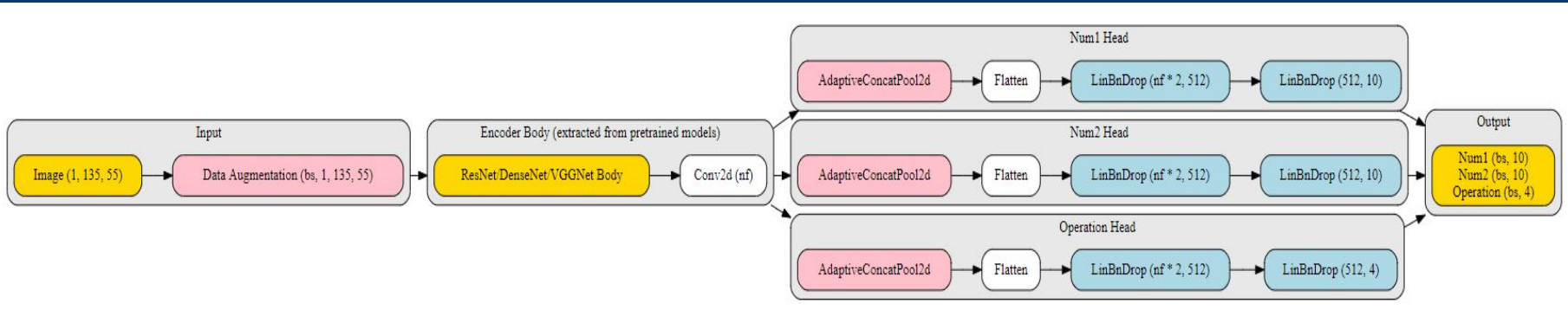
# NTU COMPUTING CHALLENGE: TEAM OMEGA

Anderson Serangoon Junior College

# RECAP OF PROBLEM STATEMENT (HANDWRITTEN MATH EXPRESSIONS)

# MODEL ARCHITECTURE (MULTI-HEAD CNN)



# “FASTAI” FRAMEWORK

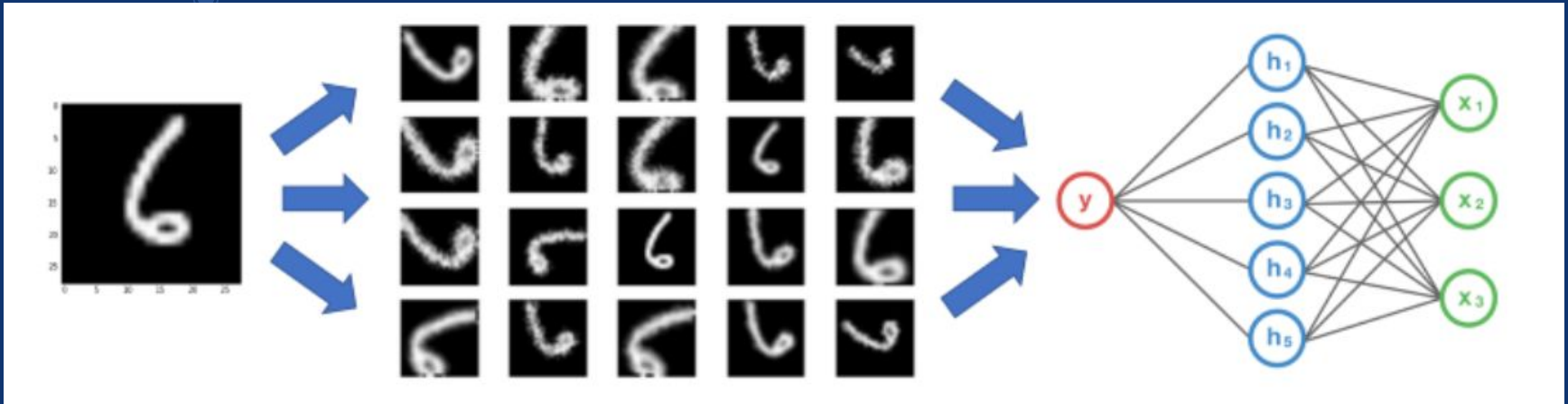
Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning. Information, 11(2), 108.

# WHAT IS “FASTAI”

fastai is a deep learning library which provides practitioners with high-level components that can quickly and easily provide state-of-the-art results in standard deep learning domains, and provides researchers with low-level components that can be mixed and matched to build new approaches.

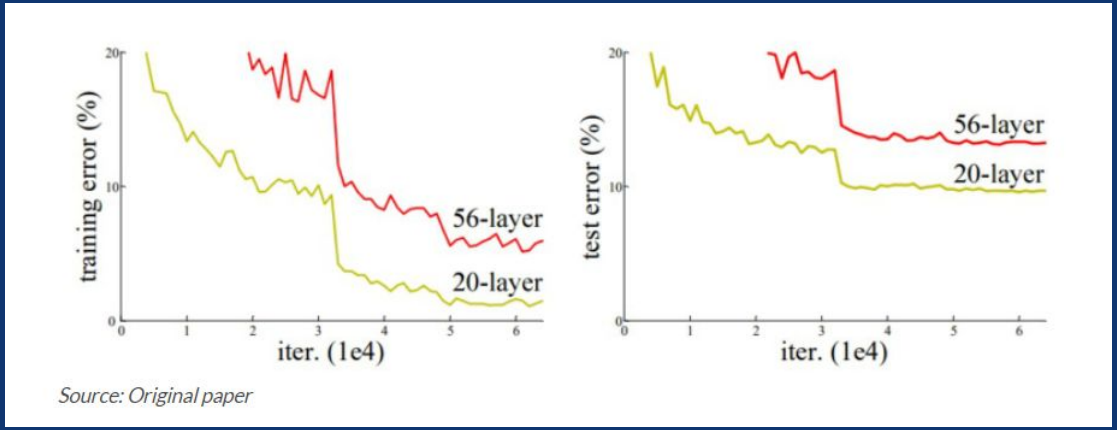
# DATA AUGMENTATION

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data.


y
h1
h2
h3
h4
h5
x1
x2
x3

# RESIDUAL NEURAL NETWORK (RESNET)

Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun (2015). Deep Residual Learning for Image Recognition



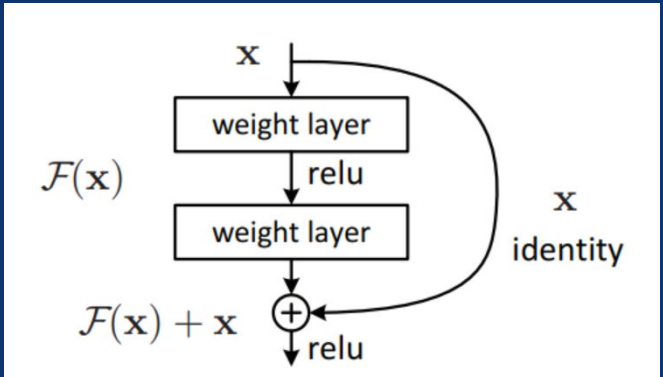

*Source: Original paper*

# ADDING MORE LAYERS ON TOP OF A NETWORK WILL DEGRADE ITS PERFORMANCE

# RESIDUAL BLOCK

## SKIP CONNECTIONS

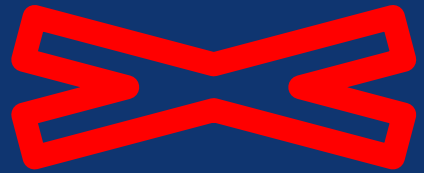
The skip connection skips training from a few layers and connects directly to the output.




## ADVANTAGE

If any layer hurts the performance of architecture then it will be skipped by regularization.

vanishing/exploding gradient problem



# DENSE CONVOLUTIONAL NETWORK (DENSENET)

Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger (2016), Densely Connected Convolutional Networks

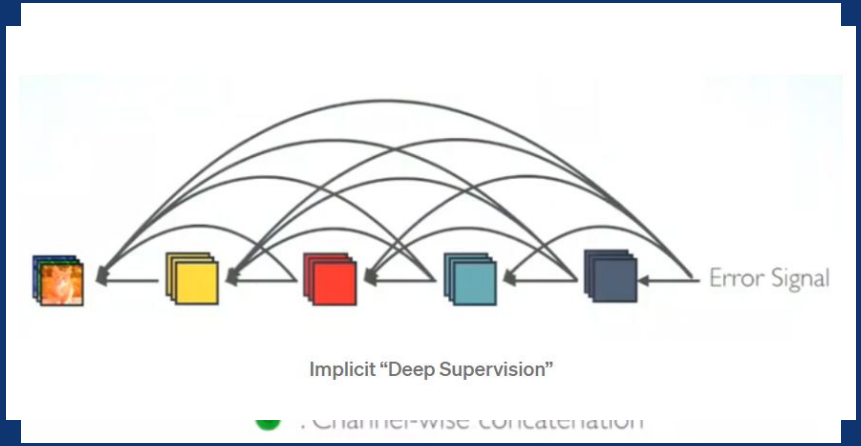
# DENSENET

## CONCATENATION

Each layer is receiving a “collective knowledge” from all preceding layers.

## GROWTH RATE

Network can be thinner and more compact. i.e number of channels can be fewer




## PARAMETER AND COMPUTER EFFICIENCY

DenseNet has a smaller size than ResNet

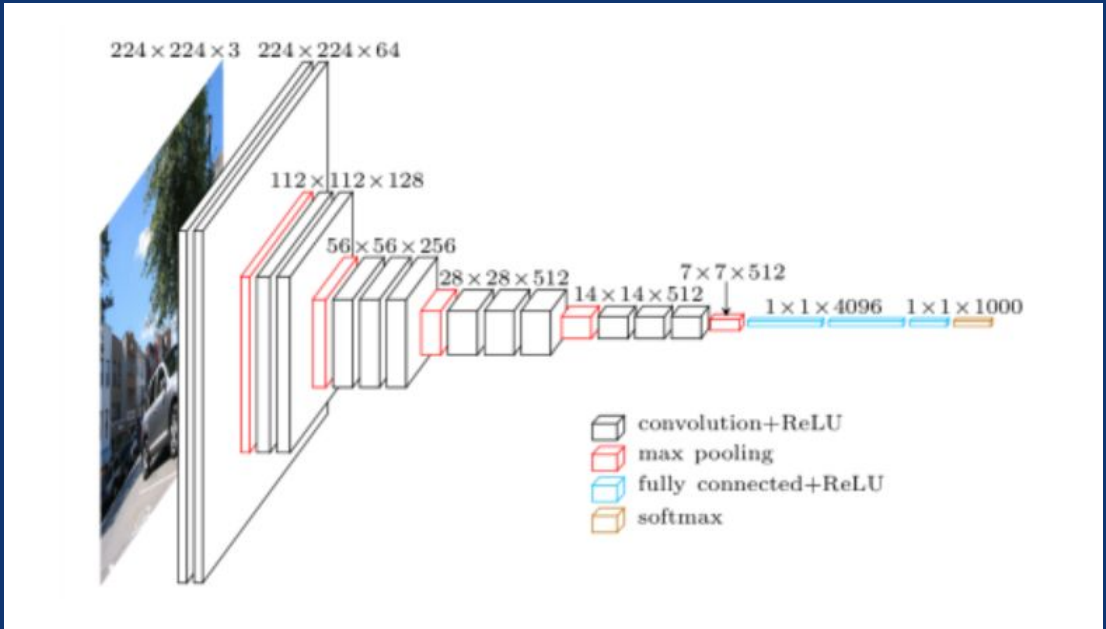
## STRONG GRADIENT FLOW

Error signal can be easily propagated to earlier layers more directly.

# VISUAL GEOMETRY GROUP (VGG) NET

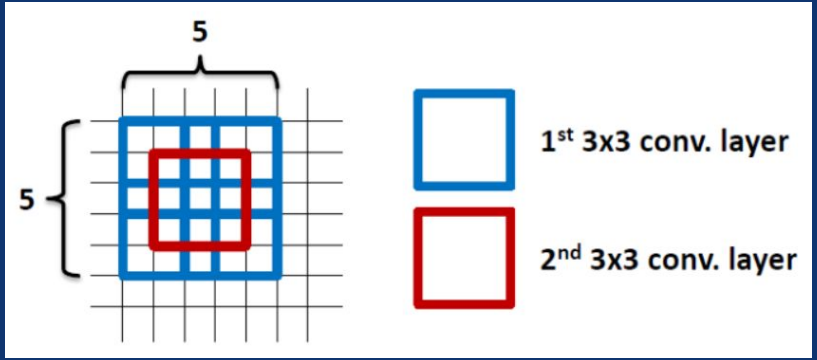
Karen Simonyan, Andrew Zisserman (2015), Very Deep Convolutional Networks for Large-Scale Image Recognition

# VGGNET




- uses only *3×3* convolutional layers stacked on top of each other in increasing depth
- reducing volume size is handled by max pooling
- two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier

# VGGNET




By using 2 layers of 3×3 filters, it actually have already covered 5×5 area. By using 3 layers of 3×3 filters, it actually have already covered 7×7 effective area.

⇩

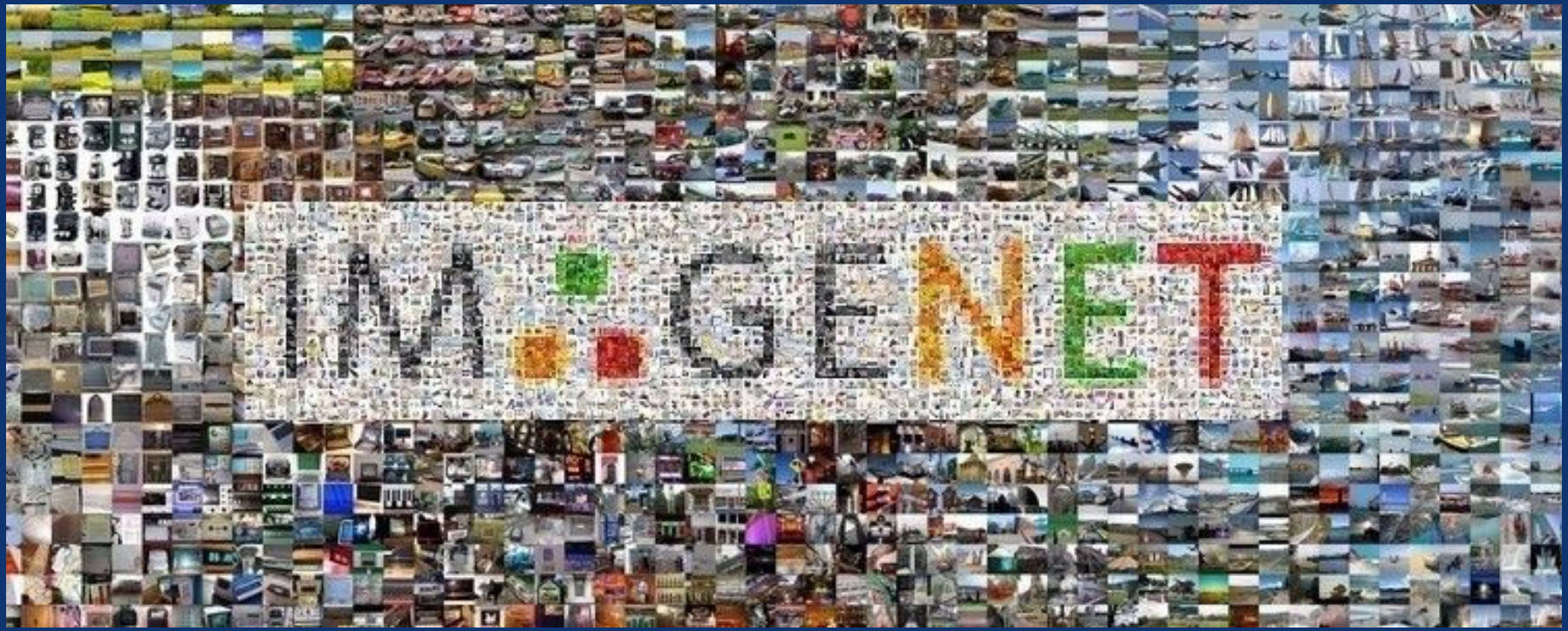
Number of parameter are fewer.

⇩

Faster convergence and reduce the overfitting problem.

# TRANSFER LEARNING

Kafeng Wang, Xitong Gao, Yiren Zhao, Xingjian Li, Dejing Dou, Cheng-Zhong Xu (2020), Pay Attention to Features, Transfer Learn Faster CNN


IMAGENET

# COMBINING ALL THREE

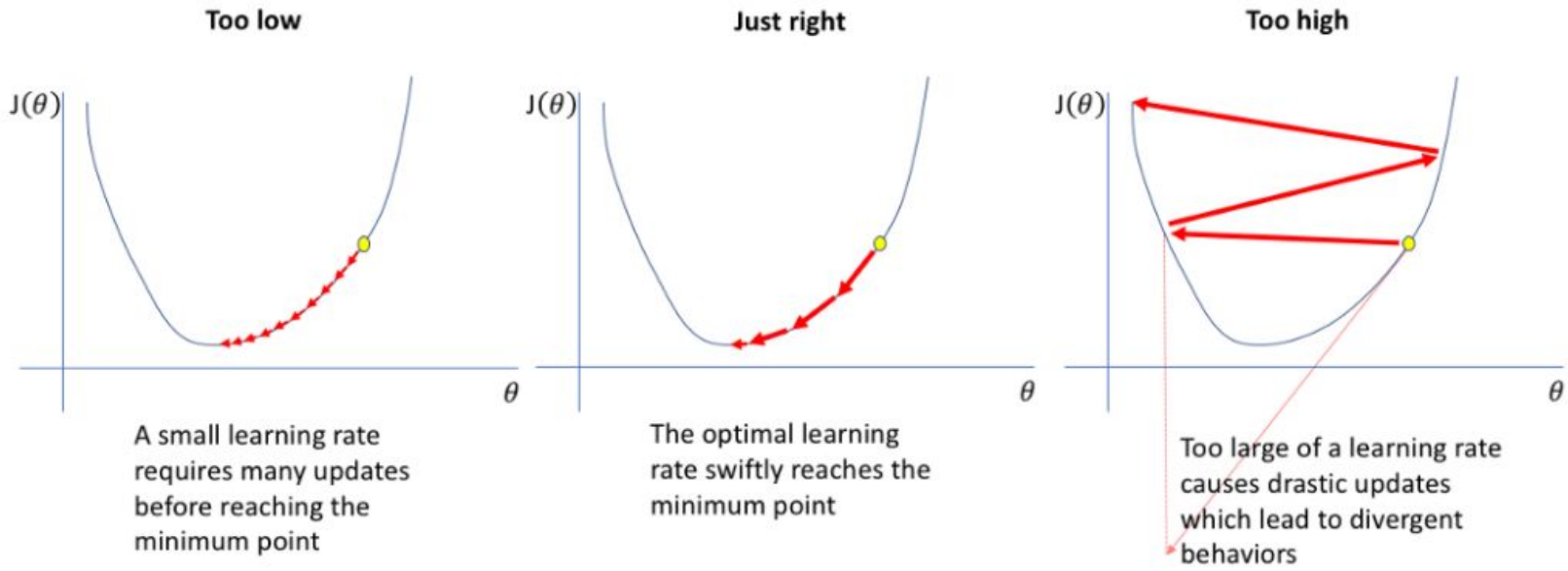
| Encoder Body | Max Accuracy (Validation) |
|---|---|
| res50 | 0.9959 |
| dense169 | 0.9956 |
| vgg19 | 0.9963 |
| all 3 combined | 0.9972 |

# WE GET A HIGH ACCURACY

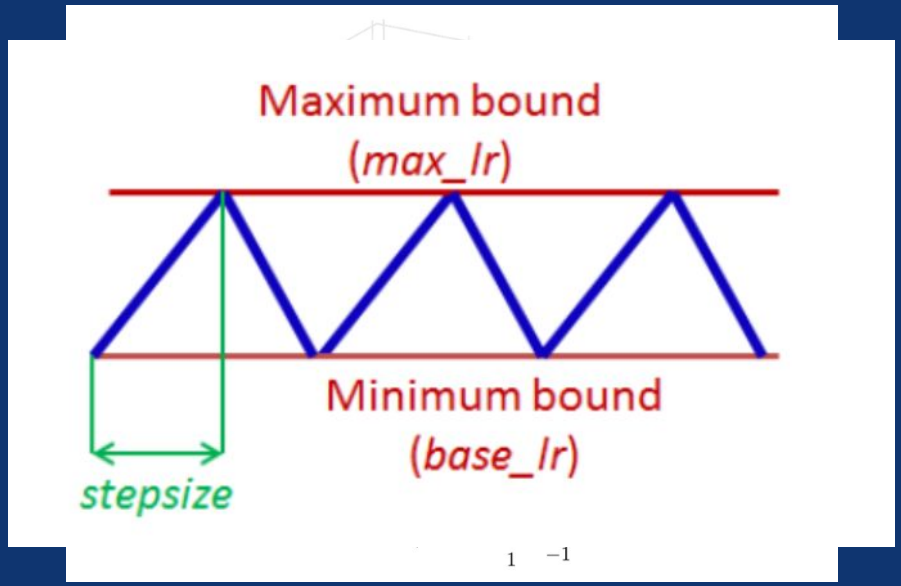
# ONE CYCLE POLICY FOR LEARNING RATE

Leslie N. Smith (2018), A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 -- Learning Rate, Batch Size, Momentum, and Weight Decay

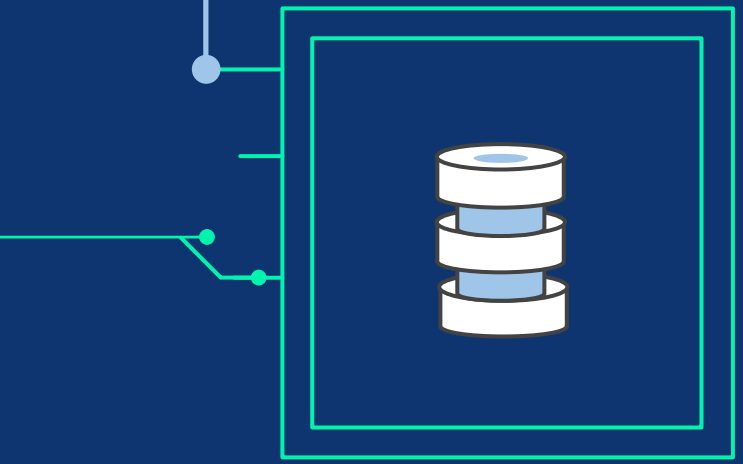
# EFFECTS OF LEARNING RATES ON TRAINING



# CYCLIC LEARNING RATES



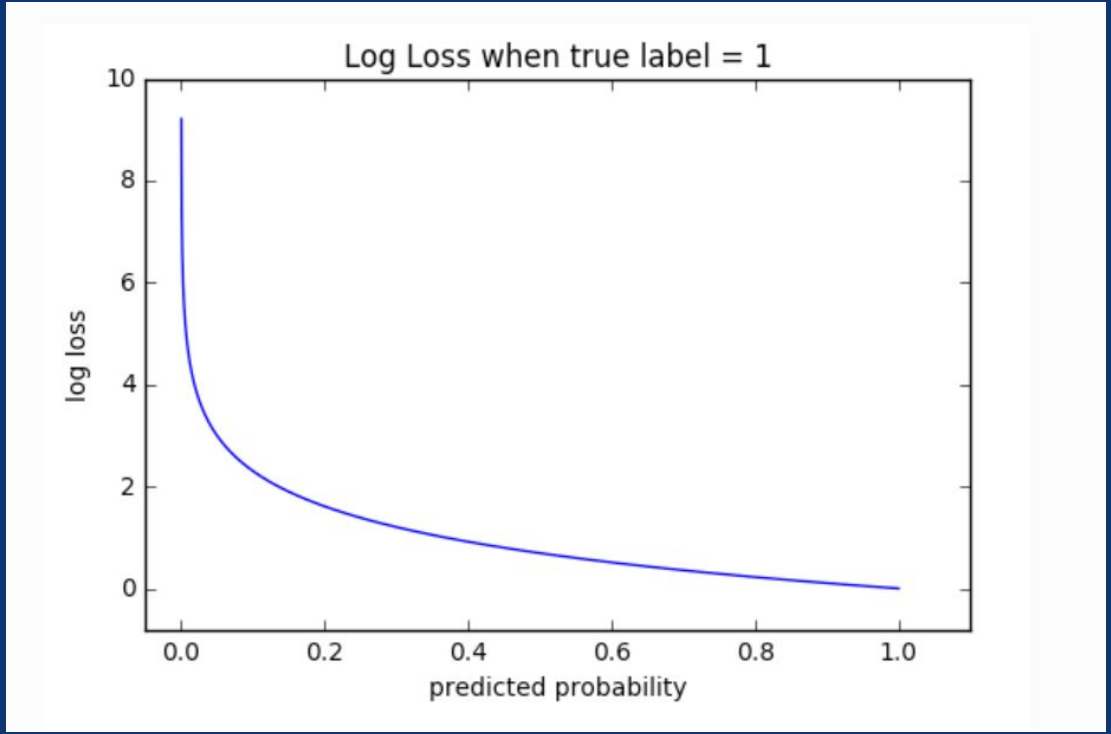

Cycle and Stepsize Saddle Point



# CROSS ENTROPY LOSS FUNCTION

Zhilu Zhang, Mert R. Sabuneu (2018),
Generalized Cross Entropy Loss for Training
Deep Neural Networks with Noisy Labels

# CROSS ENTROPY LOSS FUNCTION



# ADAM OPTIMIZER

Diederik P. Kingma, Jimmy Ba (2017), Adam: A Method for Stochastic Optimization

# HOW DOES ADAM WORK?

## 1) ADAPTIVE GRADIENT ALGORITHM (ADAGRAD)

that maintains a per-parameter learning rate that improves performance on problems with sparse gradients (eg. natural language and computer vision problems)

# HOW DOES ADAM WORK?

## 2) ROOT MEAN SQUARE PROPAGATION (RMSPROP)

that also maintains per-parameter learning rates that are adapted based on the average of recent magnitudes of the gradients for the weight (eg. how quickly it is changing). This means the algorithm does well online and non-stationary problems.

# BENEFITS OF USING ADAM

## STRAIGHTFORWARD TO IMPLEMENT

# BENEFITS OF USING ADAM

## COMPUTATIONALLY EFFICIENT

# BENEFITS OF USING ADAM

## LITTLE MEMORY REQUIREMENTS

# BENEFITS OF USING ADAM

## INVARIANT TO DIAGONAL RESCALE OF THE GRADIENTS

# BENEFITS OF USING ADAM

## WELL SUITED FOR PROBLEMS THAT ARE LARGE IN TERMS OF DATA OR PARAMETERS

## BENEFITS OF USING ADAM

# HYPER-PARAMETERS HAVE INTUITIVE INTERPRETATION AND TYPICALLY REQUIRE LITTLE TUNING

# OUR BUSINESS IDEA

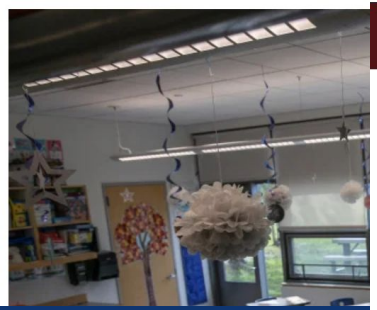
# COVID-19 PANDEMIC

TIME FOR NEW BEGINNINGS. GET 2 MONTHS FOR ONLY $8
SUBSCRIBE NOW



U.S. • COVID-19

# 'Where Are the Kids?' School Is Back in Session, but Many Kindergarteners Are Missing



By Chew Hui Min

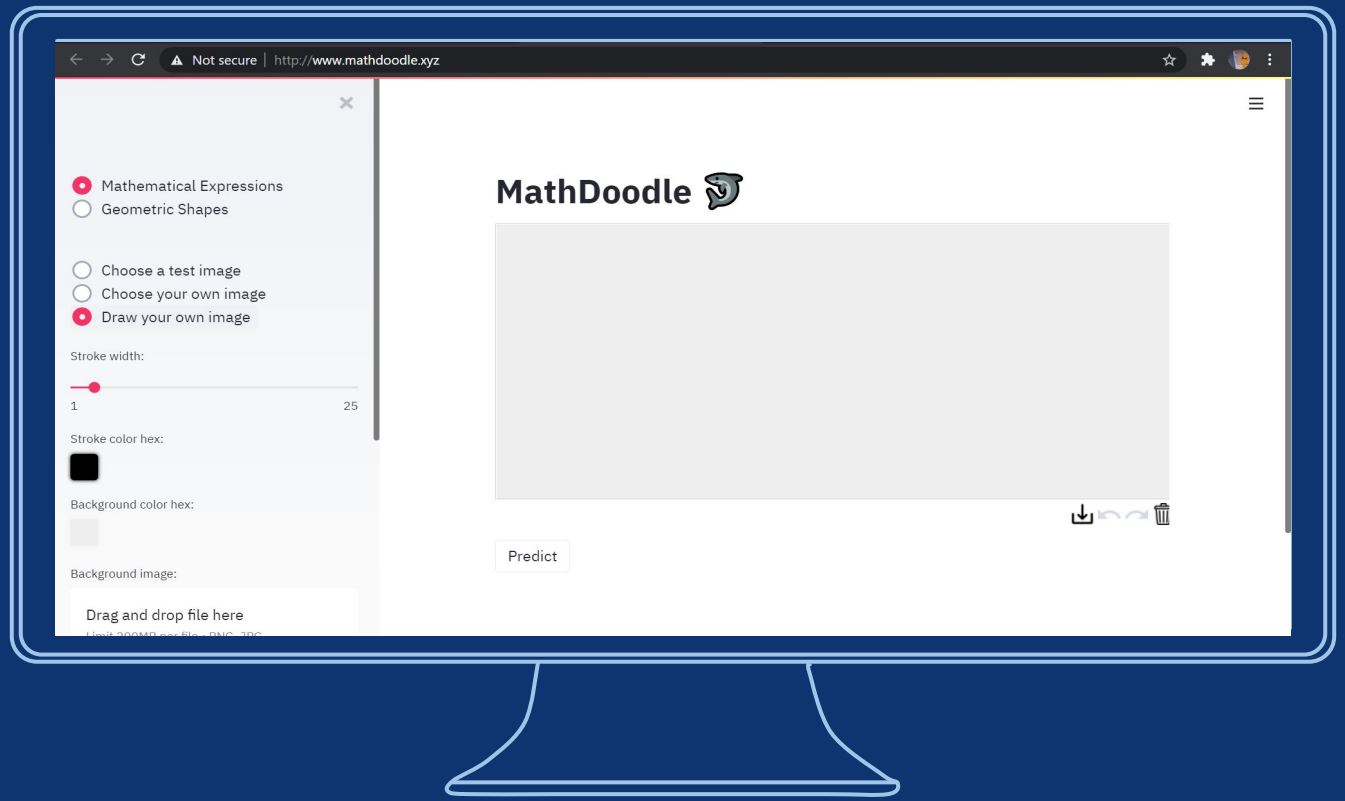
Singapore

# 'Right time' to close schools now, says Education Minister Ong Ye Kung

03 Apr 2020 05:18PM
(Updated: 03 Apr 2020 09:38PM)

Bookmark

# MATH DOODLE


Not secure | http://www.mathdoodle.xyz
Mathematical Expressions
Geometric Shapes
Choose a test image
Choose your own image
Draw your own image
Stroke width:
1
25
Stroke color hex:
Background color hex:
Background image:
Drag and drop file here
MathDoodle
Predict