Deep Learning
The AI Renaissance

Mohammad Tariqul Islam

Assistant Professor

Computer Science Department

Southern Connecticut State University

Outline

Artificial Intelligence (AI) & Machine Learning (ML)

The AI Winter

The Rise of Deep Learning

A Brief Overview of Neural Networks

Convolutional Neural Network (CNN)

Case Study 1:
Diagnosis of Diabetes using Retinal Images

Case Study 2:
Galaxy Cluster Classification

Case Study 3:
Photo-realistic Facial Image Generation using Generative Adversarial Networks

Q & A

Artificial Intelligence

What is AI?

"the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment."

Artificial Intelligence: A Modern Approach", Stuart Russell and Peter Norvig (1994)

An agent that can improve its performance through experience

Some Applications of AI

Computer Vision

- Self-driving Cars
- Face Detection & Recognition
- Security & Surveillance

Natural Language Processing

- Speech Recognition (e.g. Alexa, Siri)
- Natural Language Translation (e.g. Google Translate)

Finance

- Stock Market Prediction
- Lending Decisions
- Investment Risk Assessment

Marketing

- Advertisements based on user activity
- Promotions & discounts maximizing sales
- Purchase behavior understanding

Medical Science

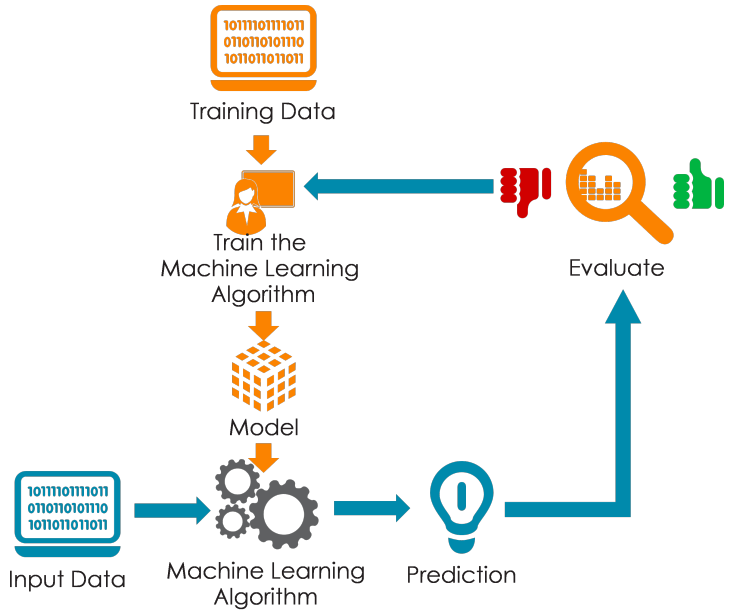
- Medical image segmentation (X-Ray, MRI, Ultrasound, etc.)
- Detection & prognosis of diseases (Pneumonia, cancers, genetic diseases)

Agriculture

- Plant disease identification
- Efficient use of pesticide

Artificial Intelligence (AI) & Machine Learning (ML)

- ML is a branch of AI
- Study, development, and application of algorithms that learn patterns from data
- A more common term for AI nowadays



https://medium.com/@tekaround/train-validation-test-set-in-machine-learning-how-to-understand-6cdd98d4a764

The AI Winter

- AI-based methods didn’t live up to the expectations
  - Computationally expensive
- Lasted from around 1973 to 2012
- Reduced funding and interest towards AI
- More focused on applying ML algorithms

The Rise of Deep Learning

- Traditional Machine Learning is a ~60 years old technique
  - SVM, DT, NN, etc.

- Deep Learning is only 9 !!
  - AlexNet: Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton (2012)
    - ImageNet Classification with Deep Convolutional Neural Networks
  - CNN: Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner (1998)
    - Gradient-based Learning Applied to Document Recognition

So… What is Deep Learning ?

a relatively modern machine learning approach that uses multi-layered (deep) artificial neural networks to solve problems that involve some form of pattern recognition.

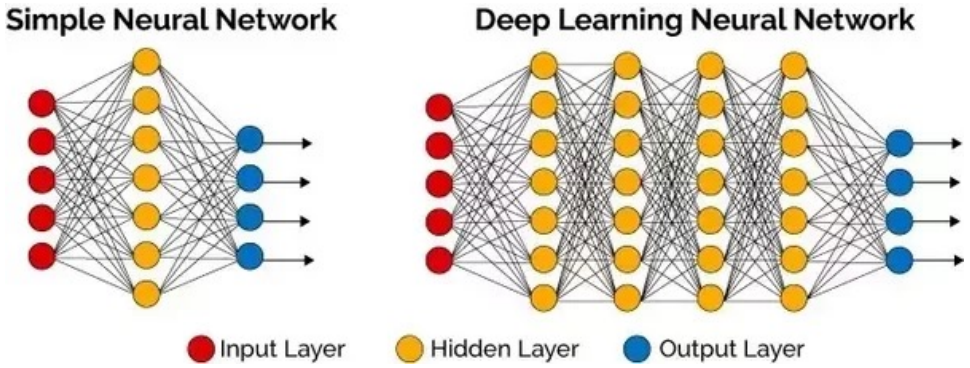


Image credit: https://www.quora.com/What-is-the-difference-between-Neural-Networks-and-Deep-Learning

Why now?

- Availability of Powerful GPUs
  - Driven by the computer and video gaming community
    - Thanks to the gamers!
  - Computer graphics requires matrix multiplications for:
    - geometric transformations
    - Shader computations
    - Rendering
  - NNs are all about matrix multiplications … lots of them.

- Better Optimization Techniques
  - Activation functions
  - Regularization techniques
  - Faster convergence algorithms

- More data!

Scalability of Traditional vs. Deep Methods

- More data doesn't help traditional learning
- Access to more data
- Sources:
  - Internet:
    - Social Media
    - Search Engines
    - Marketplaces
  - Research Labs & Studies
  - Governments
  - Institutions and Organizations

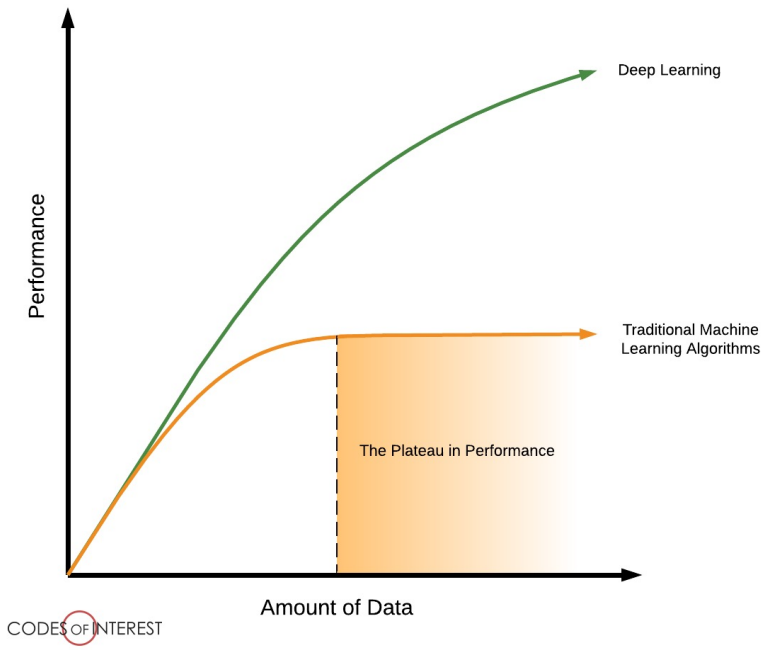
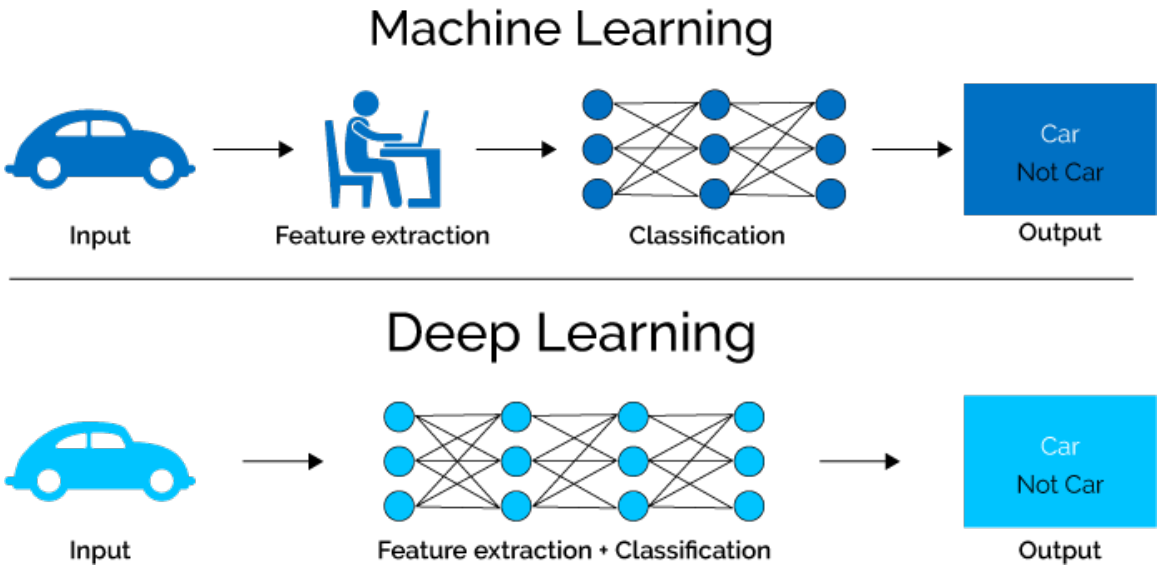


Image Credit: https://www.codesofinterest.com/2017/04/how-deep-should-deep-learning-be.html

Deep Learning vs. Traditional Learning



https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png

Deep Learning (DL) vs. Machine Learning (ML) vs. Artificial Intelligence (AI)

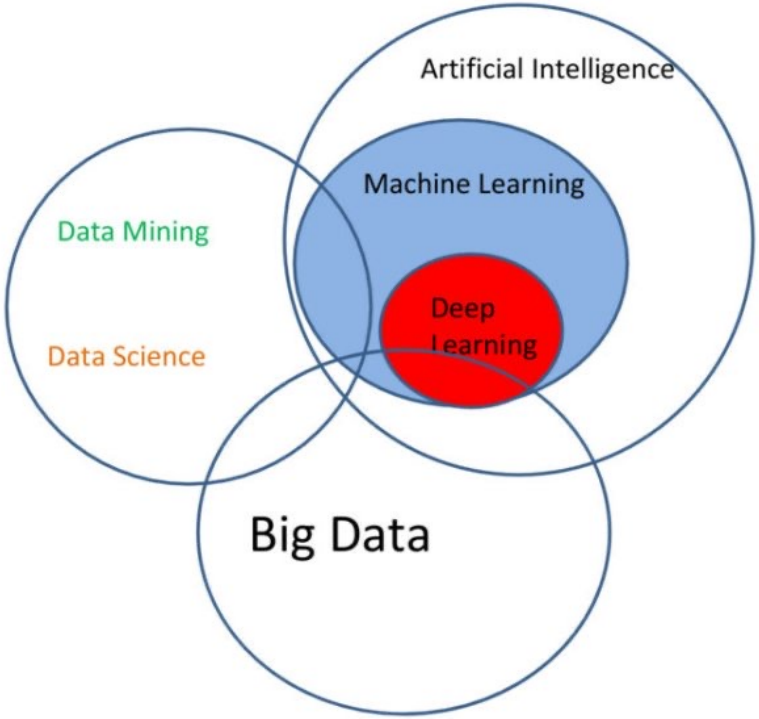


Image Credit: https://whatsthebigdata.files.wordpress.com/2016/10/ai_data-science-diagram.jpg?w=640

Neural Networks

- A network of logistic regression units (neurons)
  - Arranged in layers, each layer containing multiple neurons
  - Outputs from neurons connected to inputs of the next layer's neurons
  - Data is passed as layer zero
- Hierarchical representation of knowledge
  - Different layers encode knowledge/understanding in increasing order of complexity
- Can be stacked to learn more complex patterns !!
  - Unique in terms of extendibility

Logistic Regression as a Neuron

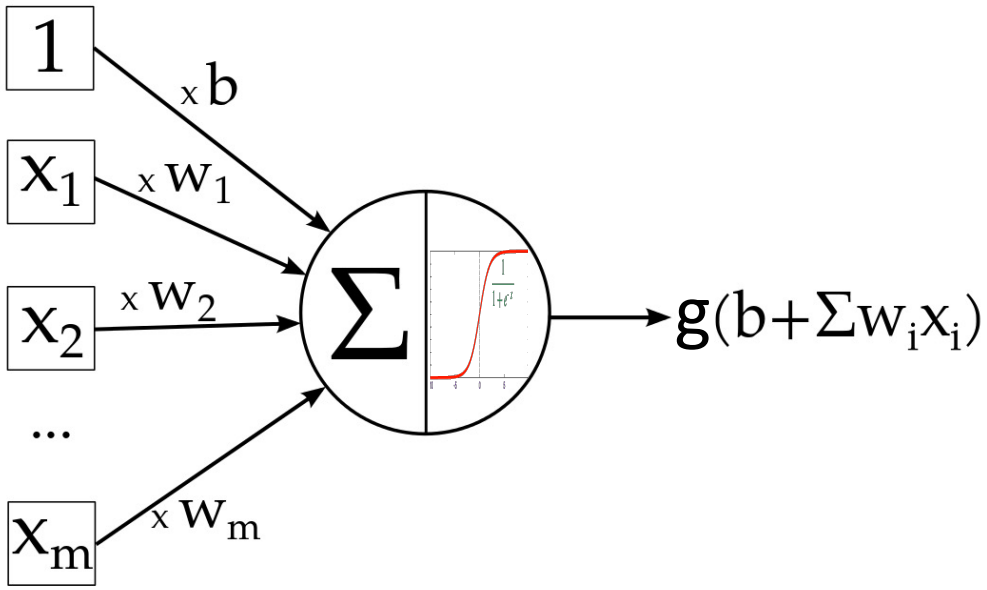


Image Credit: http://inspirehep.net/record/1300728/plots

Neural Network Representation

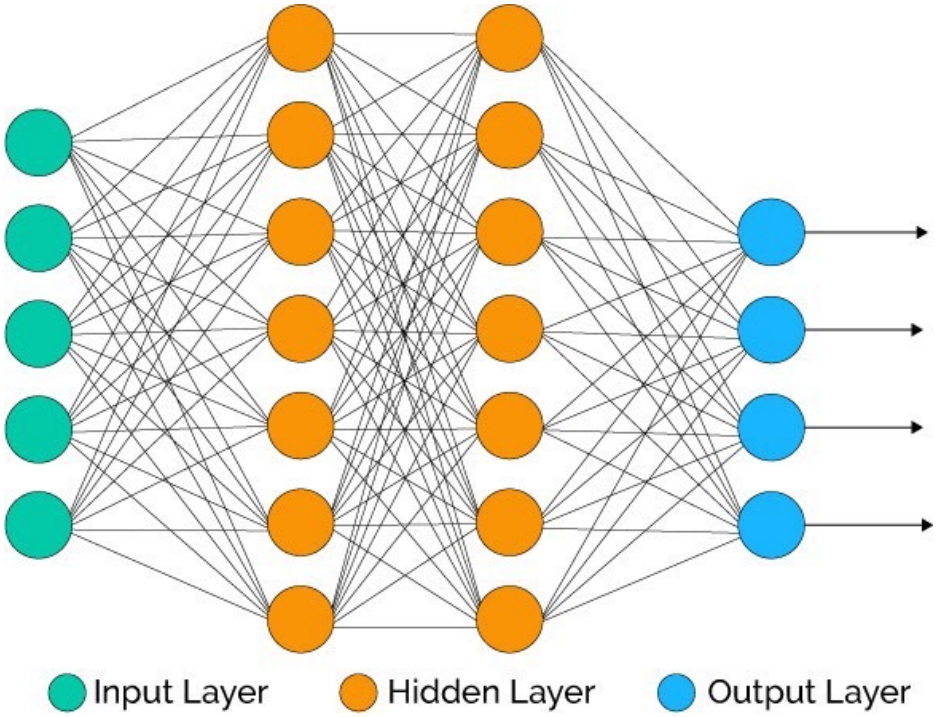
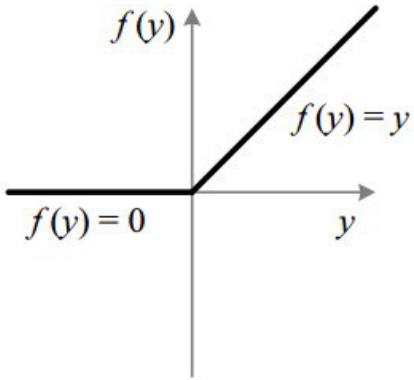


Image Credit: https://towardsdatascience.com/machine-learning-fundamentals-ii-neural-networks-f1e7b2cb3eef

Activation Functions (Non-linearity)

ReLU (Rectified Linear Unit)

$f(y) = \max(0, y)$



Leaky ReLU

$f(y) = y$ if $y>0$

ay otherwise

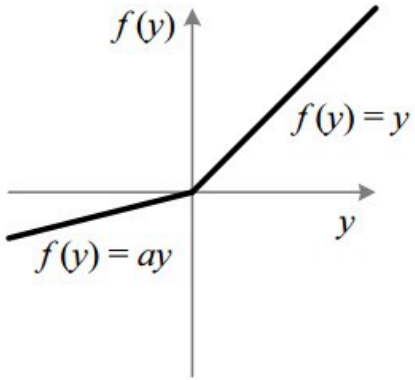


Image URL: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

Training a Neural Network

- Training: Iteratively optimize the weights (and other parameters)
  - Start with random weights, update towards the best set
- Backpropagation: Gradient propagation through the network
  - Successive application of partial differentiation and chain rule
- Gradient Descent to update the parameters:

$$\text{Repeat until convergence } \{$$

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$$\}$$

Image Credit: https://hackernoon.com/gradient-descent-aynk-7cbe95a778da

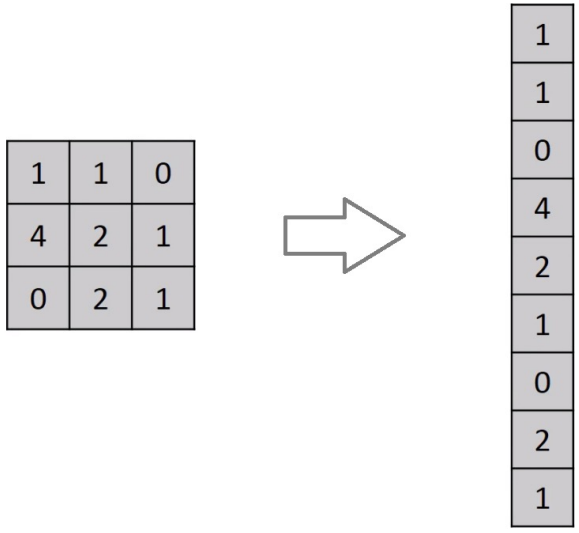
Convolutional Neural Networks: Motivation

Fully-Connected Networks are not suitable for all kinds of data

- Good for low-dimensional, vector-based data.

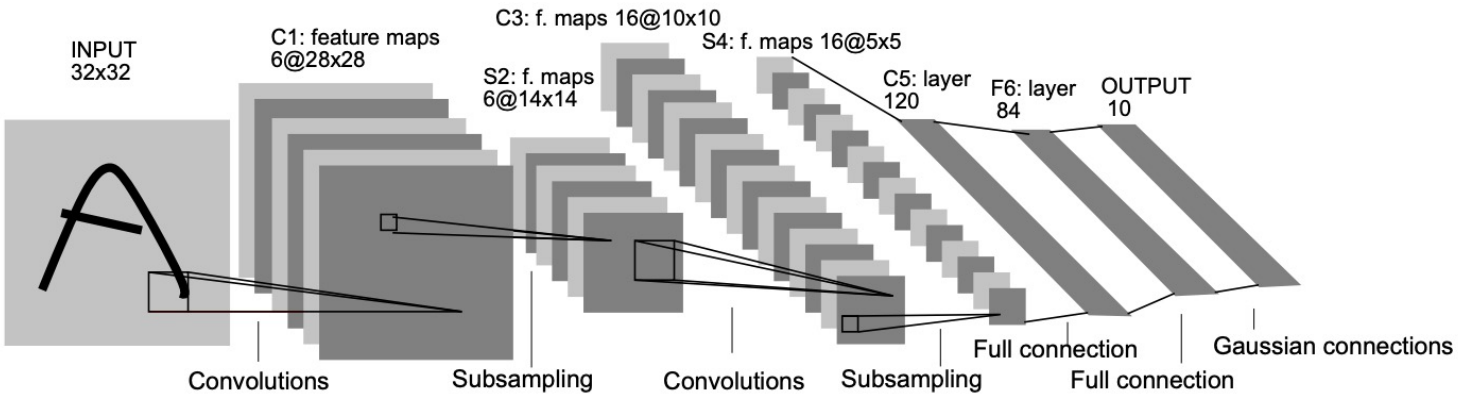
Drawbacks of MLP for Image Data Analysis:

- Number of connections (weights) from the input layer to the first layer increases drastically
- Number of trainable parameters in the order of hundreds of millions
  - Causes overfitting



Slide Credit : **Ming Li** (https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture5CNN.ppt)

Convolutional Neural Networks



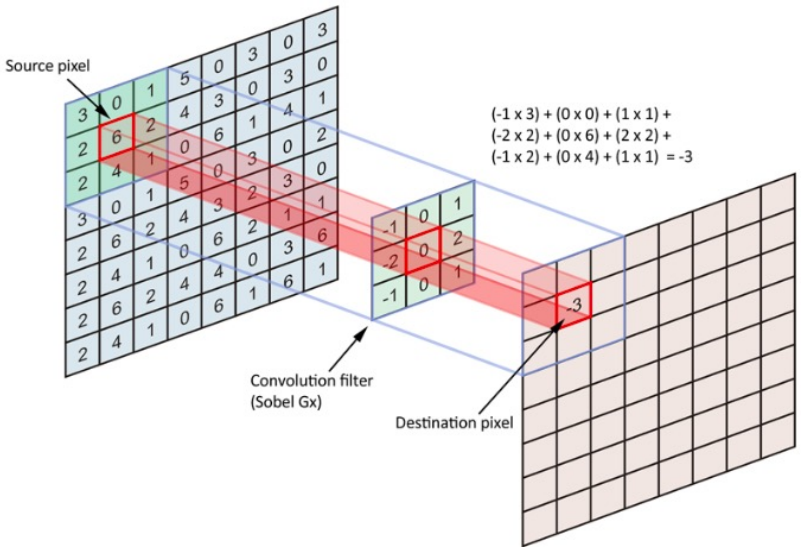
Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Intelligent Signal Processing, 306-351, IEEE Press, 2001

Convolutional Neural Networks

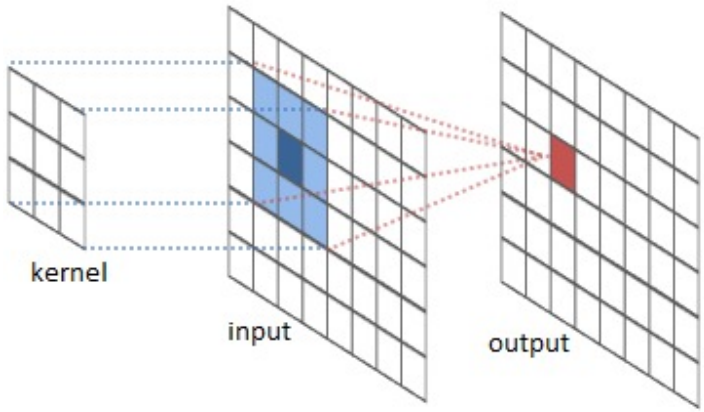
- Components of a Simple Convolutional Neural Network
  - Convolution Layer
  - Max Pooling Layer
  - Fully-connected Layer

- Activation Functions (Non-linearity):
  - ReLU (Rectified Linear Unit)
  - Leaky ReLU
  - Sigmoid
  - Softmax

Convolution



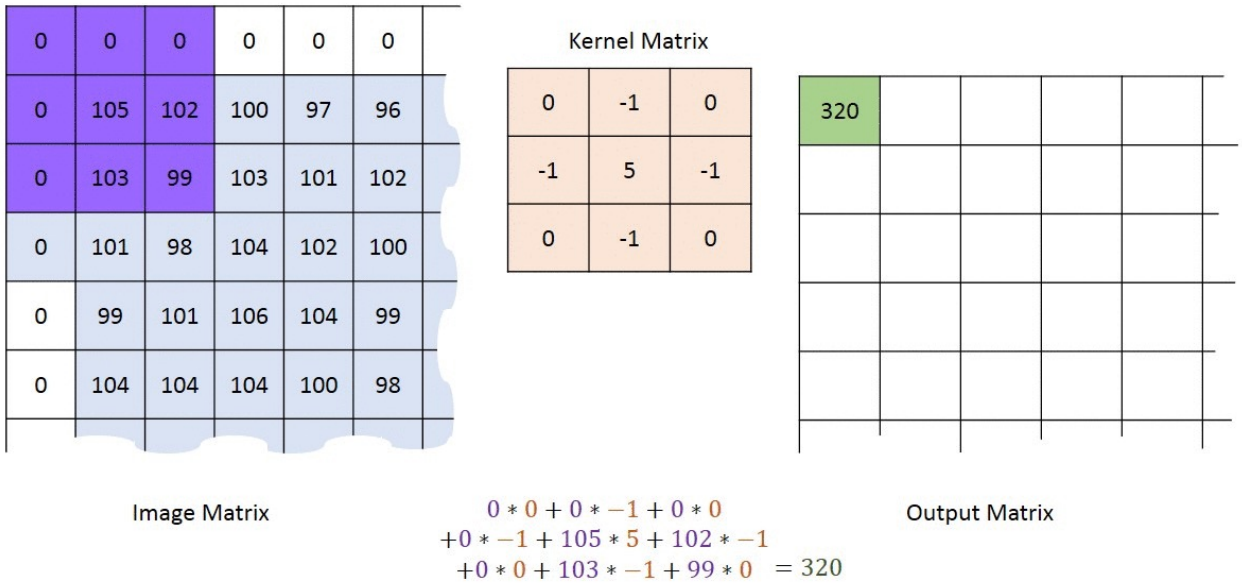
Convolution Operation



Scope and target of operation

Image Credit: https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050, http://colah.github.io/posts/2014-07-Understanding-Convolutions/

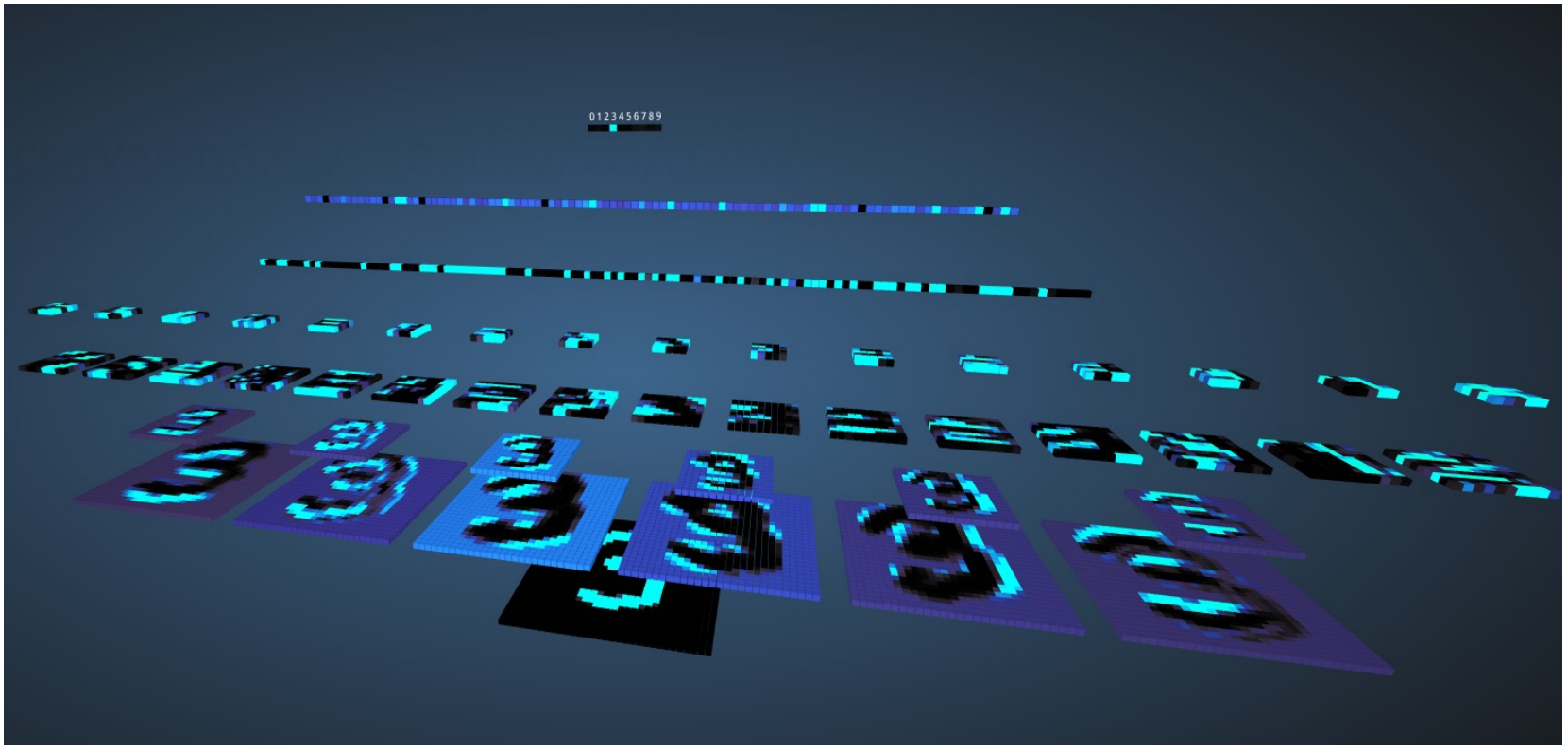
Convolution in Action



Convolution with horizontal and vertical strides = 1

http://machinelearninguru.com/computer_vision/basics/convolution/convolution_layer.html

Convolution Response in Different Layers



http://scs.ryerson.ca/~aharley/vis/conv/

3D Convolution for Multi-Channel Images

- Works on volume instead of surface
- Looks at multiple channels (feature maps) of the input
- Produces a single 2D response map from the input

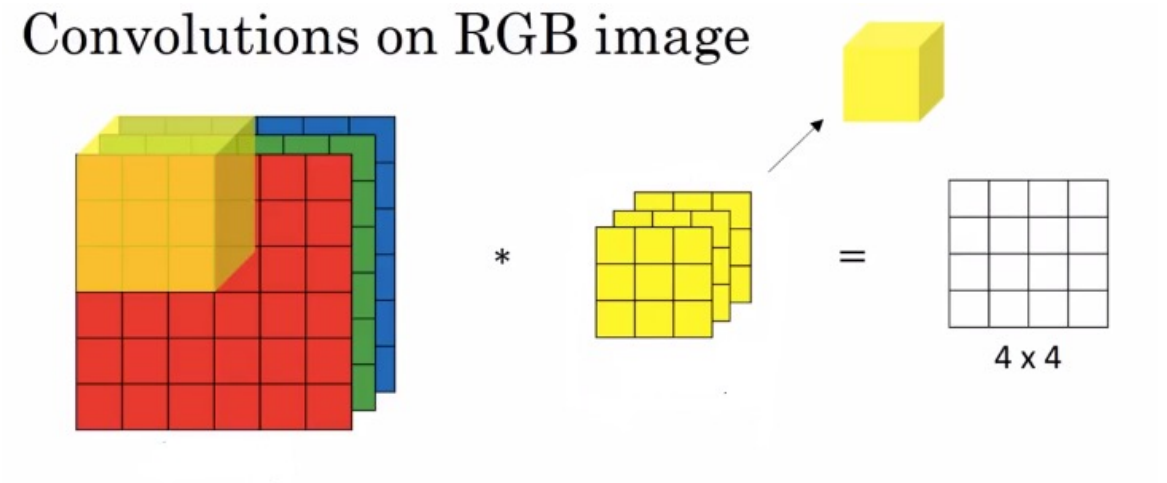


Image Credit: https://stackoverflow.com/questions/47982594/how-a-convolutional-neural-net-handles-channels

Max & Average Pooling

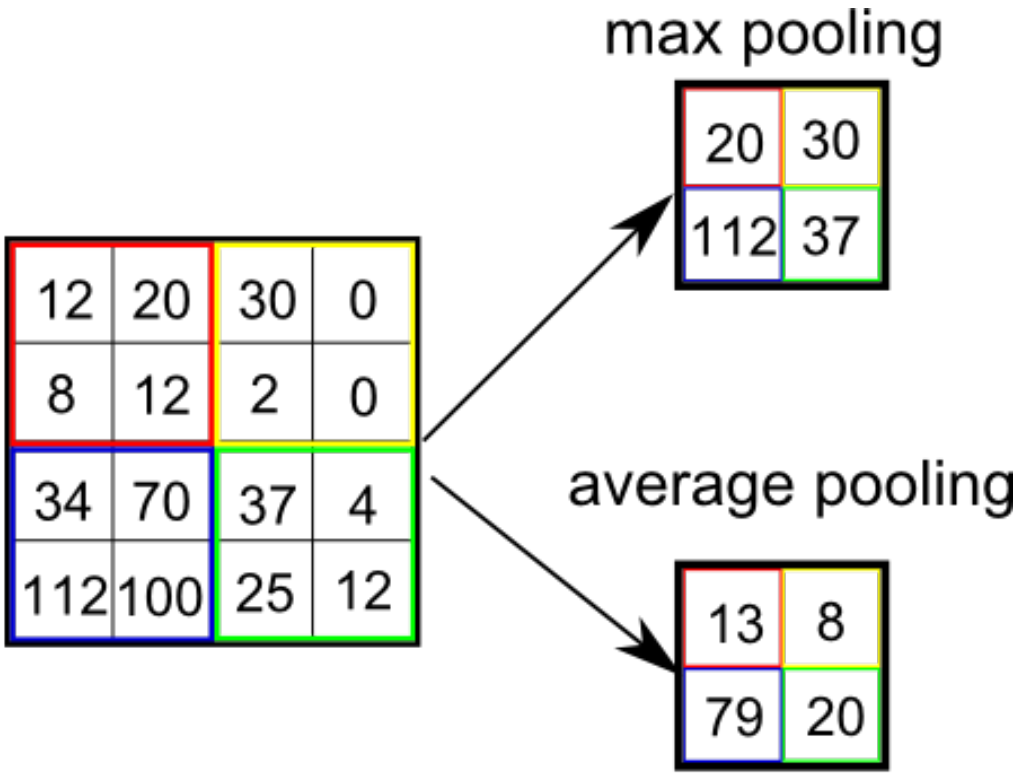


Image URL: https://medium.com/data-science-group-iitr/building-a-convolutional-neural-network-in-python-with-tensorflow-d251c3ca8117

A Convolutional Neural Network

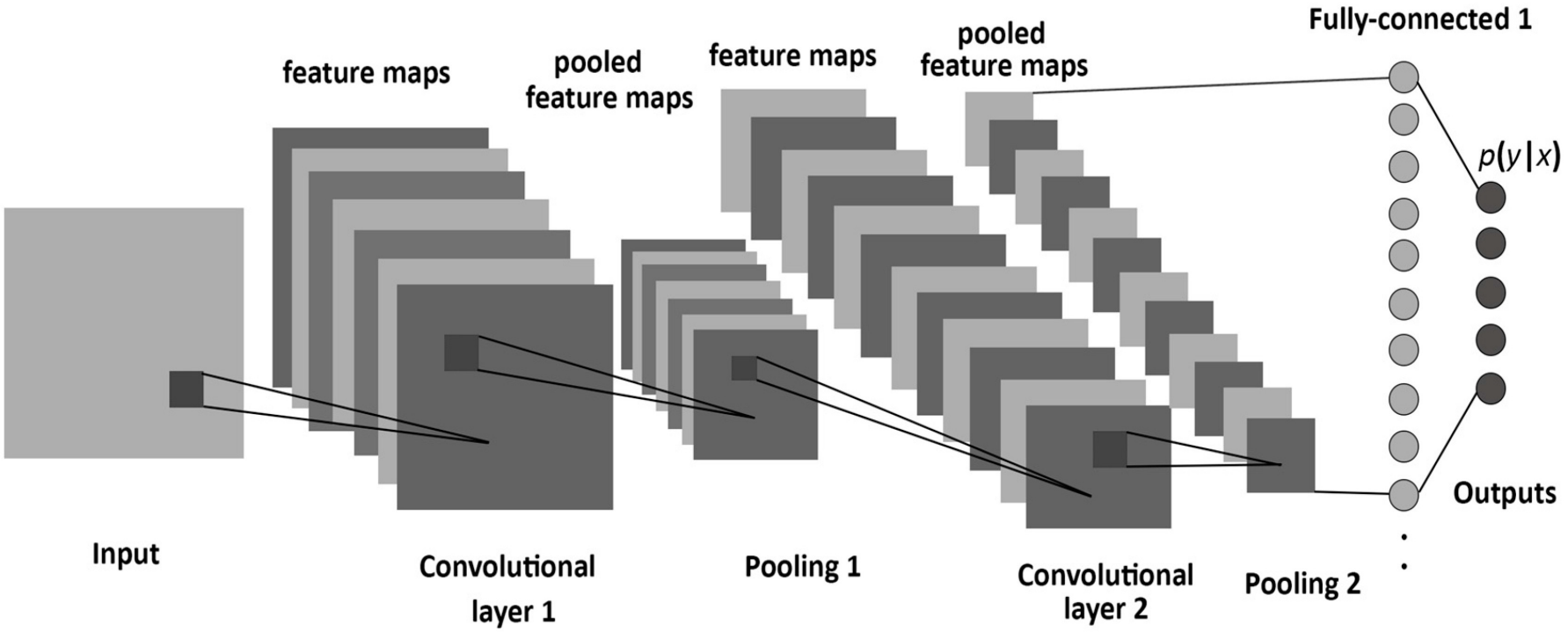


Image credit: Albelwi, S.; Mahmood, A. A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy* **2017**, *19*, 242

Some notable CNN Architectures

- AlexNet
- VGG-16
- ResNet
- DenseNet
- GoogleNet (aka Inception)
- EfficientNet

Image Classification

- Given an image, classify it into a category that represents the most significant object/theme in the image

Cat

Dog

Train

Object Detection

Image => multiple objects + their bounding boxes

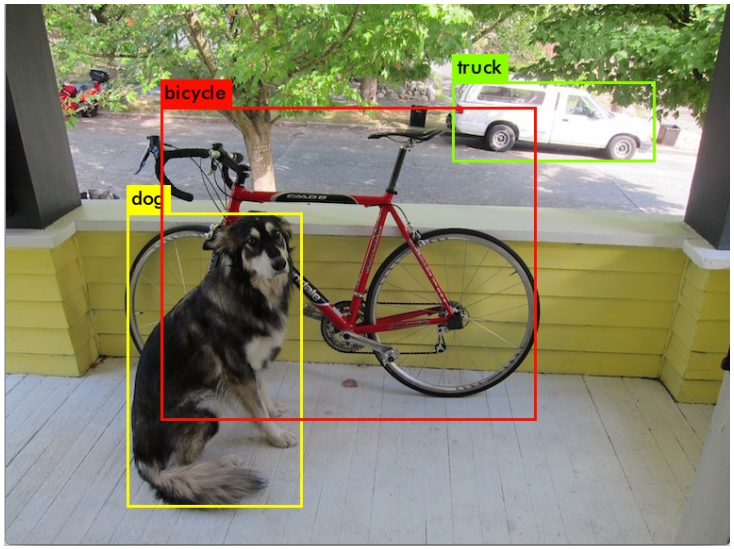
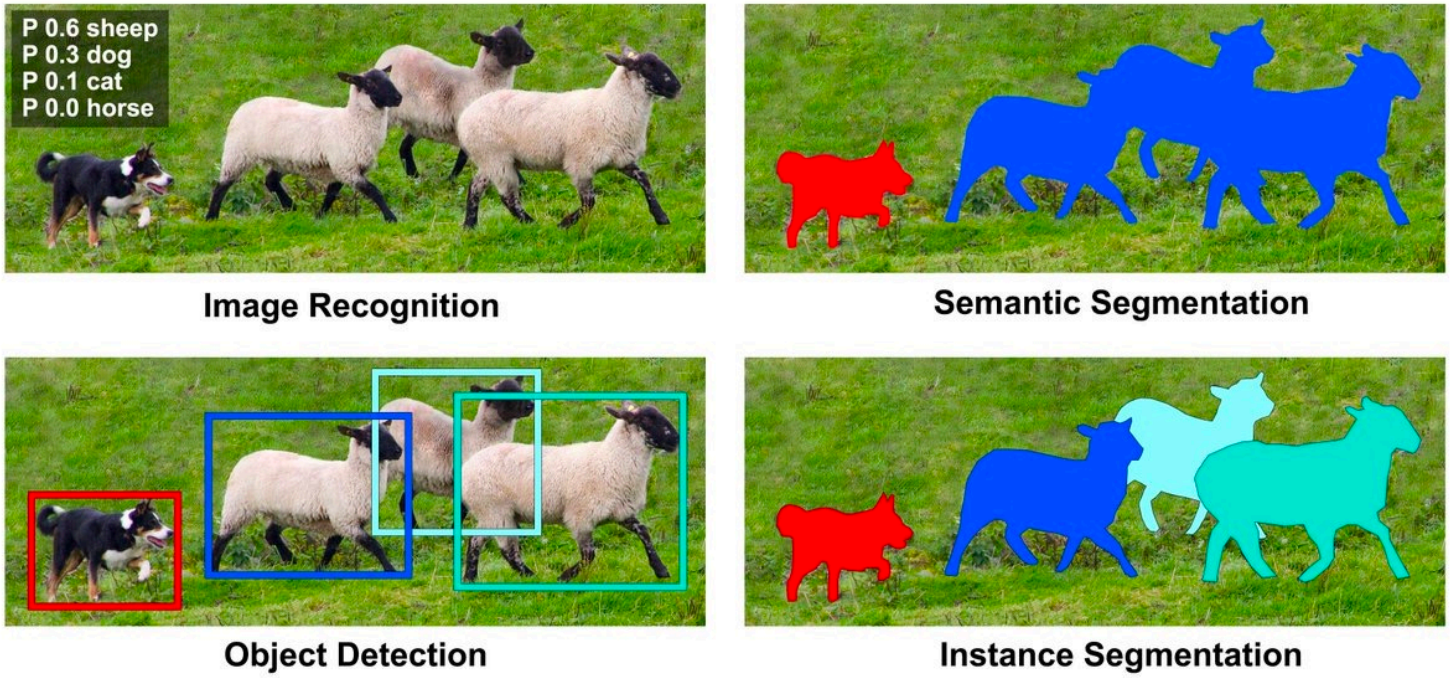


Image Credit: https://pjreddie.com/darknet/yolo/

Image Segmentation



https://ai-pool.com/d/could-you-explain-me-how-instance-segmentation-works

Neural Style Transfer

https://github.com/jcjohnson/neural-style

Deep Learning is More Than Just CNNs …

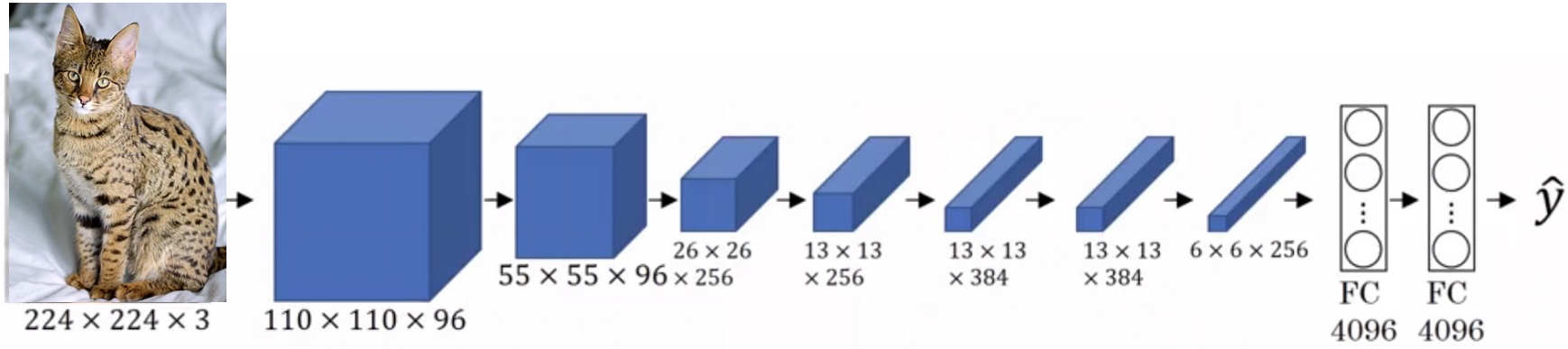
- Recurrent Neural Networks (RNN)
  - LSTM
  - GRU
  - Attention-based models

- Transformers and its variants
  - BERT (Bidirectional Encoder Representations from Transformer)
  - RoBERTa (A Robustly Optimized BERT Pretraining Approach)
  - GPT (Generative Pre-trained Transformer) 2/3

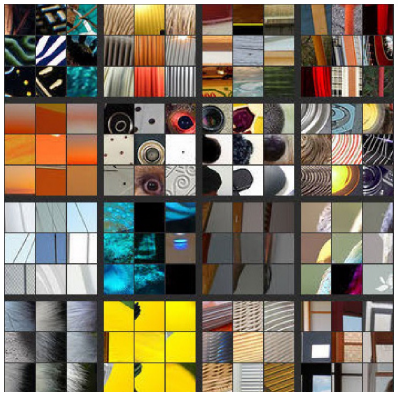
Deep Learning is More Than Just CNNs …

- Generative Adversarial Networks (GANs)
- Capsule Networks
- Variational Autoencoder
- Graph Convolutional Networks
- Deep Belief Network
- Deep Botzman Machine

Content Understanding



Top - 9 image patches that maximize first layer's particular neuron (9)

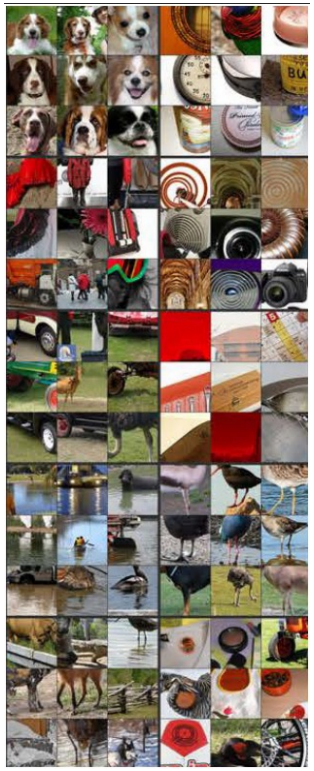


Top - 9 image patches that maximize second layer's particular neuron (9)

Visualizing and Understanding Convolutional Networks. Matthew D. Zeiler and Rob Fergus

Visual Concepts Learned by CNNs

Layer 3

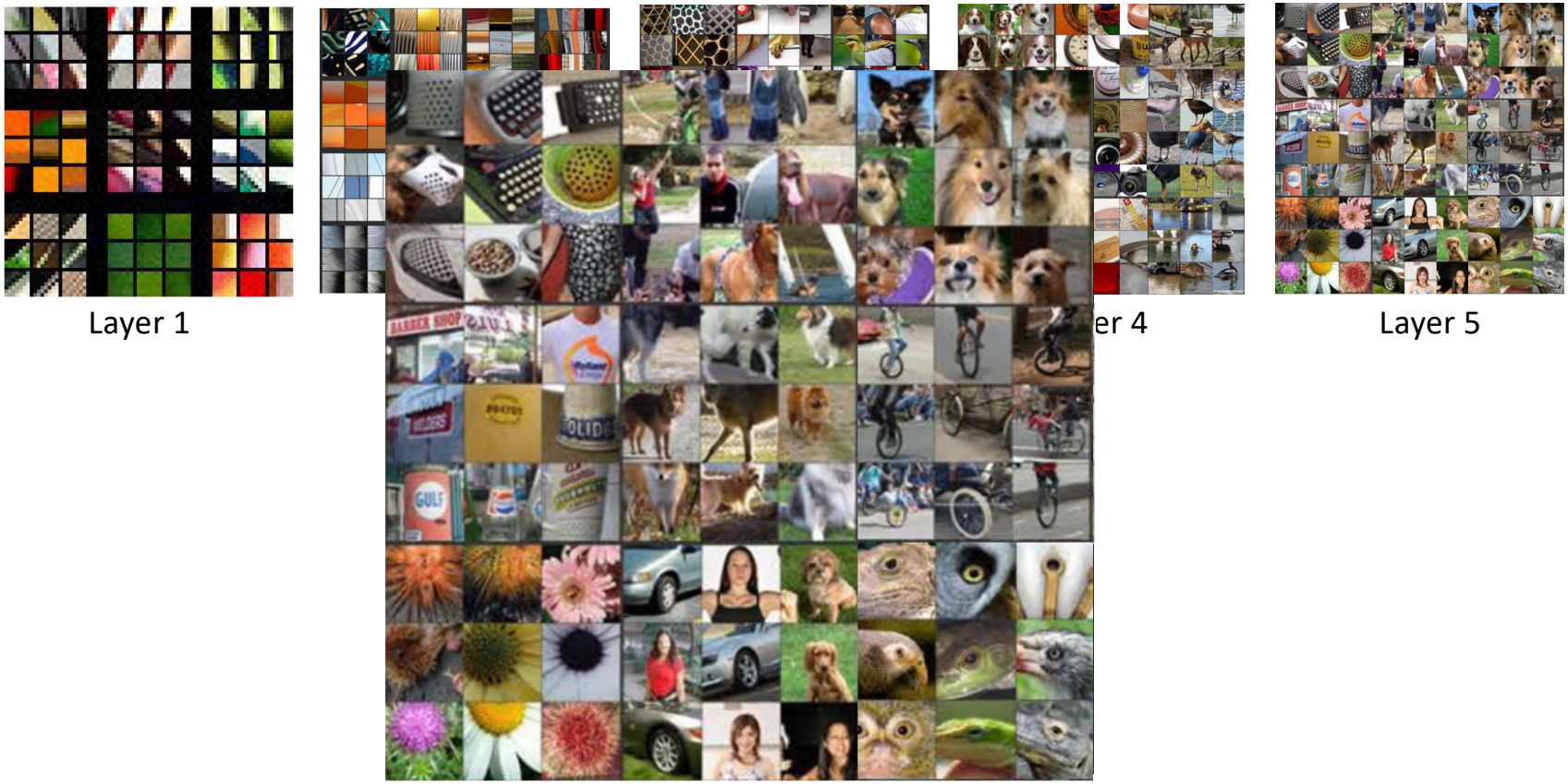


Layer 4

Layer 5

Visualizing and Understanding Convolutional Networks. Matthew D. Zeiler and Rob Fergus

Visual Concepts Learned by CNNs



Layer 1

er 4

Layer 5

Visualizing and Understanding Convolutional Networks. Matthew D. Zeiler and Rob Fergus

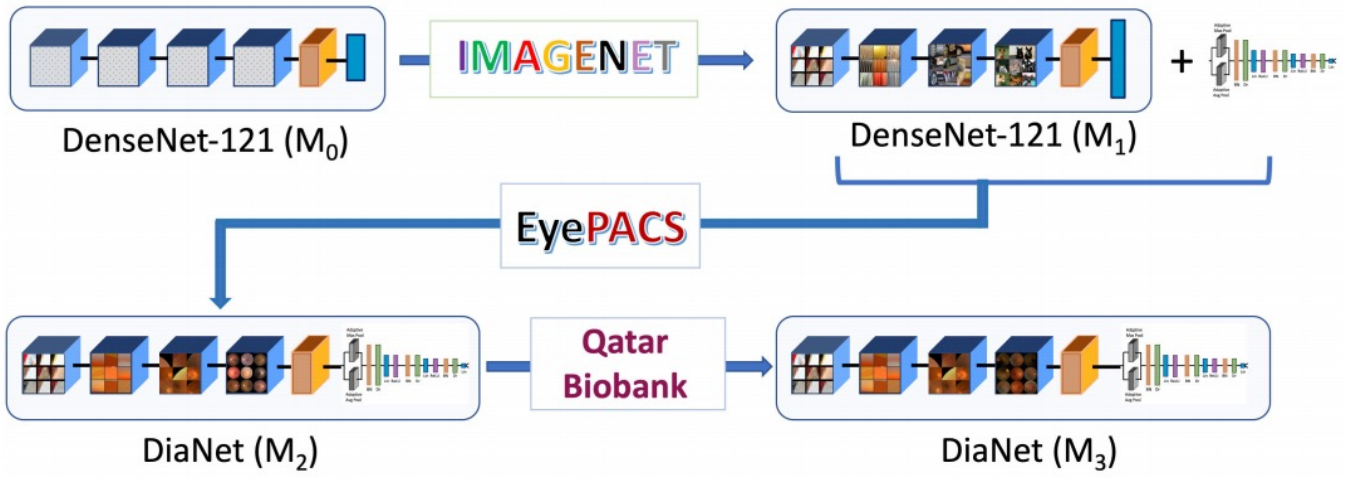
Transfer Learning

- Allows a deep neural network to perform well using only a few data points
- Works by transferring the knowledge learned on a task similar to the target task.
- Pre-trained network architectures available at model zoos.
- Pre-trained models are trained (fine-tuned) on the smaller dataset

Case Study 1:
Diagnosis of Diabetes using Retinal Images

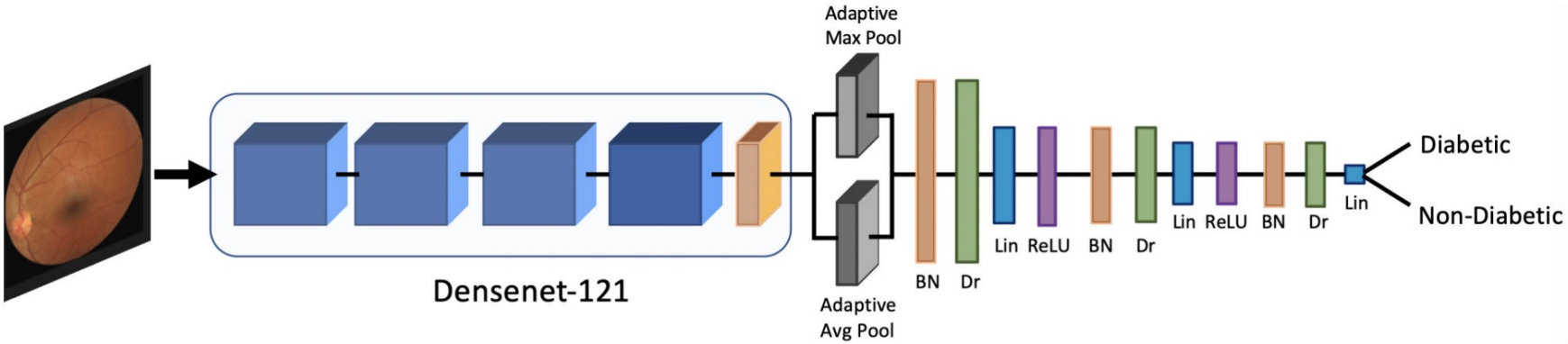
- Diabetes is one the leading fatal diseases
- Usually diagnosed using clinical biomarkers
- Retina Image => Diabetic Retinopathy (DR) exists but
  Retina Image => Diabetes does not
- Proposed a CNN-based approach
  - Multi-stage fine-tuning

Case Study 1:
Diagnosis of Diabetes using Retinal Images



- M. T. Islam, H. R. H. Al-Absi, E. A. Ruagh and T. Alam, **"DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only,"** in *IEEE Access*, vol. 9, pp. 15686-15695, 2021, doi: 10.1109/ACCESS.2021.3052477.

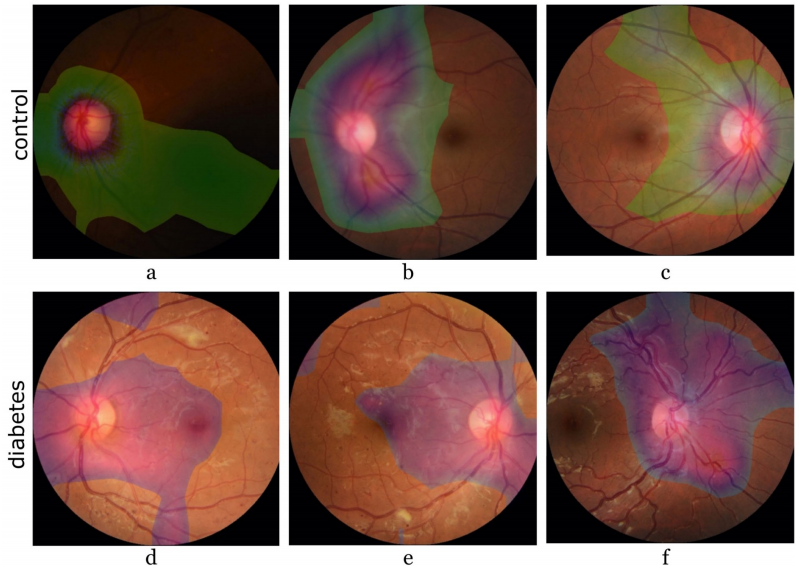
Case Study 1:
Diagnosis of Diabetes using Retinal Images



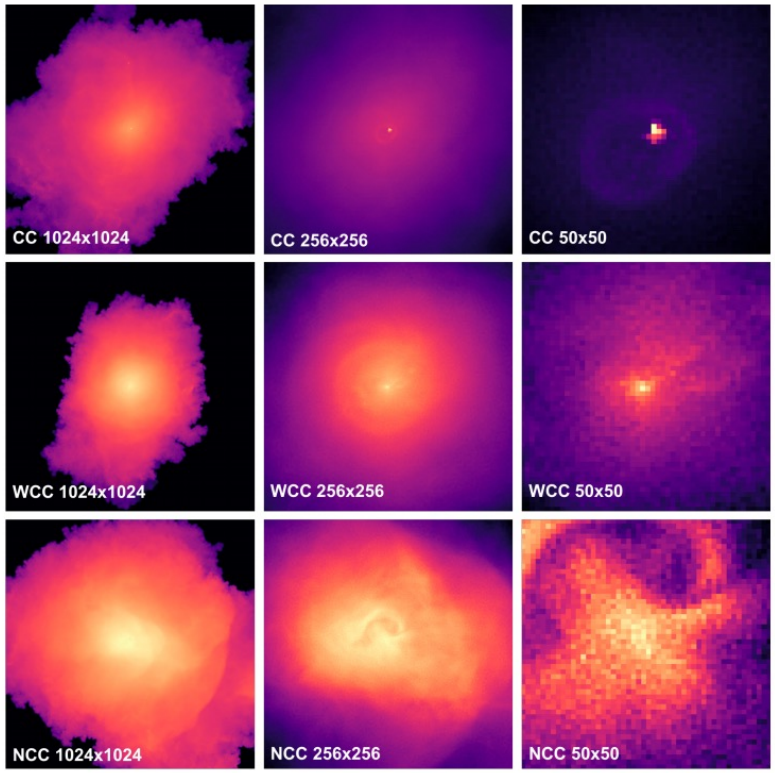
Layer Name	Output Size	Number of Parameters
AdaptiveMaxPool2D	[1024, 1, 1]	0
AdaptiveAvgPool2D	[1024, 1, 1]	0
Flatten + Concat	[2048]	0
BatchNorm1D	[2048]	4096
Dropout	[2048]	0
Linear	[512]	1,049,088
ReLU	[512]	0
BatchNorm1D	[512]	1024
Dropout	[512]	0
Linear	[256]	131,328
ReLU	[256]	0
BatchNorm1D	[256]	512
Dropout	[256]	0
Linear	[2]	512

Case Study 1:
Diagnosis of Diabetes using Retinal Images

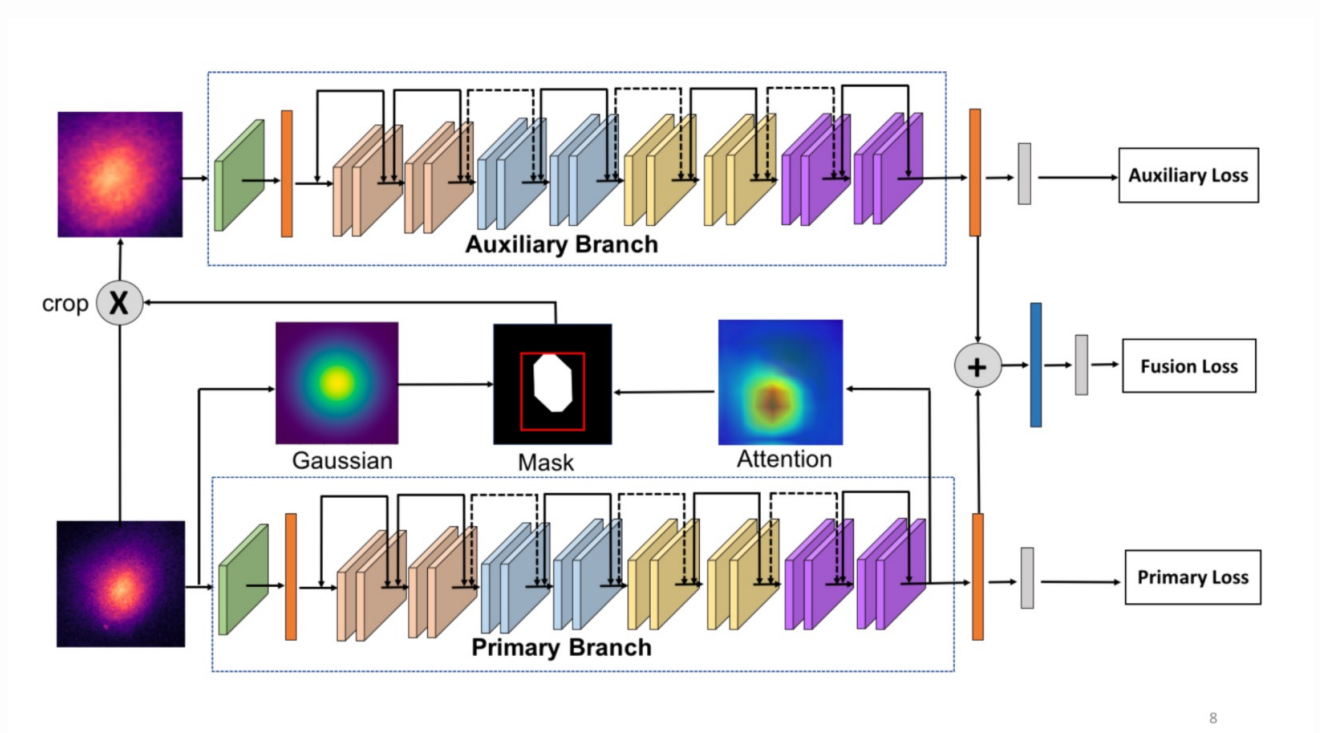
Metric	Accuracy	Precision	Sensitivity/Recall	Specificity	F1 Score	AUC ROC
QBBNet_Res50+ XGB	80.65	79.27	83.15	78.14	81.16	80.64
DiaNet_Res50 + XGB	82.01	80.41	84.78	79.23	82.53	82.08
QBBNet + XGB	76.56	76.63	76.63	76.5	76.63	76.56
DiaNet + XGB	83.92	81.09	82.58	79.23	84.67	83.91
QBBNet_Res50	80.38	80.94	76.63	79.15	79.66	80.39
DiaNet_Res50	83.1	81.77	85.32	80.87	83.51	83.1
QBBNet	79.02	78.01	80.97	77.04	79.47	79.01
DiaNet	**84.47**	**83.59**	**85.86**	**83.06**	**84.71**	**84.46**



Case Study 2:
Galaxy Cluster Classification



Case Study 2:
Galaxy Cluster Classification



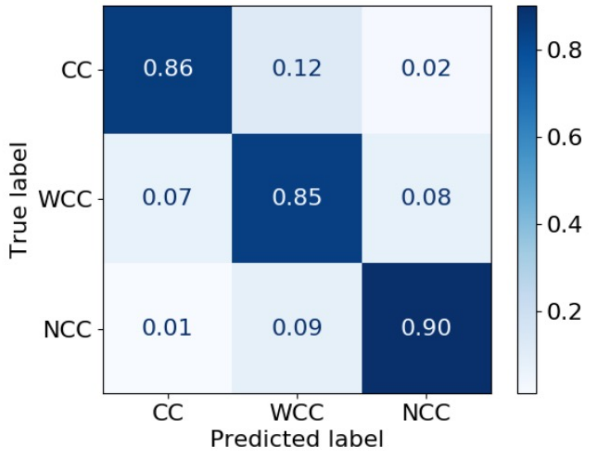
Multi-Branch Attention Networks for Classifying Galaxy Clusters. Zhang et al. ICPR 2020.

Case Study 2:
Galaxy Cluster Classification

- Both the primary and auxiliary branches are classification networks that predict the core type of the galaxy cluster
- Consist of a portion of ResNet-18 architecture
- In the primary branch, ResNet architecture is used for high-level feature map extraction.
  - feature map as is a 8×8 tensor with 512 channels (output of the last residual block before the Global Average Pooling (GAP) layer)
  - It is used to generate an attention map
    - Class Activation Mapping is performed to localize the discriminative regions used by the CNN to identify the cluster type. We compute a weighted sum of M to obtain a class activation map (CAM) for the input image.
- For the input image, a parametric function is formulated using bivariate Gaussian distribution.
- A binary mask is created by taking the union of the attention mask and the Gaussian mask.
- The binary mask to crop an informative region from the input image and pass it to the auxiliary branch for classification.
- The output vectors from the GAP layers from the branches are concatenated and passed through a fully-connected layer.

Case Study 2:
Galaxy Cluster Classification

Approach	Attention	Gaussian	Regression	macro-avg. f1	class	precision	recall	f1
Baseline	✗	✗	✗	0.803	CC	0.59	0.79	0.68
					WCC	0.92	0.85	0.88
					NCC	0.84	0.86	0.85
Ours(Att)	✓	✗	✗	0.823	CC	0.62	0.81	0.70
					WCC	0.93	0.86	0.89
					NCC	0.86	0.90	0.88
Ours(Gauss)	✗	✓	✗	0.813	CC	0.58	0.79	0.67
					WCC	0.93	0.85	0.89
					NCC	0.86	0.90	0.88
Ours(Att+Gauss)	✓	✓	✗	0.827	CC	0.67	0.79	0.73
					WCC	0.91	0.86	0.89
					NCC	0.84	0.88	0.86
Ours(all)	✓	✓	✓	**0.830**	CC	0.65	0.86	0.74
					WCC	0.94	0.85	0.89
					NCC	0.83	0.90	0.86



Case Study 3:
Generative Adversarial Networks (GANs)

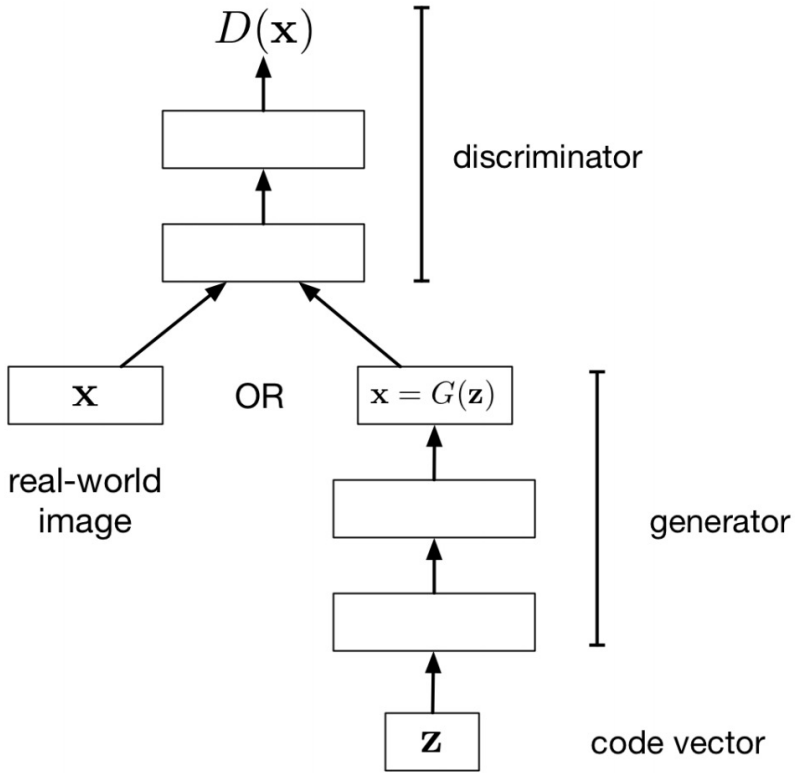


Image credit: CSC321 Lecture 19: Generative Adversarial Networks, Roger Grosse https://www.cs.toronto.edu/~rgrosse/courses/csc321_2018/slides/lec19.pdf

Applications of GANs: Generating Realistic Human Faces

https://thispersondoesnotexist.com/

Applications of GANs: Generating Realistic Human Faces

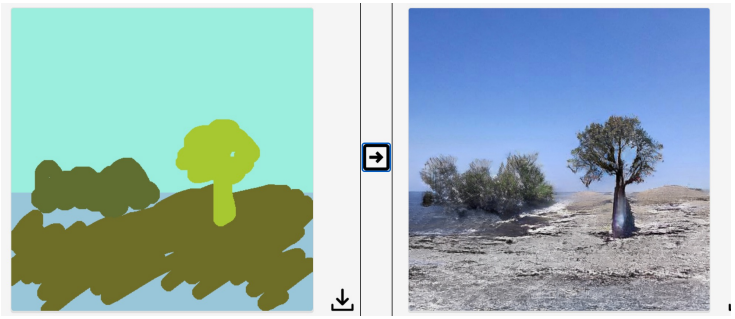
https://thispersondoesnotexist.com/

Applications of GANs: Object Generation

Progressive Growing of GANs for Improved Quality, Stability, and Variation. Karras et al. 2017.

Applications of GANs: From Sketch to Landscape

- GauGAN: Changing Sketches into Photorealistic Masterpieces



- Interactive Demo:
  http://nvidia-research-mingyuliu.com/gaugan

- Demonstration from the Developers
  https://www.youtube.com/watch?v=p5U4NgVGAwg

Applications of GANs: Animating Still Photographs

Deep Nostalgia™ by MyHeritage:
https://www.myheritage.com/deep-nostalgia

Trying MyHeritage Deep Nostalgia™
https://www.youtube.com/watch?v=_W47RHLpMsA

Introducing Deep Nostalgia™ Special Animations
https://www.youtube.com/watch?v=WJ7Jn1Zp58o

Questions

Neural Style Transfer

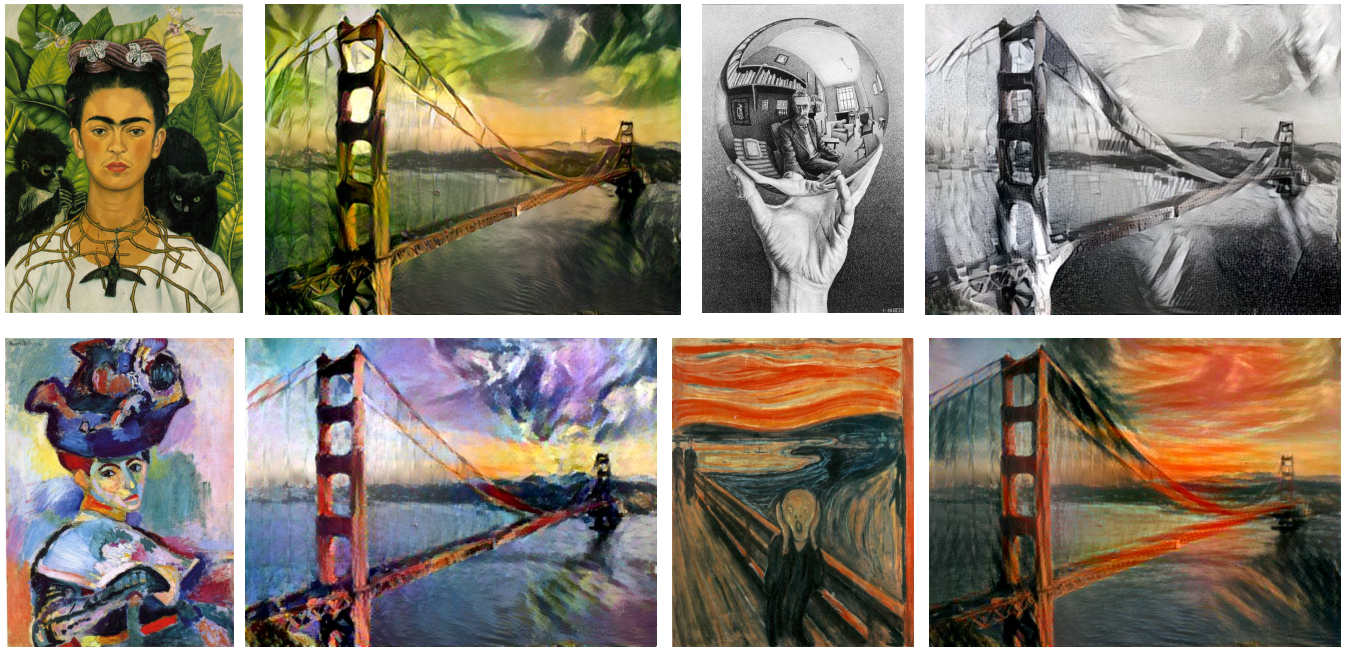
https://www.youtube.com/watch?v=Khuj4ASldmU

https://github.com/jcjohnson/neural-style

Neural Style Transfer

https://github.com/jcjohnson/neural-style

Neural Style Transfer



https://github.com/jcjohnson/neural-style

Neural Style Transfer: Algorithm

- Start with a random noise image as the image to be generated (G)
- In each iteration
  - compare the style of the style image (S) and G – it is the style loss
    $J_{style}(C,G)$
  - compare the content of the content image (C) and G – it is the content loss
    $J_{content}(S, G)$
  - Combine the two loss using appropriate weights:
    $J(C, S, G) = \alpha * J_{content}(C,G) + \beta * J_{style}(S, G)$
- Find gradients of the total cost w.r.t. the input
- Alter the input according to the gradients
- Repeat until some criteria is met (i.e. # of iterations, # cost threshold)

Content Cost

- Select a layer l to be used to compare contents
- Use a pre-trained CNN (e.g. VGG trained on ImageNet)
- Compute the response of C and G at layer l: $R_{[l][C]}$ and $R_{[l][G]}$
  - Responses should be similar if C and G are similar (content-wise)
- Define $J(C, G) = Norm_{L2}(R[l][C], R[l][G])$

Style Cost

- What is style?

  Correlation (cross-covariance) between responses for different convolution filters

- For a layer l:
  - Gram[l][S] = Cross_Covariance($R_k^{[l][S]}$, $R_{k'}^{[l][S]}$)
  - Gram[l][G] = Cross_Covariance($R_k^{[l][G]}$, $R_{k'}^{[l][G]}$)
  - $J_{style}^{[l]}(S, G) = Norm_{L2}(G[l][S], G[l][G])$
- Sum over multiple layers for better results:
  - $J_{style}(S, G) = \Sigma J_{style}^{[l]}(S, G)$

Overall Cost & Minimization

Overall Neural Style Transfer cost:

$J(C, S, G) = \alpha * J_{content}(C,G) + \beta * J_{style}(S, G)$

Iteratively minimize J(C, S, G)

- Update G using gradients