Image Processing with Deep Learning

Michael Yuhas and Subrat Panda

Cyber-Physical Systems Research Group
Computer Science & Engineering, NTU

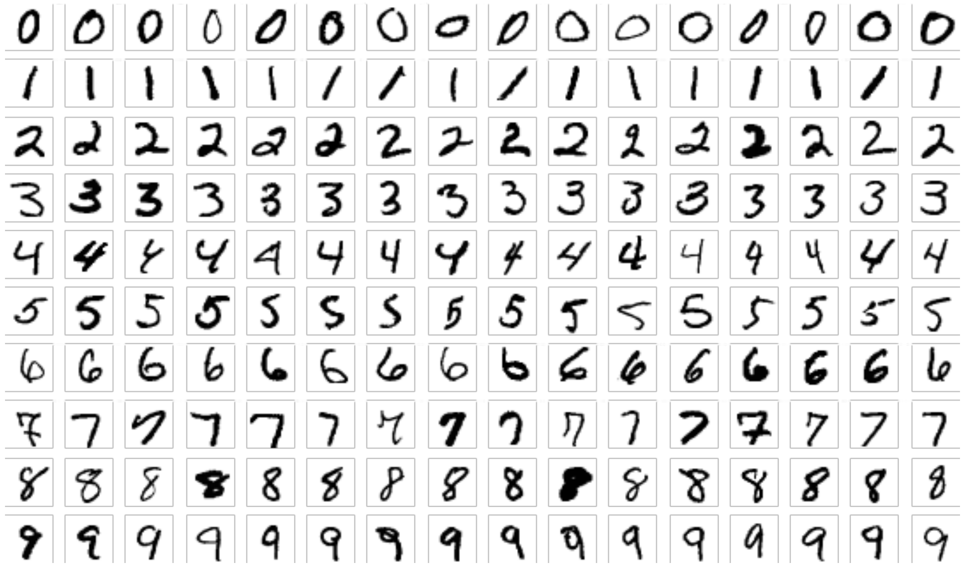
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agenda

Time	Item
2:00-2:15	Introduction to Convolution
2:15-2:45	Convolutional Neural Networks (CNNs)
2:45-3:15	Classification with CNNs
3:15-3:30	Break
3:30-4:00	Transfer Learning
4:00-4:30	Semantic Segmentation with CNNs
4:30-5:00	Other Techniques

A Real-World Problem…

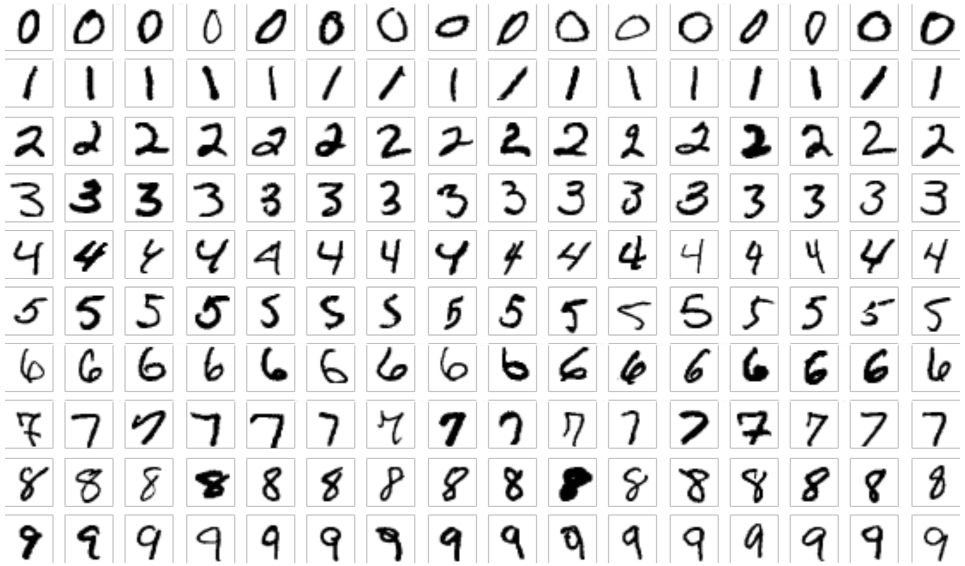
- Imagine you run a post office
- You receive millions of letters and packages a day and need to decide where to send them
- Human labor is costly and time consuming
- How can you automate this task?



[1] S. Janprasai, MNIST Examples Modified

A Real-World Problem…

- This was the problem facing USPS in 1989
- How would you solve this problem?

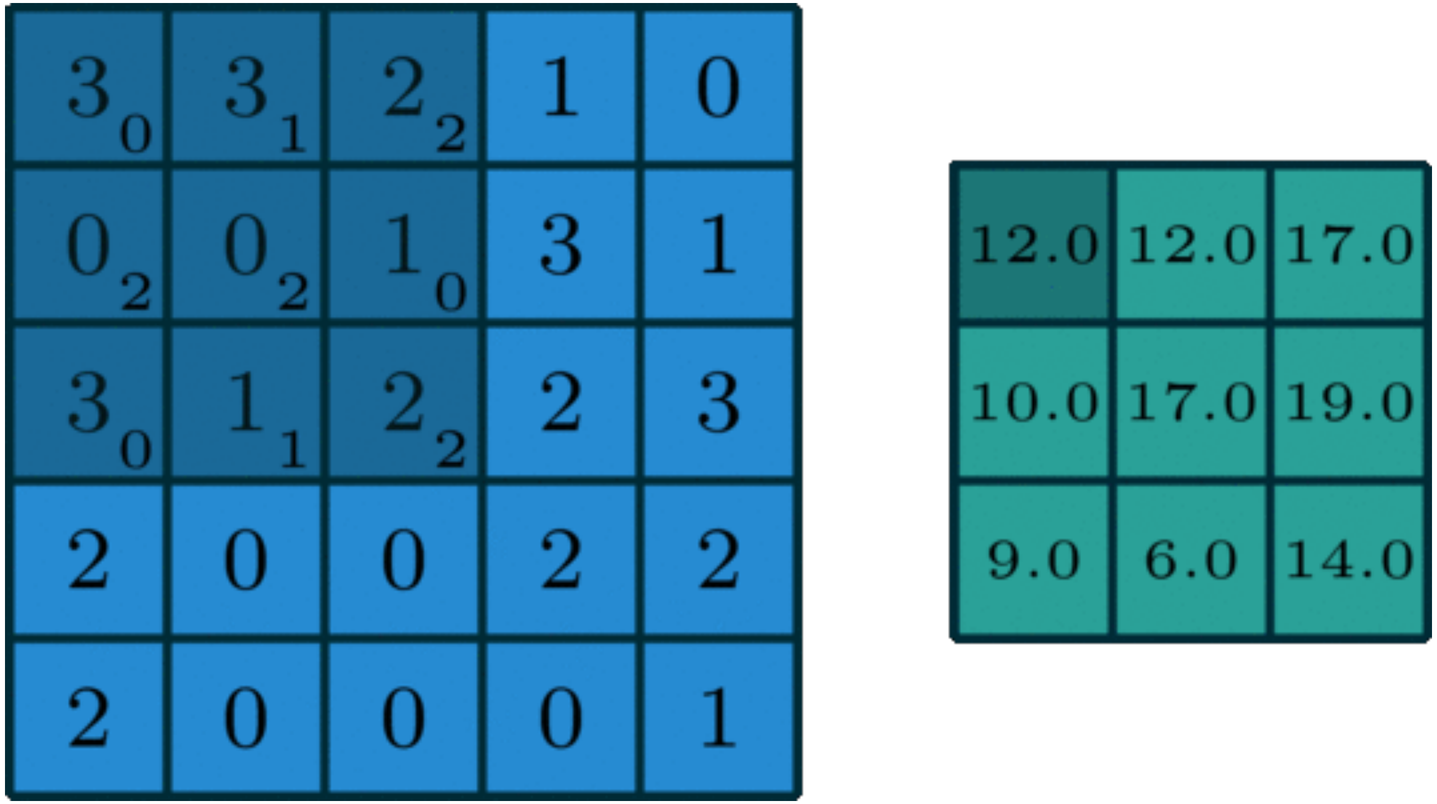


[1] S. Janprasai, MNIST Examples Modified

Introduction to Convolution

Convolution

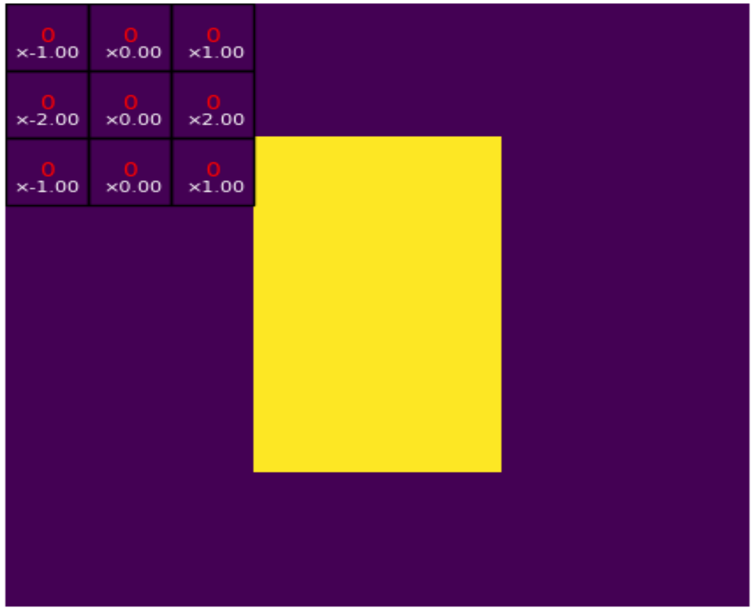
- An image is represented as a matrix of numbers
  - Higher value = More intesity
- A *kernel* is also a matrix containing some pattern
- Slide a kernel along an image an multiply the overlapping values, and sum the result
- $y[m,n] = \sum_{j=-\infty}^{+\infty} \sum_{i=-\infty}^{+\infty} x[i,j]h[m-i,n-j]$

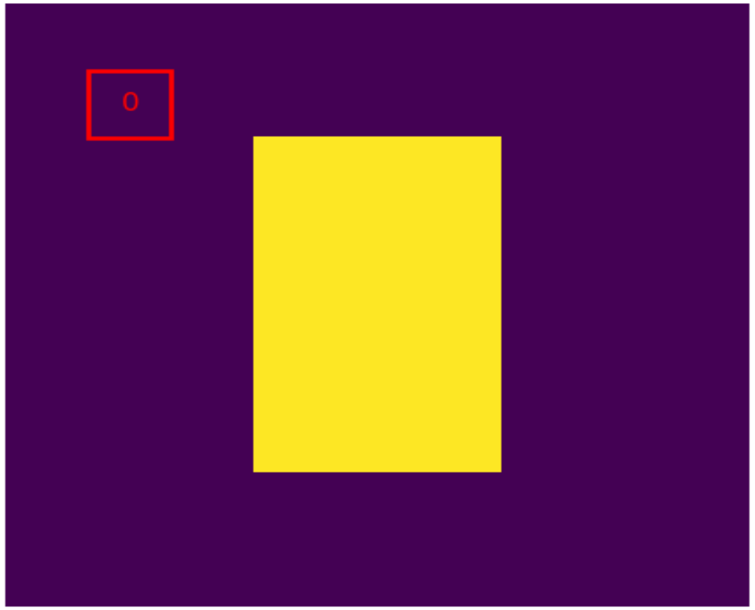


[2] Intuitively Understanding Deep Convolution, Towards Data Science (2018)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Edge Detection

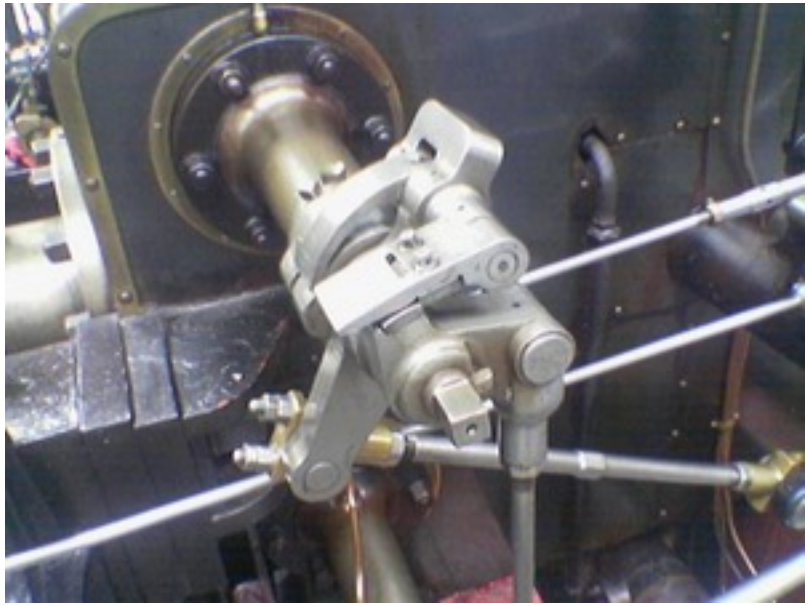




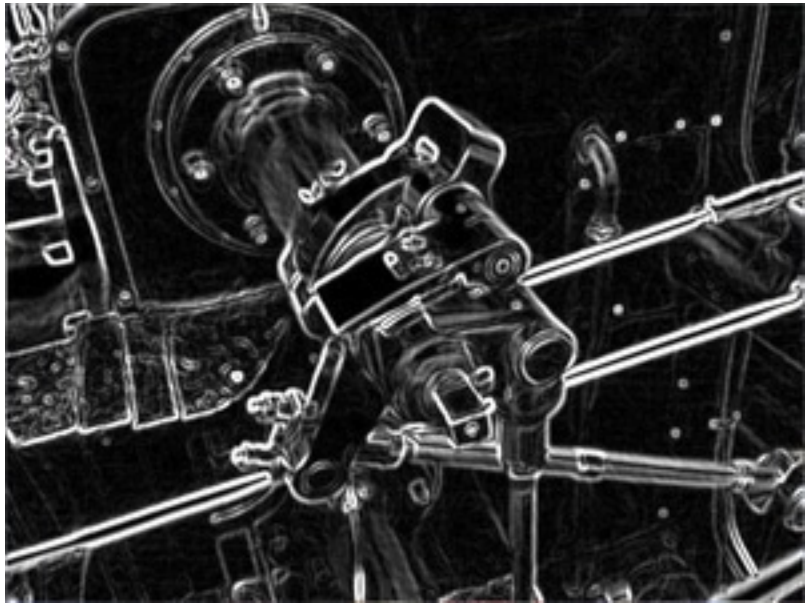
- *Sobel kernel* – a special kernel used for edge detection
- Yield a high value when a gradient is present in one direction

Exercise 1

- Write your own edge detector…



[3] Simpsons Contributor, Valve Original

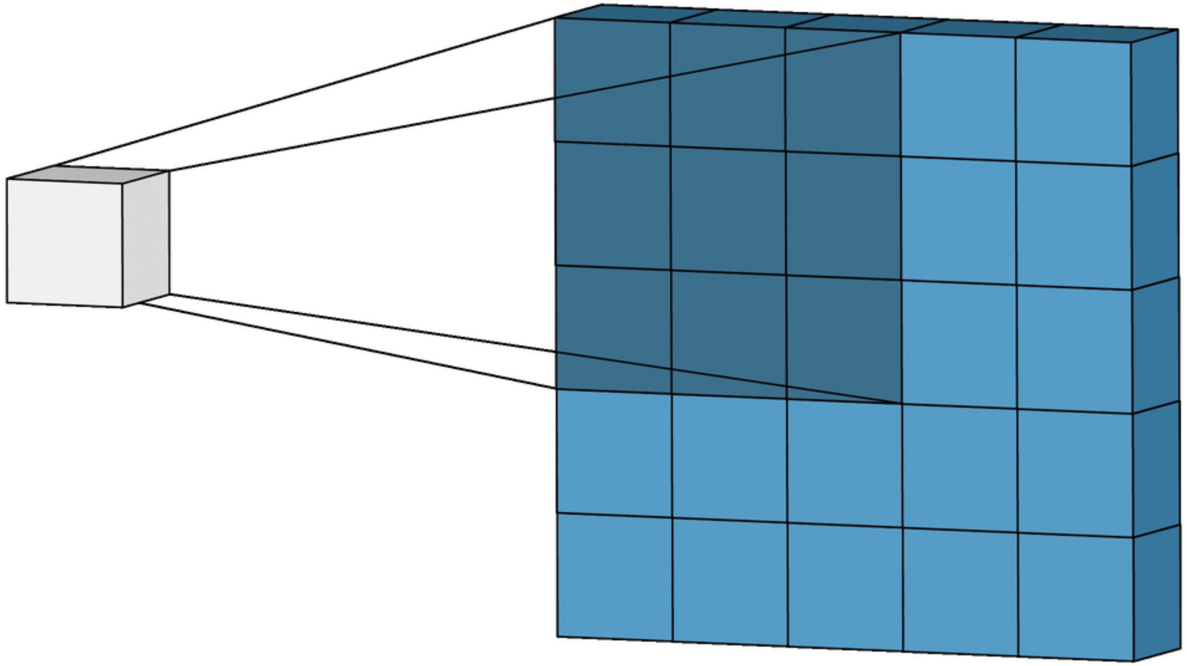


[3] Simpsons Contributor, Valve Sobel

Image Classification with CNNs

Learnable Filters

- What if we were able to learn kernel weights instead of crafting them ourselves
- This pattern, summing the multiplication of weights is the same thing that happens in a neural network!
- Not just limited to three color channels: can go from any arbitrary sized tensor to another



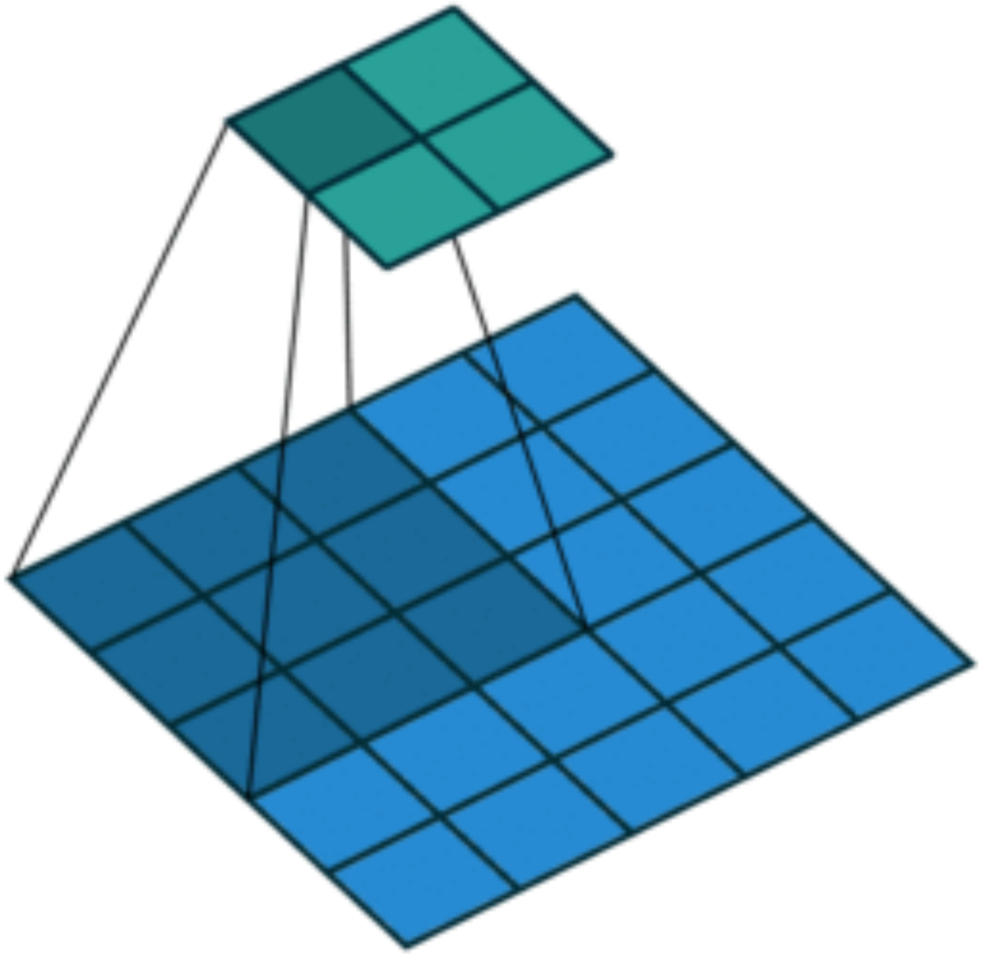
[4] Intuitively Understanding Deep Convolution, Towards Data Science (2018)

Padding

- https://ezyang.github.io/convolution-visualizer/
- Reasons to pad:
  - Information loss at edges of image
  - Need to match output dimensions
- Types of padding:
  - 0's
  - 1's
  - Reflective

Stride

- Skip pixels while dragging the kernel
- A higher stride let’s us…
  - Process data faster
  - Reduce dimensionality
- At the expense of…
  - Ignoring information in the input



[4] Intuitively Understanding Deep Convolution, Towards Data Science (2018)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

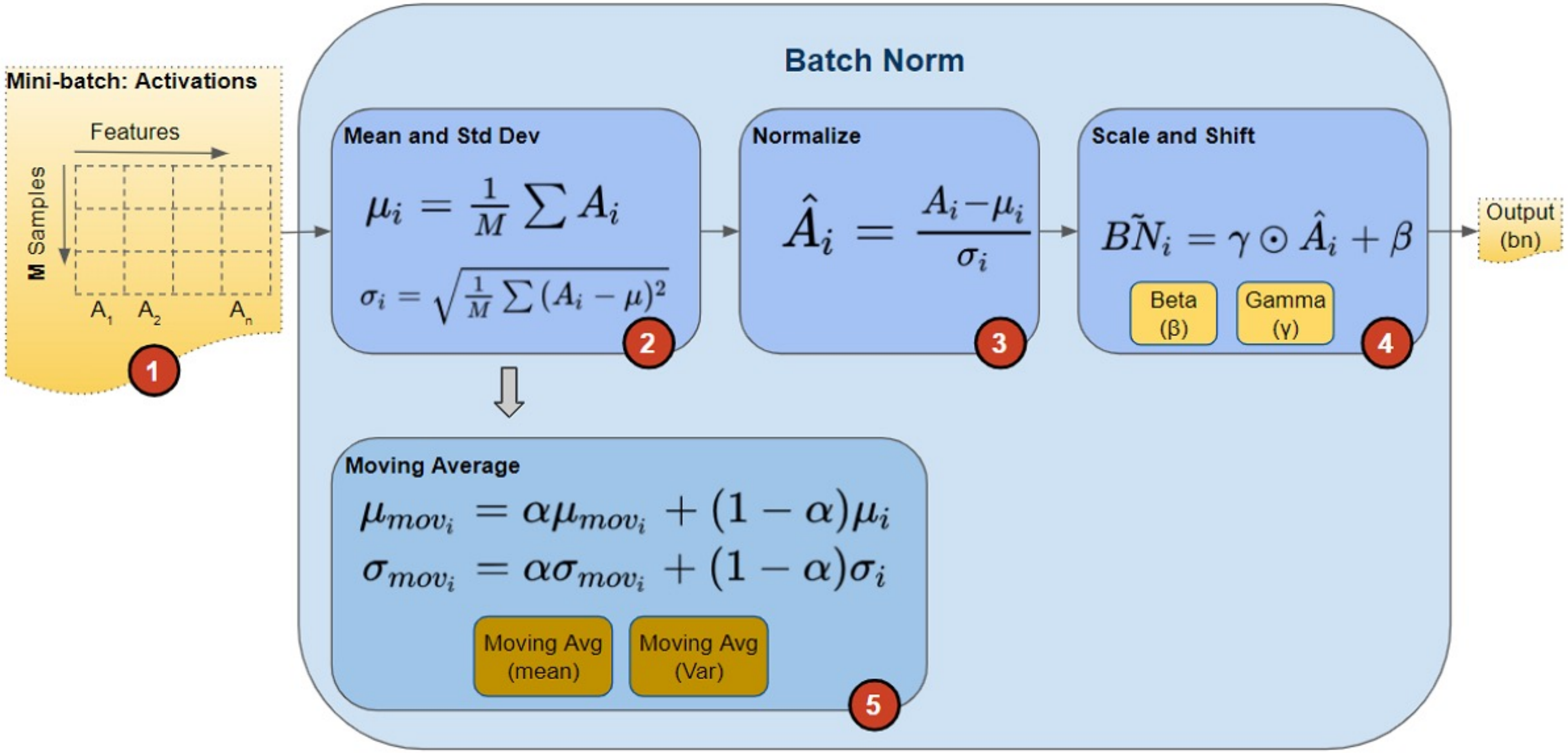
Dilation

- https://ezyang.github.io/convolution-visualizer/
- Increase the receptive field of a neuron without increasing number of computations

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Normalization

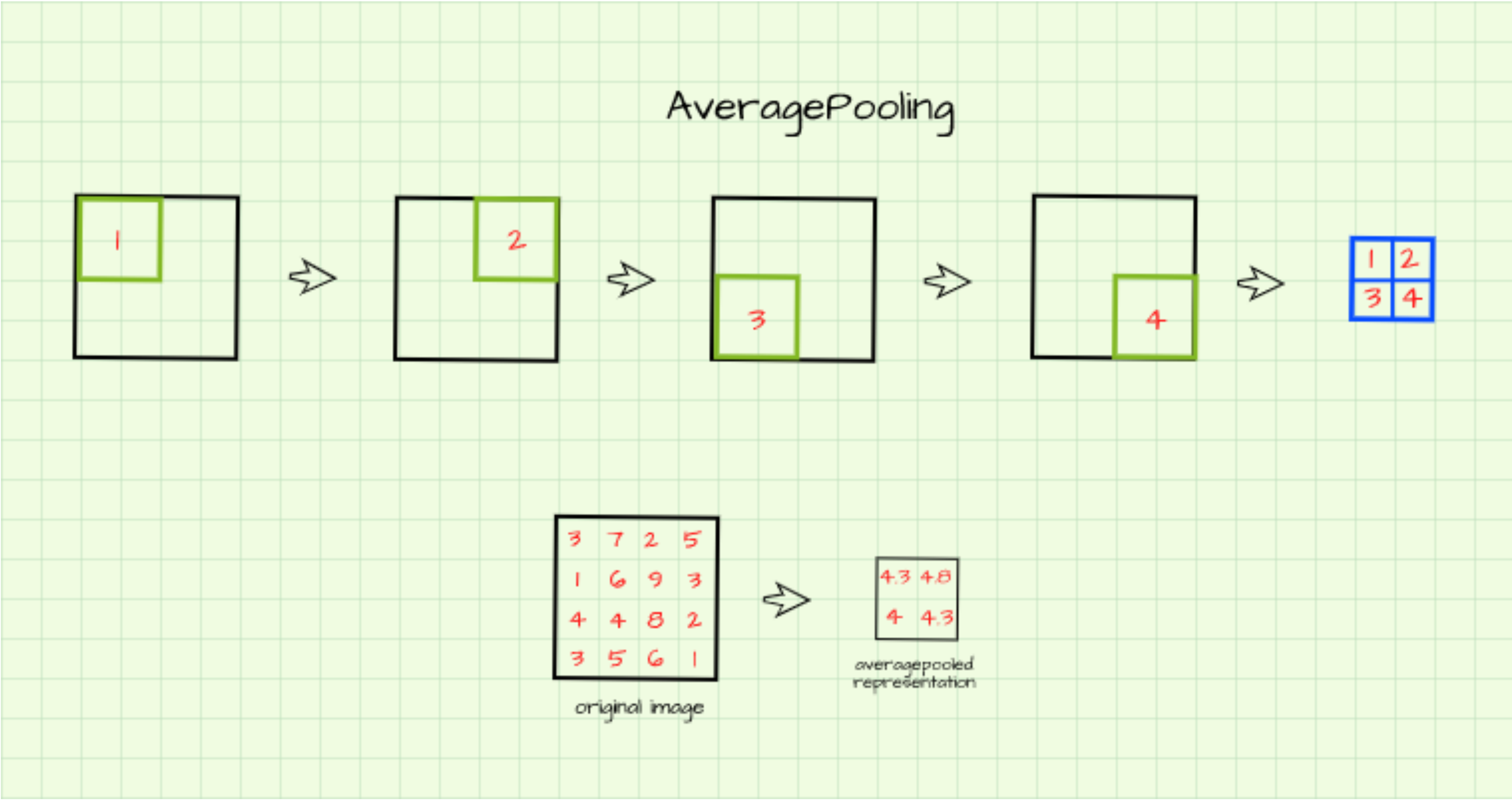
- We want to normalizae feature values to help gradient descent converge
- Two learnable parameters
  - $\beta = E[x]$
  - $\gamma = \sqrt{var(x)}$
- Can place before or after activation function



[5] K. Doshi, *Batch Norm Explained Visually*, Towards Data Science (2021)

Pooling

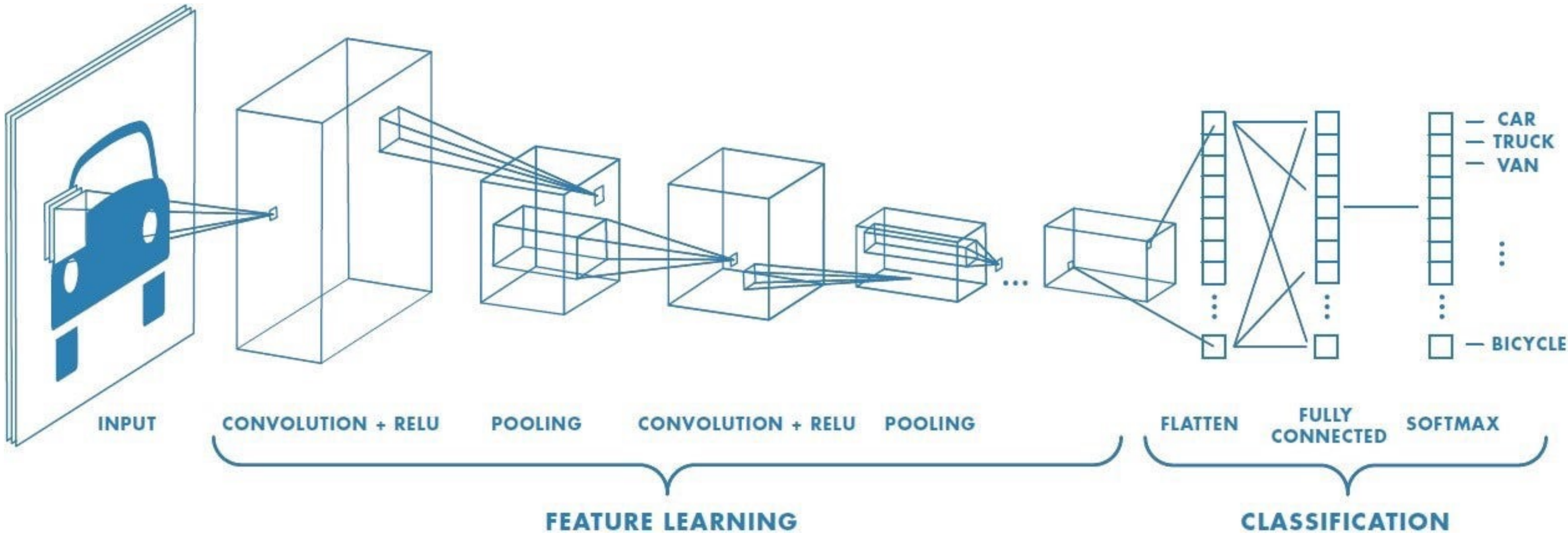
- Sometimes we want to reduce dimensionality without learning any new parameters
- Speed up computation cost
- Types:
  - Average
  - Max
  - Min



[6] O. Olu-Ipinlaye, *Pooling in Convolutional Neural Networks*, PaperSpace (2023)

Adding Depth

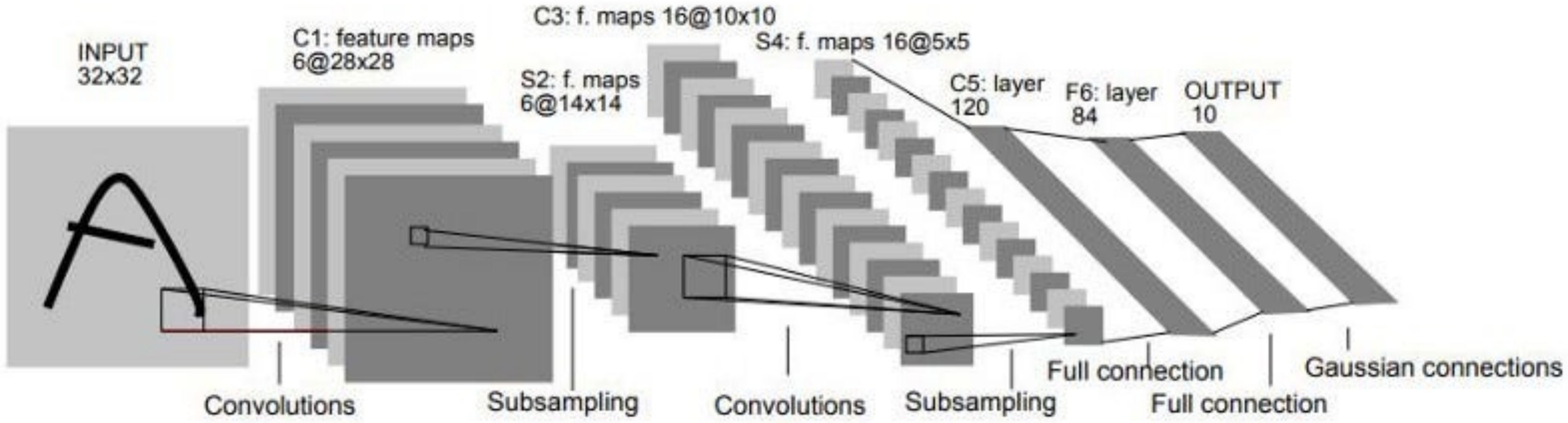
- In shallow networks we learn a function that maps a feature space to an output
- In deep learning, we don't know the features, so we use one conv layer to learn them
- Repeat *ad infinitum*



[7] S. N. Gupta, *Deep Convolutional Neural Networks Explained in Layman's Terms*, Medium (2022).

LeNet

- Invented by Yann LeCun in 1989 to classify postcode digits
- 2 convolutional layers followed by 2 fully connected layers
- 28x28 black and white input image
- Surprisingly powerful (can use this to make a car follow lanes)

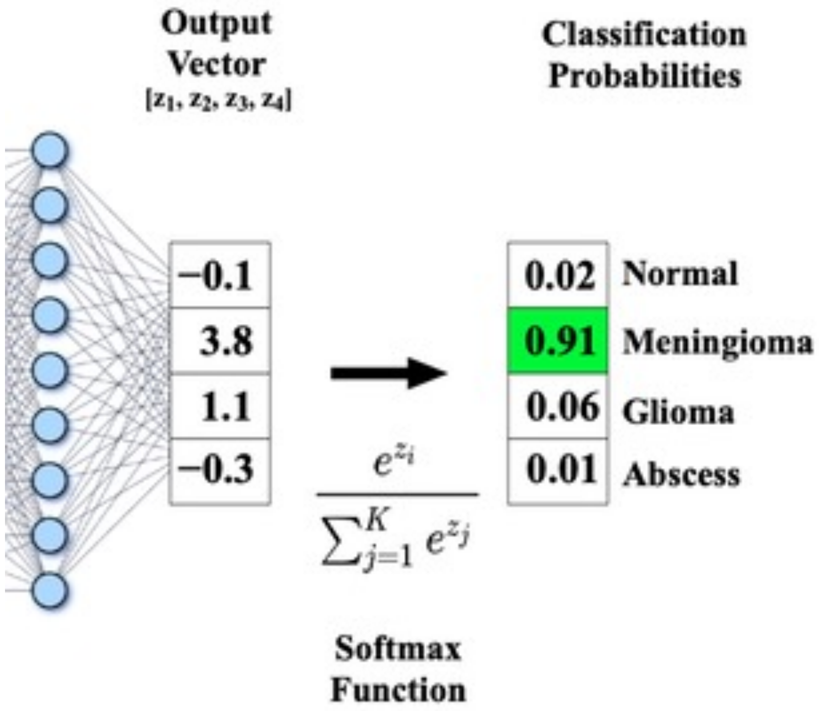


[8] S. Bangar, *LeNet5 Architecture Explained*, Medium (2022)

NLLLoss and Softmax

[9] Questions and Answers in MRI

- Sometimes we want to know not just which class is most likely, but what is its probability
- With this probability value, Softmax let's us create a loss function based on probability
- We take the log of the probability to make this play nicely with optimizers (like Adam)

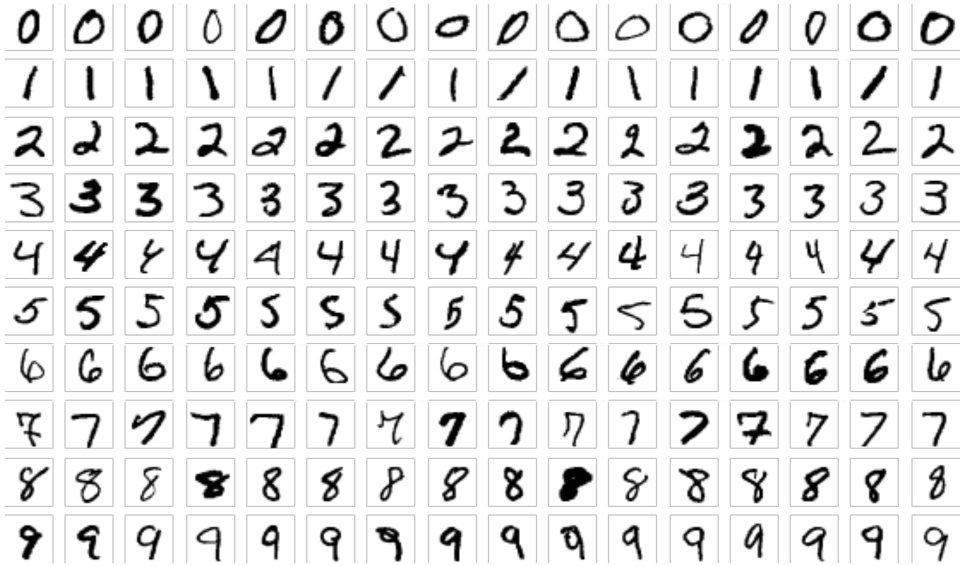


$$\log \mathbb{P}(\mathcal{D}|\theta) = \sum_{i=1}^{n} \left(y_i \log \hat{y}_{\theta,i} + (1 - y_i) \log (1 - \hat{y}_{\theta,i}) \right)$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Exercise 2: Image Classification

- Now let’s put everything we learned into practice…
- Can you solve the character recognition problem?



[1] S. Janprasai, MNIST Examples Modified

Break Time

Cyber-Physical Systems Research Group
Computer Science & Engineering, NTU

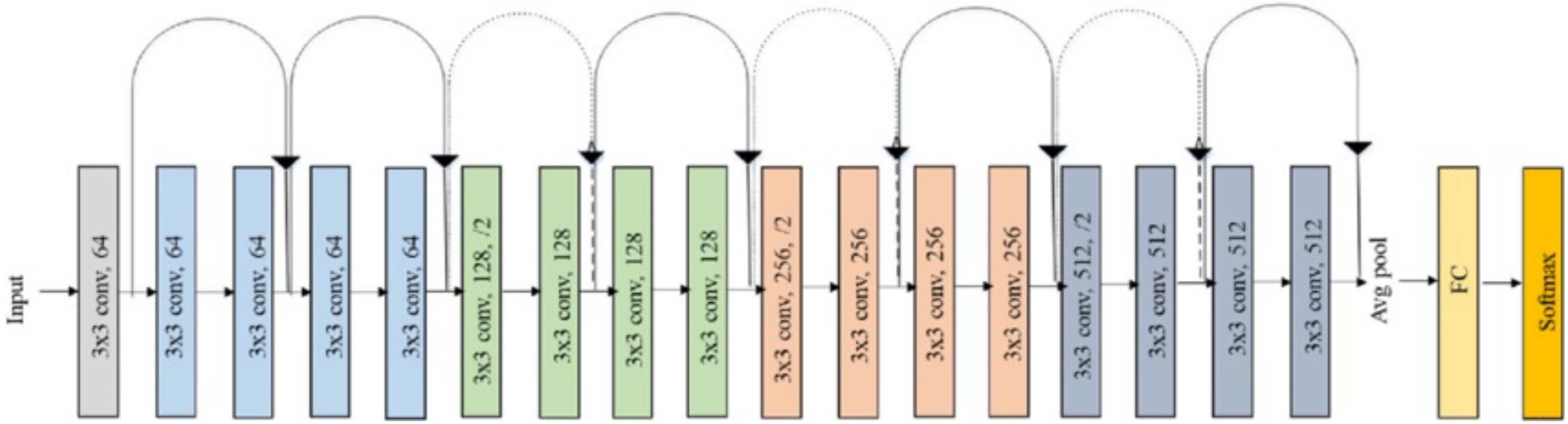
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Using Pretrained Models

Cyber-Physical Systems Research Group
Computer Science & Engineering, NTU

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Resnet

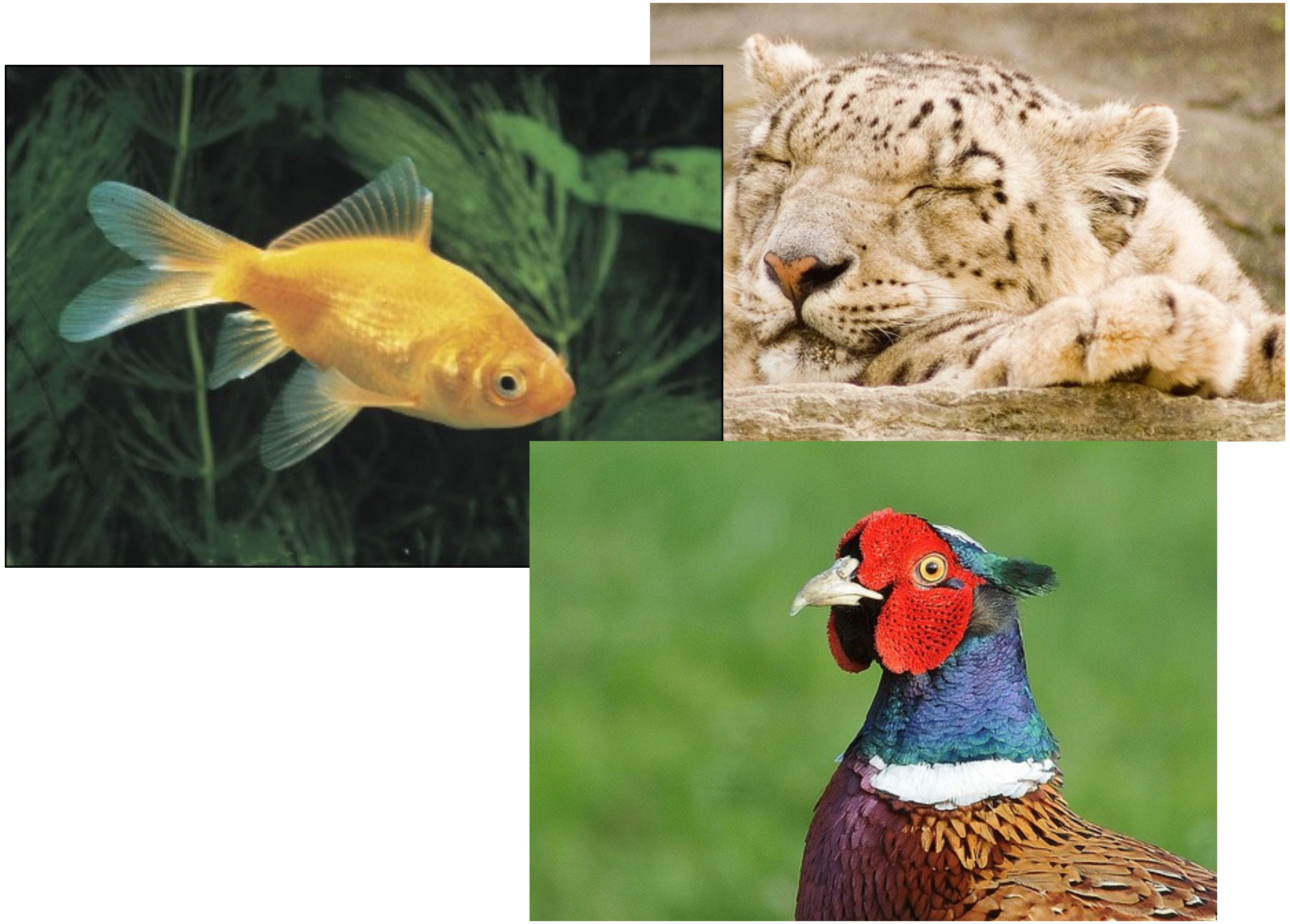


[10] S. Igbal, [Original Resnet-18 Architecture]

- We can go deeper, but we suffer the vanishing gradient problem
- Resnet solves this with skip connections
- Comes in variants from 18 layers to over 100
- However, this comes at a cost of training time

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

ImageNet



[11] ImageNet

- Benchmark dataset for color image classification
- Over 1000 classes
- 1,281,167 training images
- 50,000 validation images
- 100,000 test images

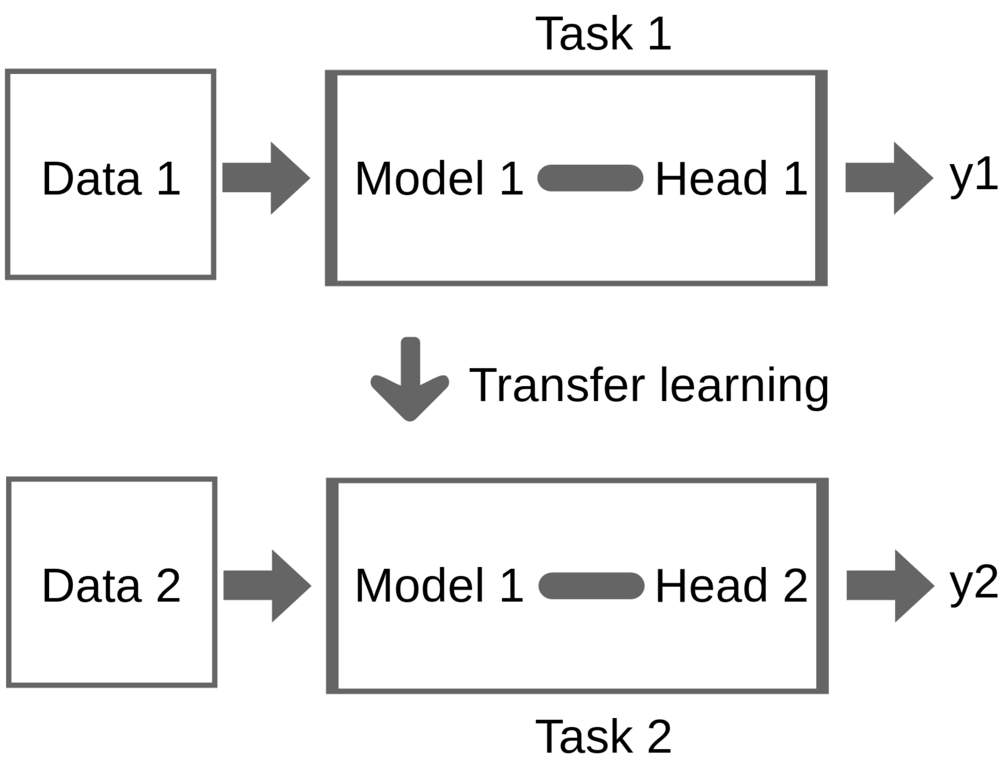
Exercise 4: Load a Pretrained Model from TorchHub

- Let’s use a more complex classification model…

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Transfer Learning

- Use a model trained on one dataset to make predictions on another dataset
- We need to fine tune on some data from the second dataset
- Can greatly reduce training time



[12] Biggerj1, *Illustration of Transfer Learning*

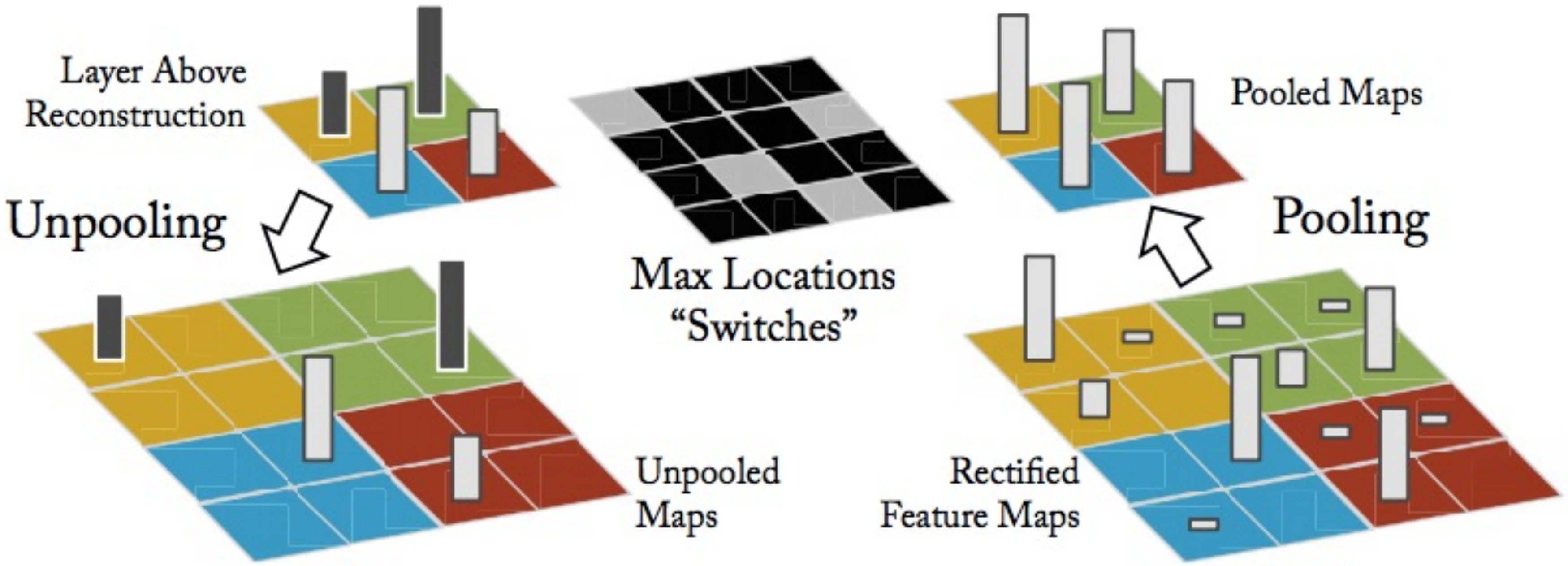
Exercise 5: Transfer Learning

- Can you fine-tune Resnet to classify images of Pokemon?

[13] R. Banik, *The Complete Pokemon Dataset*, Kaggle (2018)

Semantic Segmentation with CNNs

Unpooling

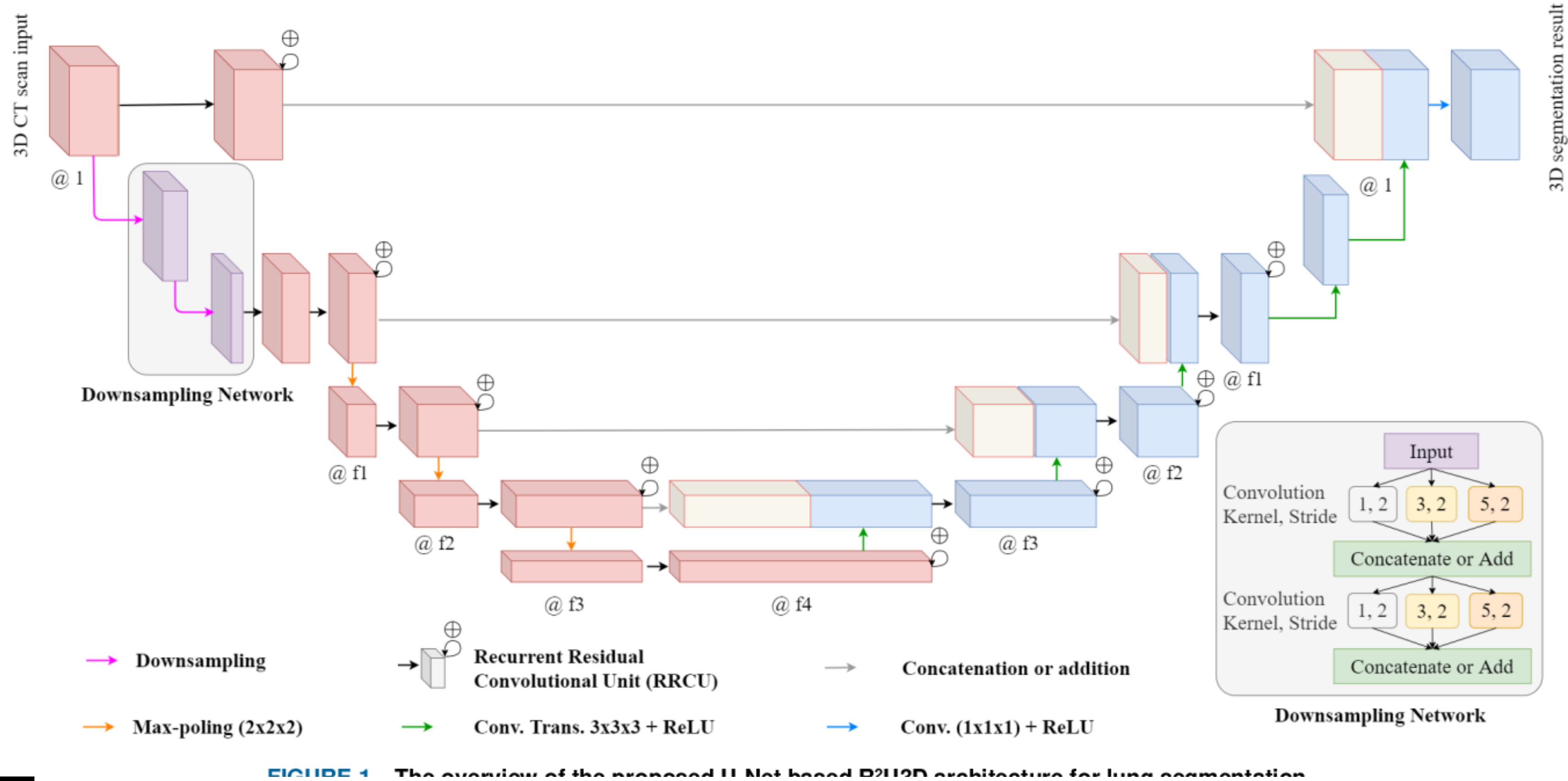


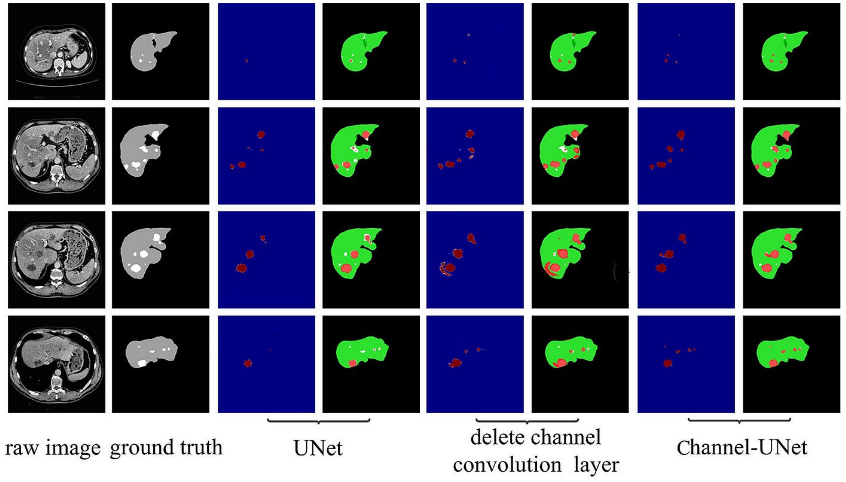
[14] [Visualizing and Understanding Convolutional Neural Networks](), CMU (2016)

- If we can reduce dimensions, we can also increase them
- Usefull for image generation, augmentation, and segmentation

UNET

- An Image Segmentation Architecture
- Classify each pixel in an input image
- Useful for lane detection, medical imaging, etc.





[15] A. Quesada, *U-Net: A Versatile Deep Learning Architecture for Image Segmentation*, Medium (2023)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Image Generation with Diffusion Models