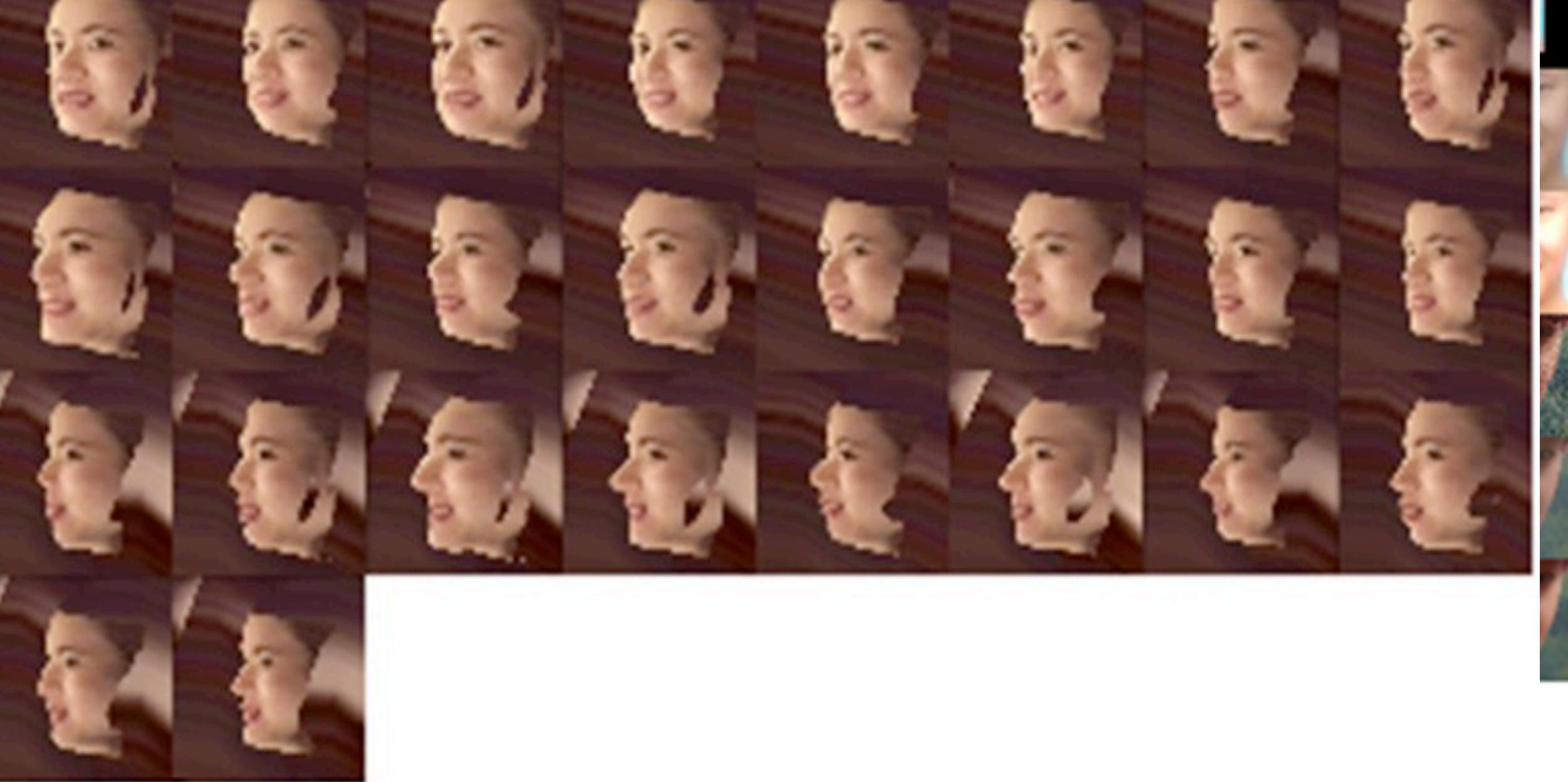


Classes

To query
database

Preprocessing Done

- Data augmentation -- <https://github.com/aleju/imgaug>
http://www.openu.ac.il/home/hassner/projects/augmented_faces/Masietal2016really.pdf 
- Face Alignment (dlib) –Openface 
- Pose alignment -- <http://cvlab.cse.msu.edu/project-pifa.html>
https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Jourabloo_Pose-Invariant_3D_Face_ICCV_2015_paper.pdf 
- Compute image means 
- Image resizing



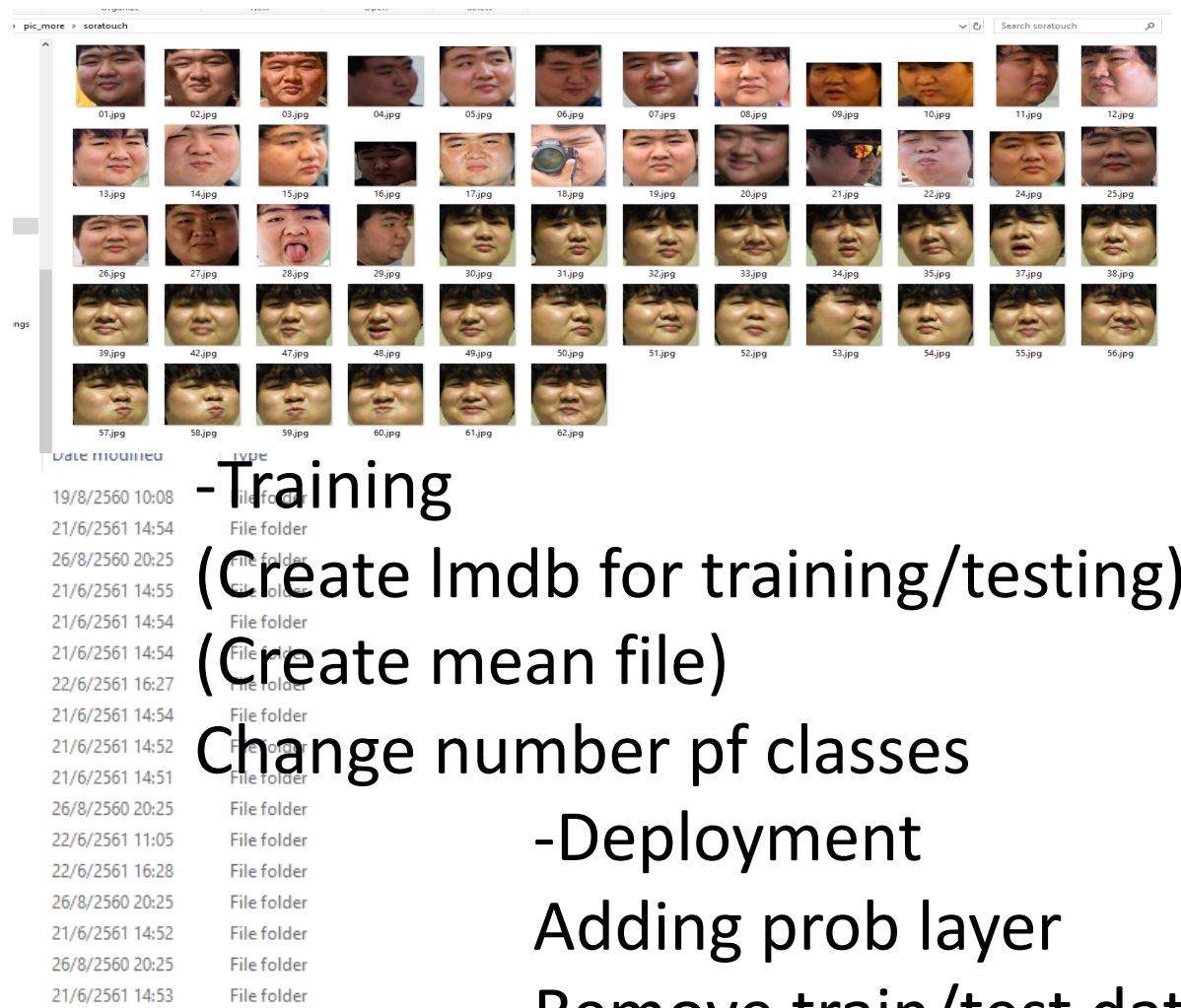
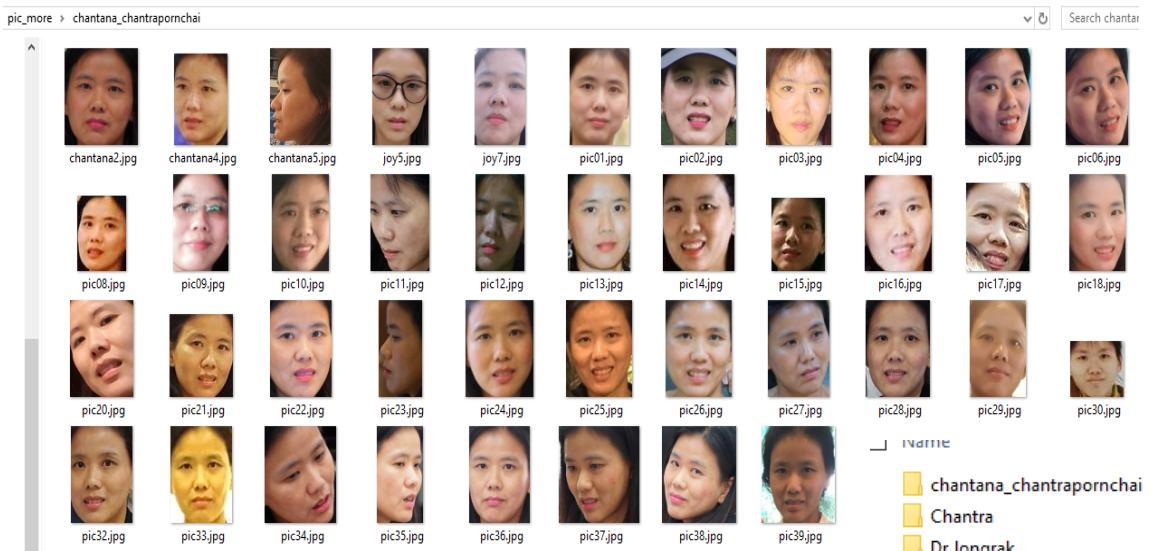
https://github.com/iacopomasi/face_specific_augm

<https://github.com/aleju/imgaug>

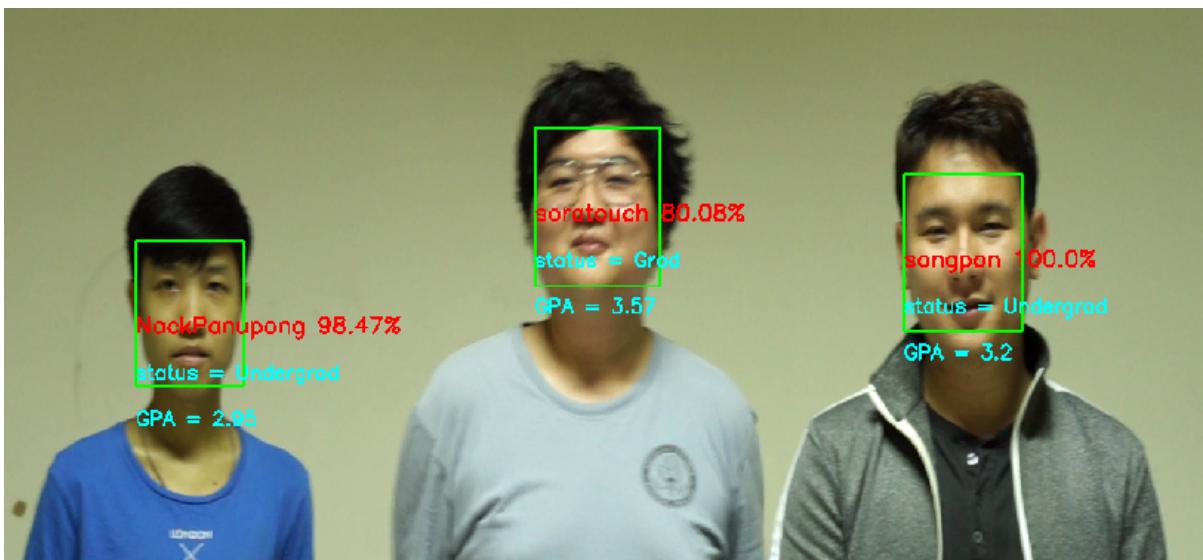
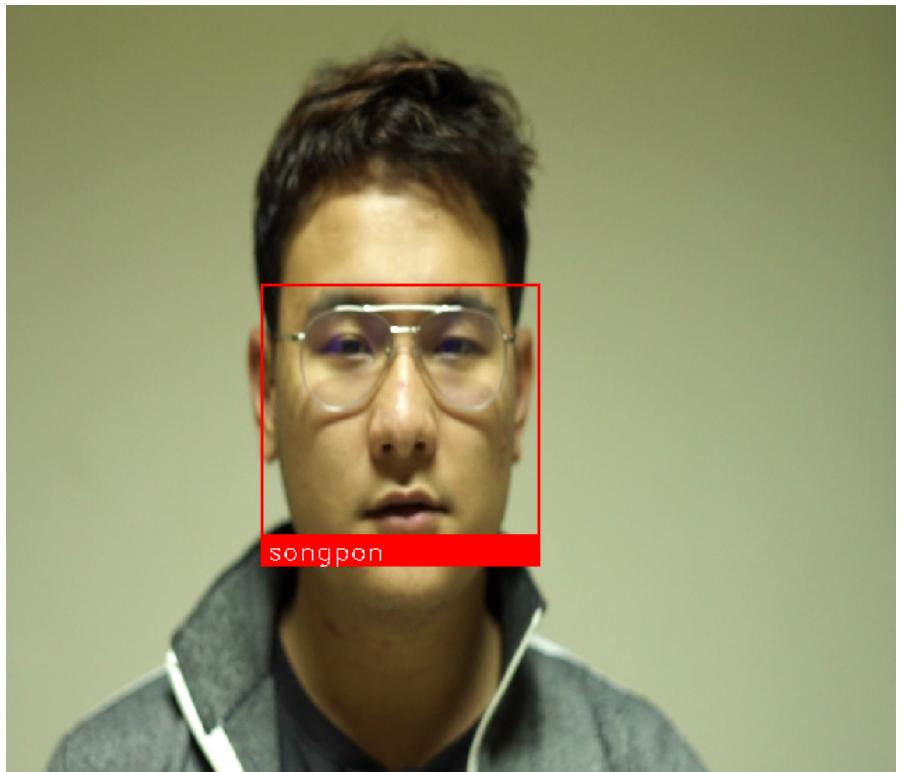
Transfer learning parameters

- Number of classes
- Number of layers
- Layer parameters: filter size, pretrained weight, module, batch size, number of filters, etc.

Adding new classes/data images



-Training
(Create Imdb for training/testing)
(Create mean file)
Change number pf classes
-Deployment
Adding prob layer
Remove train/test data s



Discussions

Method	Accuracy	Size	Simplicity	Speed	Scalability	Sensitivity
Deep learning (MTCNN+SQ)	Medium	Constant	Medium	High	High	Med
Openface	High	Varied	High	Low	Low	High
Dlib+knn	High	Varied	High	Low/ Med	Low	High

- Simplicity in retraining
- Speed of inference
- Scale to large data sets
- Sensitivity to smaller size/ Correctness

In summary

- Explore various face recognition tools including deep learning
- Study preprocessing method
- Study transfer learning approach

Deep learning approach can scale very well with large data sets.

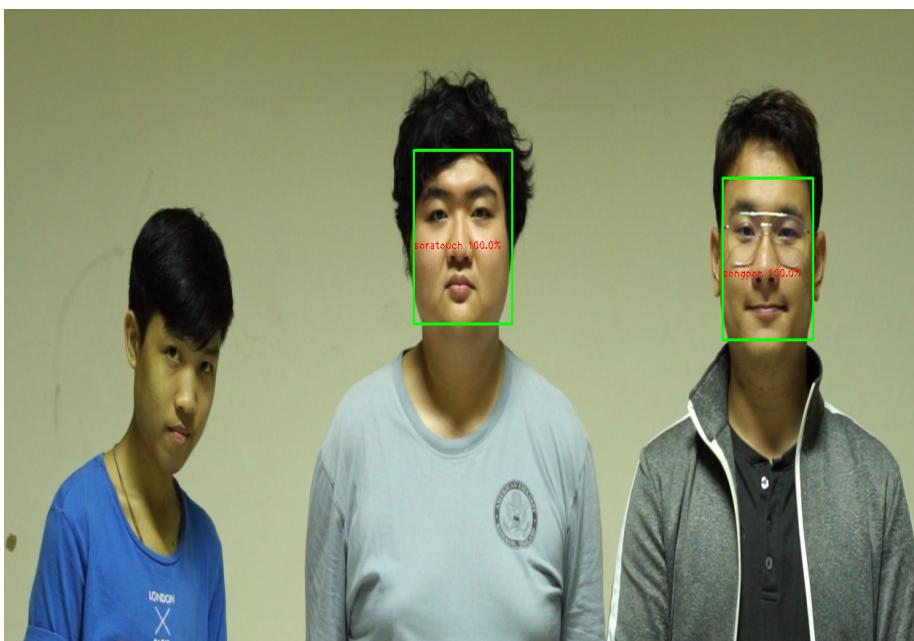
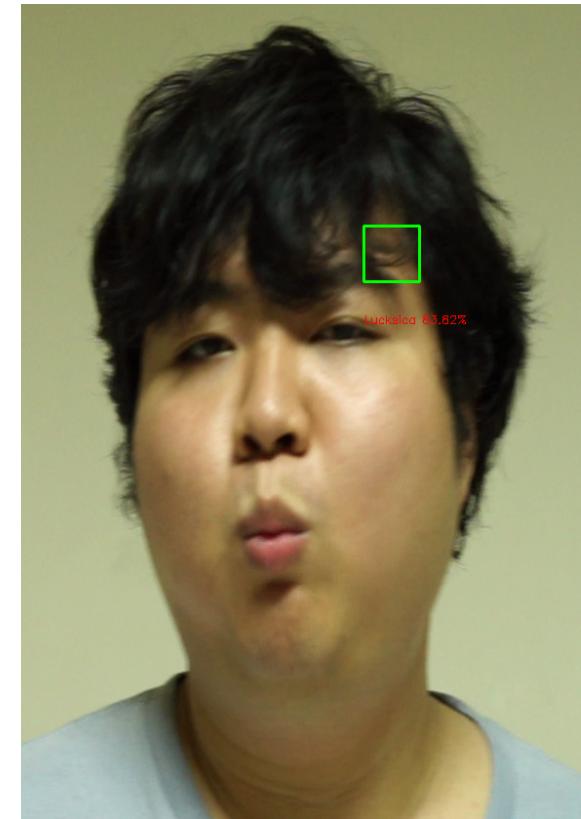
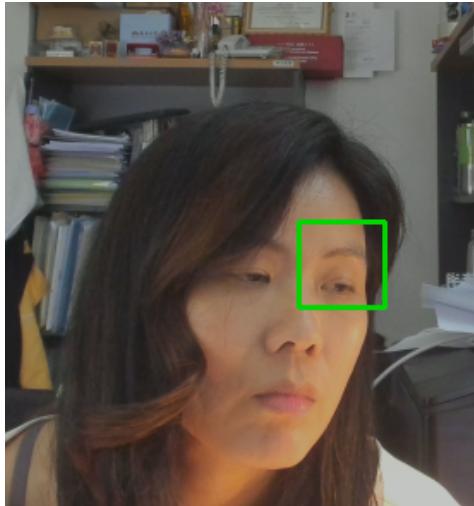
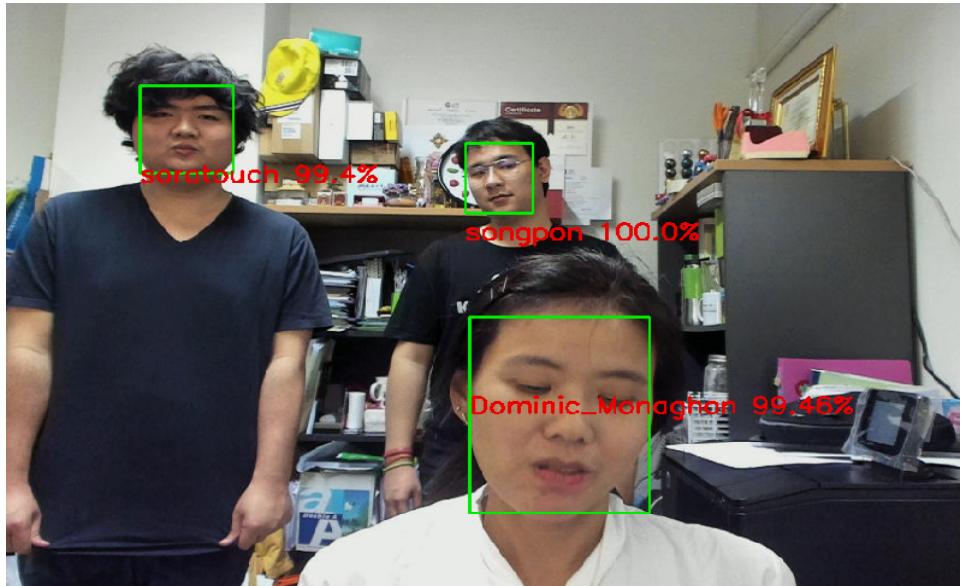
-The more training data/ the more accuracy

OpenFace/dlib is limited to number of training images. High accuracy is obtained with small training set.

Suggestion to work with

- Fixed environment; lights and path
- Fixed pose for seconds, acceptable resolution.
- More training images, more accuracy
- Drawback: unclear image (blur/resolution), image with less landmark for detection,
 - False face detection (size threshold)
- In case with accuracy less than threshold: human judgement
 - In case of video, average predictions over frames

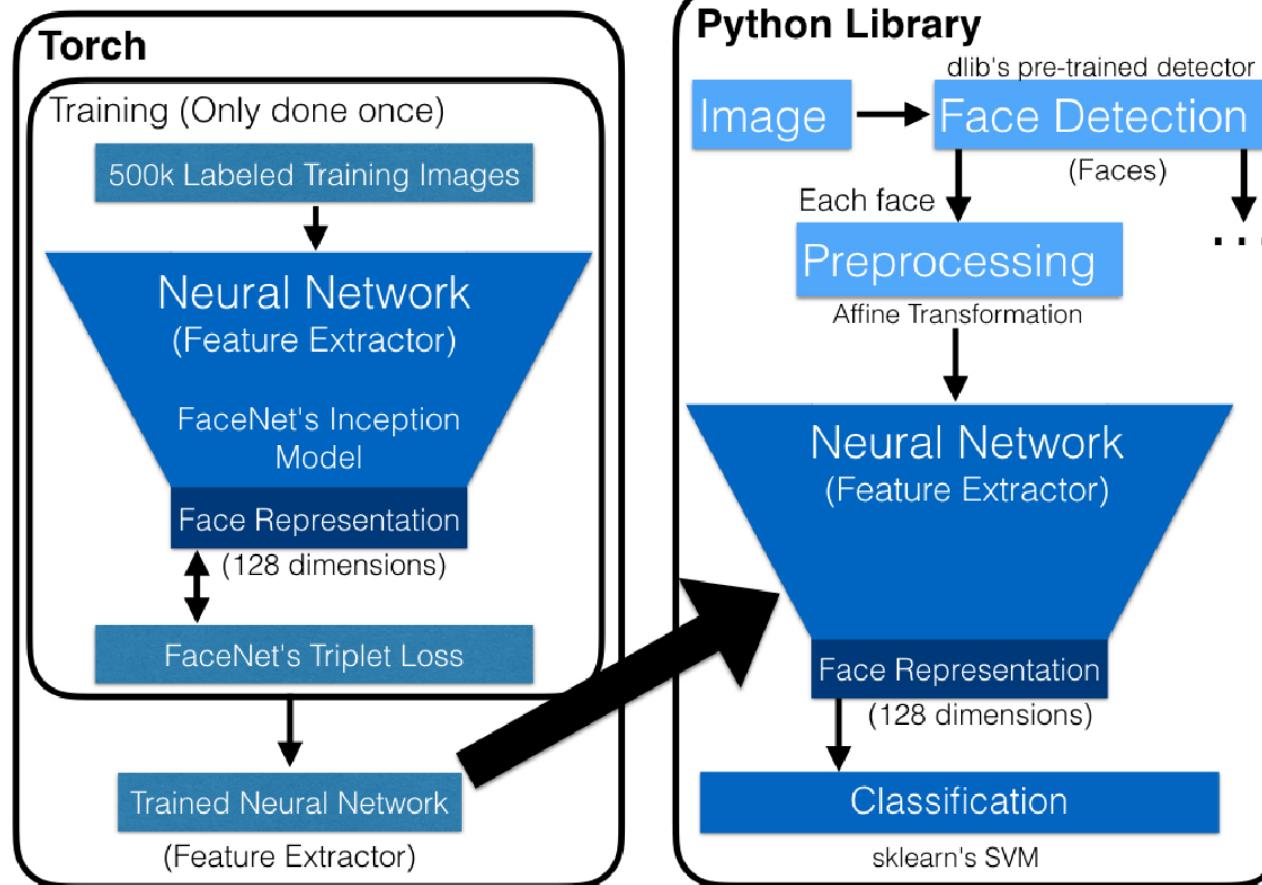
In case of unknown class: open set training approach



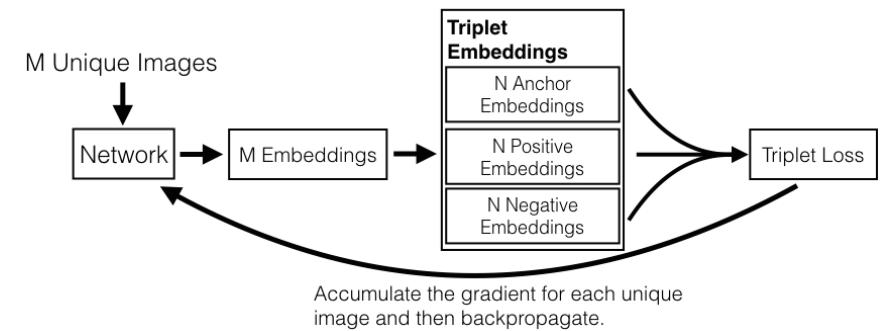
More about tools...

Openface

<https://blog.algorithmia.com/understanding-facial-recognition-openface/>



Use dlib face detection
Use openCV affine transformation (align)
Feature extraction (128 embeddings)
SVM training (clustering)



<https://www.cs.cmu.edu/~satya/docdir/CMU-CS-16-118.pdf>

Triplet Loss (FaceNet)

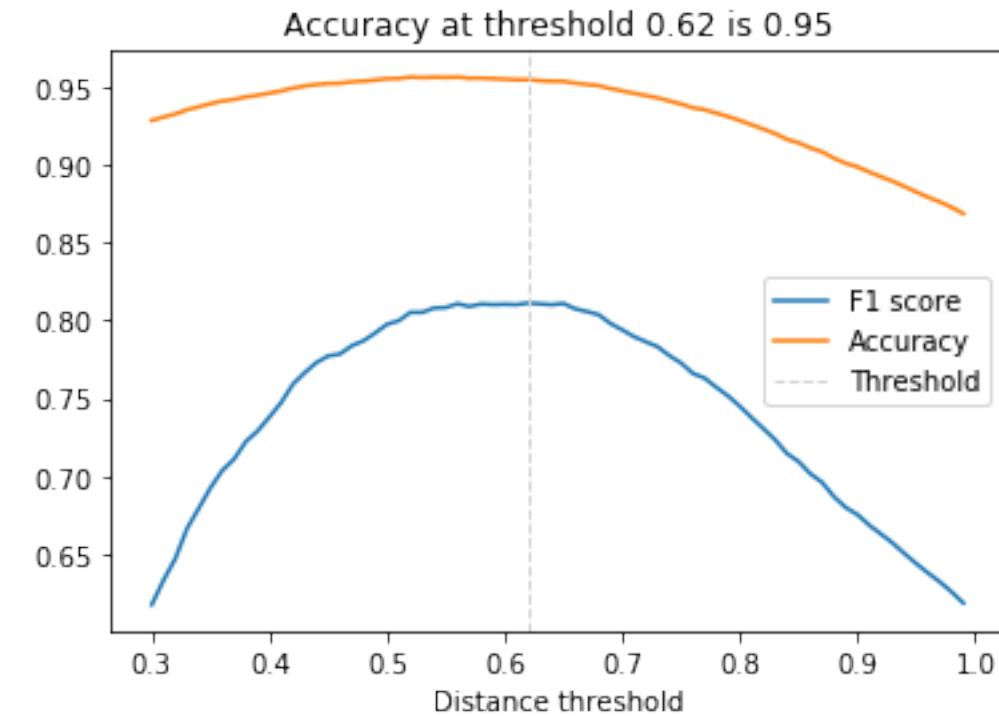
- from anchor, positive and negative embedding vectors
 - Squared L2 distance between all faces of the same identity is small and the distance between a pair of faces from different identities is large.
 - when the distance between an anchor image x_i^a and a positive image x_i^p (same identity) in embedding space is smaller than the distance between that anchor image and a negative image x_i^n (different identity) by at least a margin α .

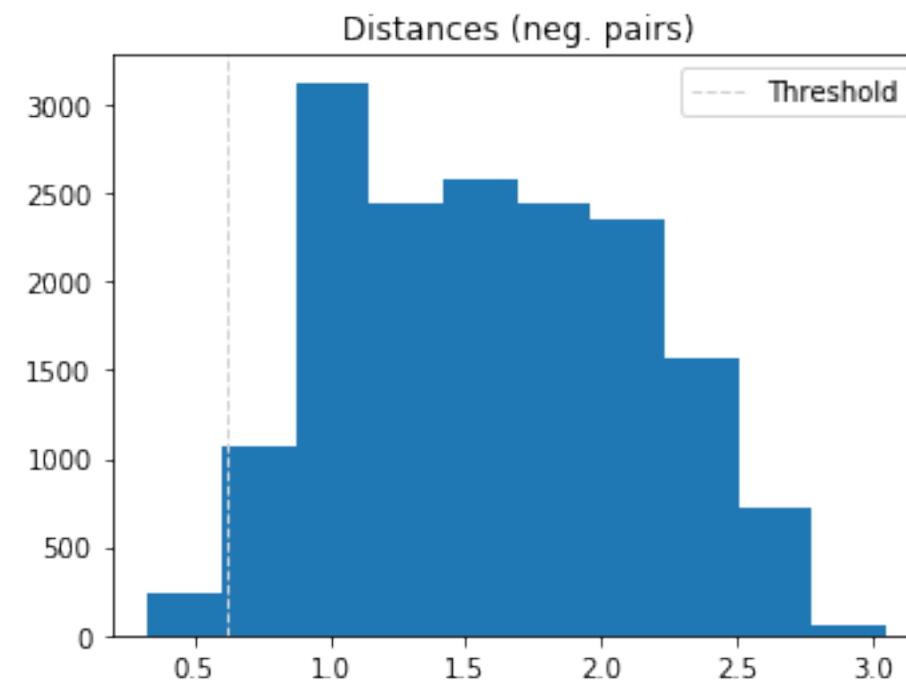
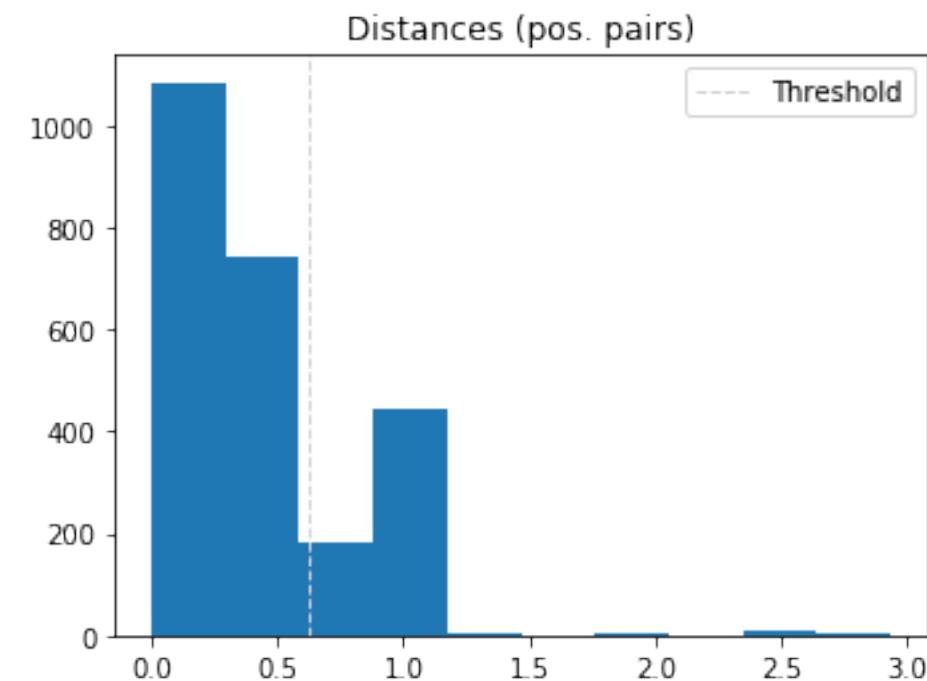
- Minimize loss

$$L = \sum_{i=1}^m [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha]_+$$

Steps

- Align faces
- Find embeddings
- Find distances between pairs for all possible pairs
 - Find list of the same images (identical images)
- Find F1 for identical images with less than given thresholds (Varying threshold)
 - skewed classes (much more negative pairs than positive pairs),
- Find accuracy for identical images with less than given thresholds (Varying threshold)





Distance distribution between positive pairs and negative pairs

- Given an estimate of the distance threshold τ , face recognition is as calculating the distances between an input embedding vector and all embedding vectors in a database.
- The input is assigned the label (i.e. identity) of the database entry with the smallest distance if it is less than τ or label *unknown* otherwise.

<https://github.com/krasserm/face-recognition>

- face-recognition1.ipynb
- github.com/cchantra/face_training

<https://github.com/habrman/FaceRecognition>

- `python3 main.py ./models/models.pb ./picture_train/`

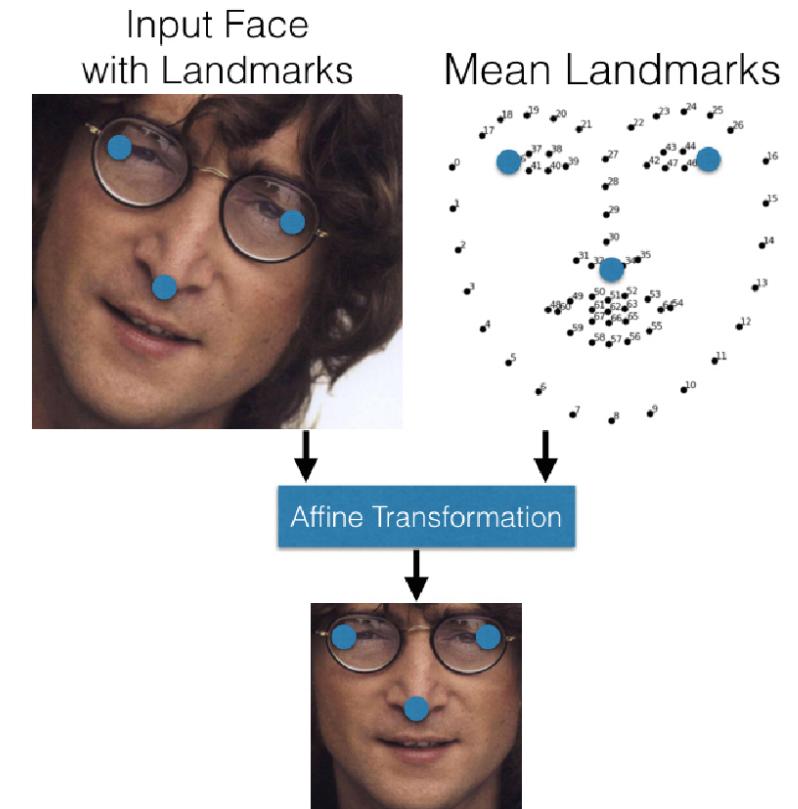
<https://cmusatyalab.github.io/openface/>

demo

OpenCV face recognition helper libraries

- Dlib face detection (HOG Histogram of Oriented Gradient, Support Vector Machine or OpenCV's Haar cascade classifier)
- Find face landmark

Affine transform



Other demos

- <https://github.com/chen0040/keras-face>
 - /Users/admin/face_classification
- https://github.com/oarriaga/face_classification
 - /Users/admin/face_classification