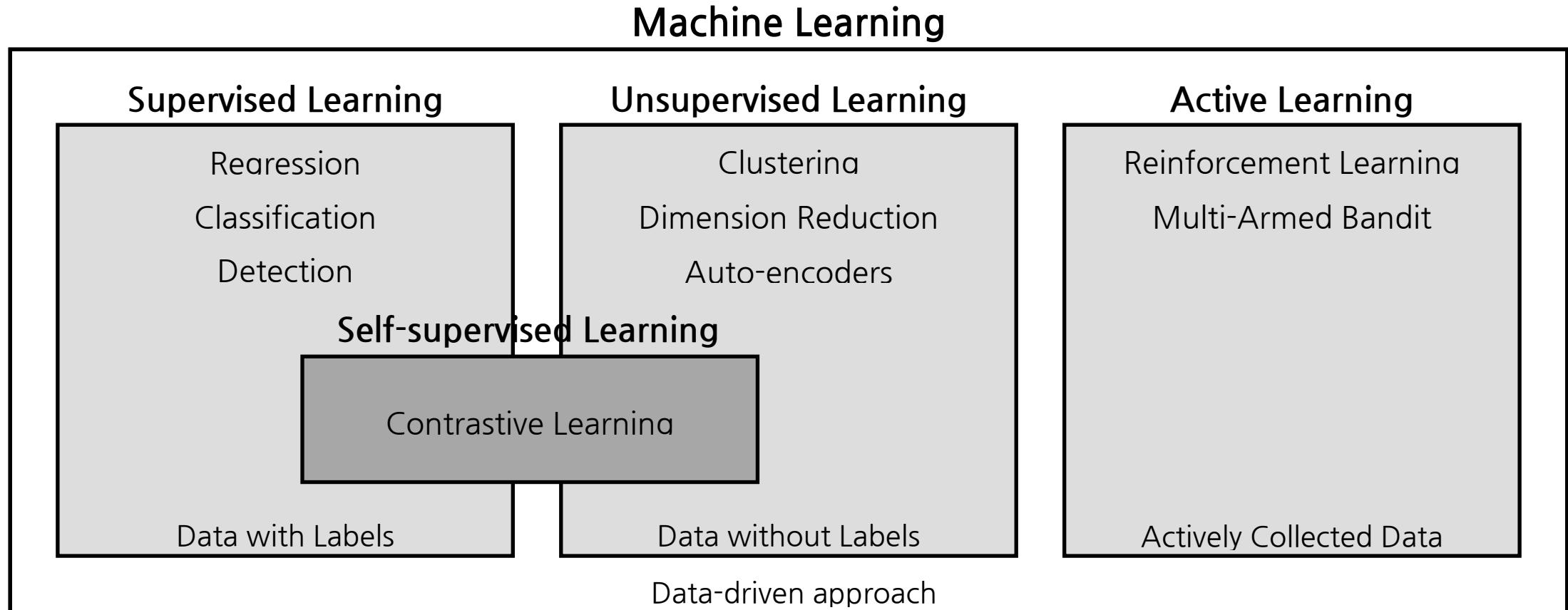
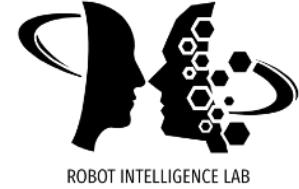




Historical Review of Deep Learning

Sungjoon Choi, Korea University

Introduction





Introduction

- Key Components of Deep Learning
 - The data that we can learn from.
 - A model of how to transform the data.
 - A loss function that quantifies the badness of the model.
 - An algorithm to adjust the parameters to minimize the loss.

Data

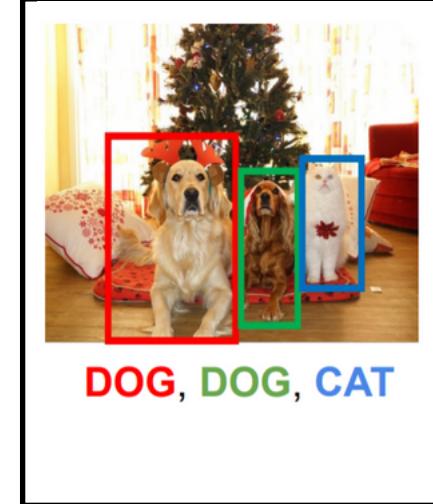
Classification



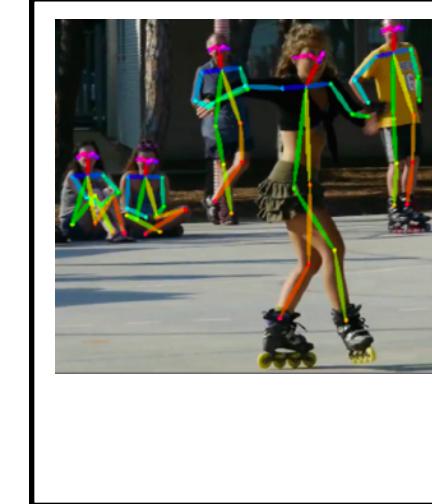
Semantic Segmentation



Detection



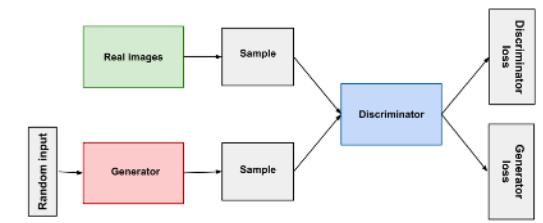
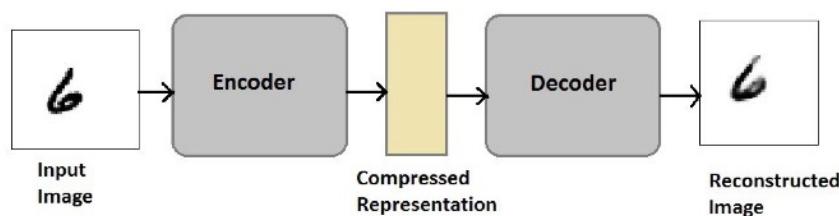
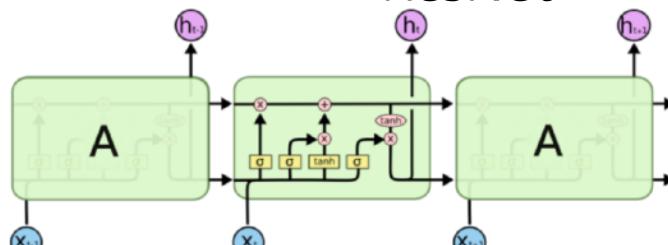
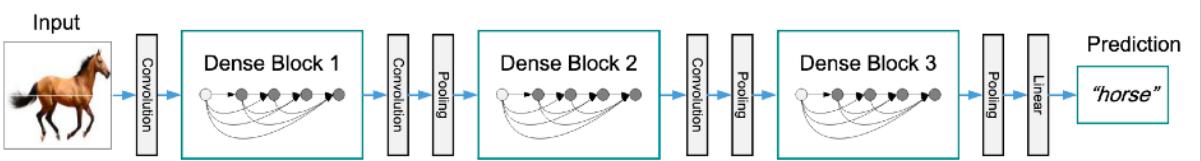
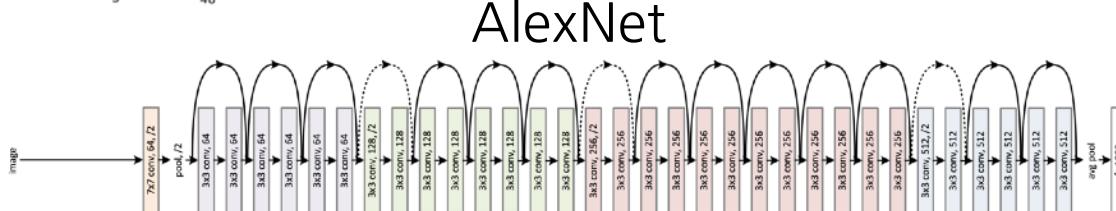
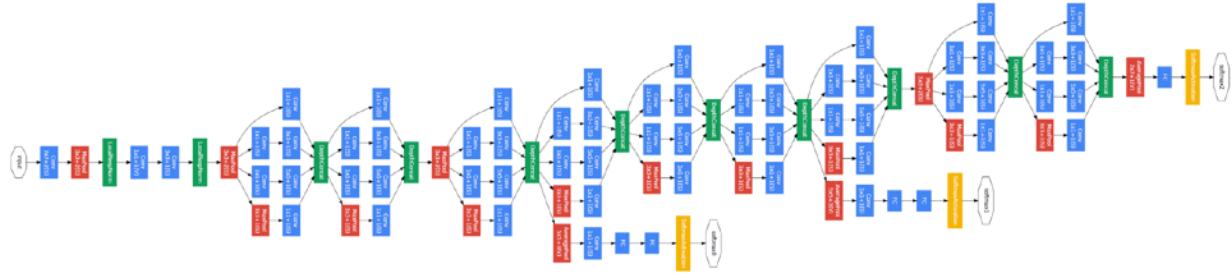
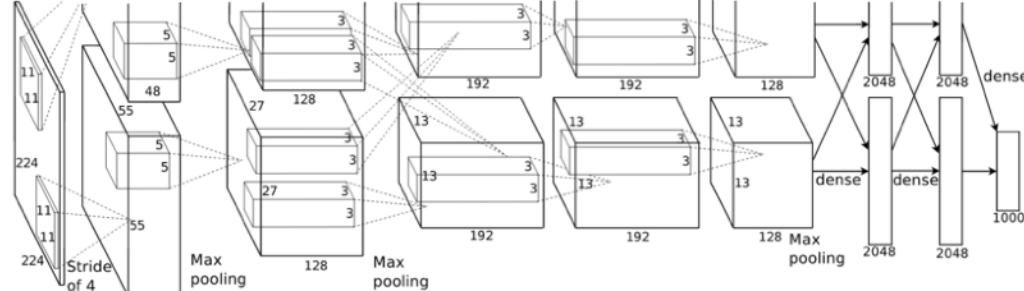
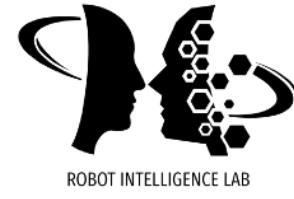
Pose Estimation



Visual QnA



Model Architectures



Loss



Regression Task

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D (y_i^{(d)} - \hat{y}_i^{(d)})^2$$

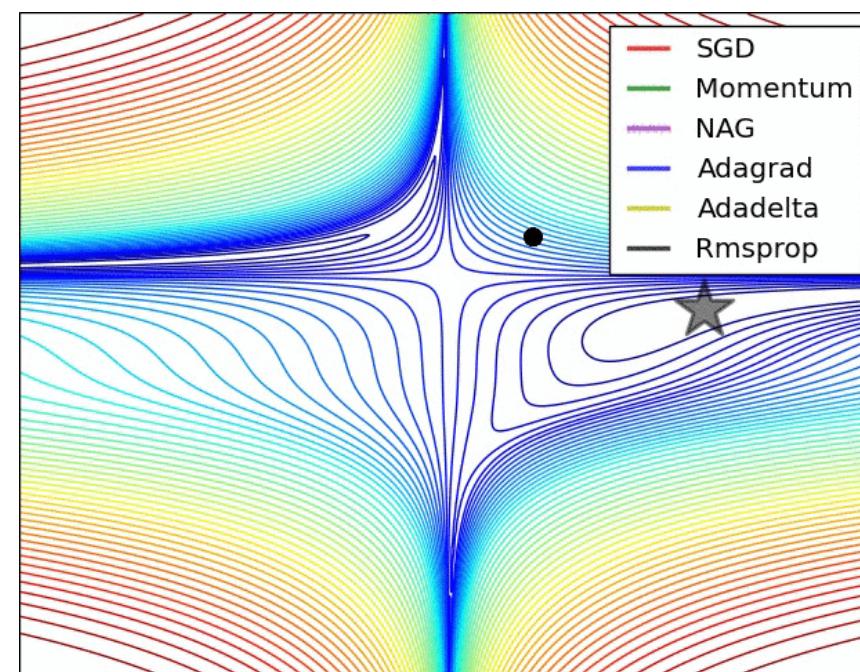
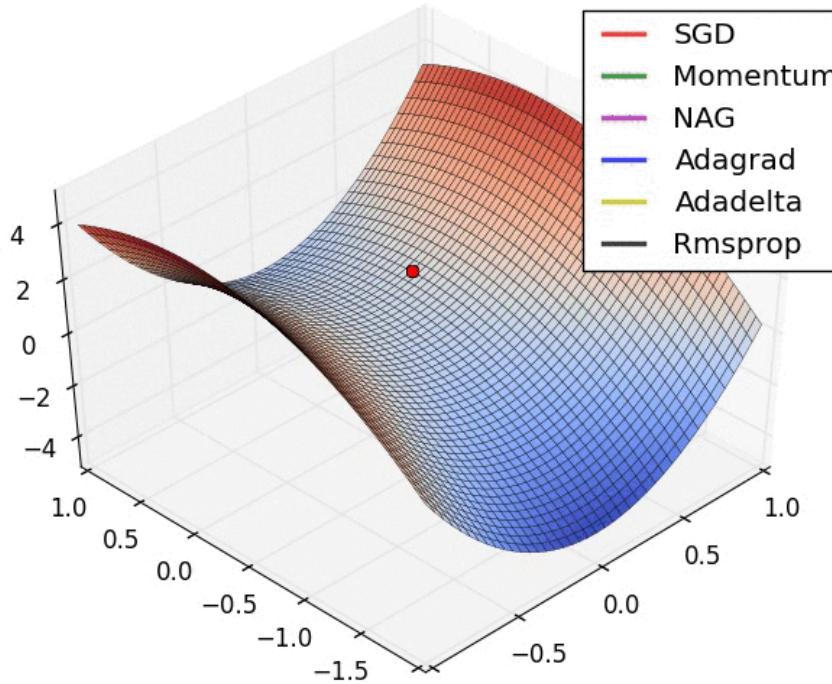
Classification Task

$$\text{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D y_i^{(d)} \log \hat{y}_i^{(d)}$$

Probabilistic Task

$$\text{MLE} = \frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D \log \mathcal{N}(y_i^{(d)}; \hat{y}_i^{(d)}, 1) \quad (= \text{MSE})$$

Algorithm



Dropout
Early stopping
k-fold validation
Weight decay
Batch normalization
MixUp
Ensemble
Bayesian Optimization



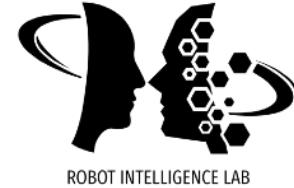
Historical Review

Deep Learning's Most Important Ideas - A Brief
Historical Review

Denny Britz

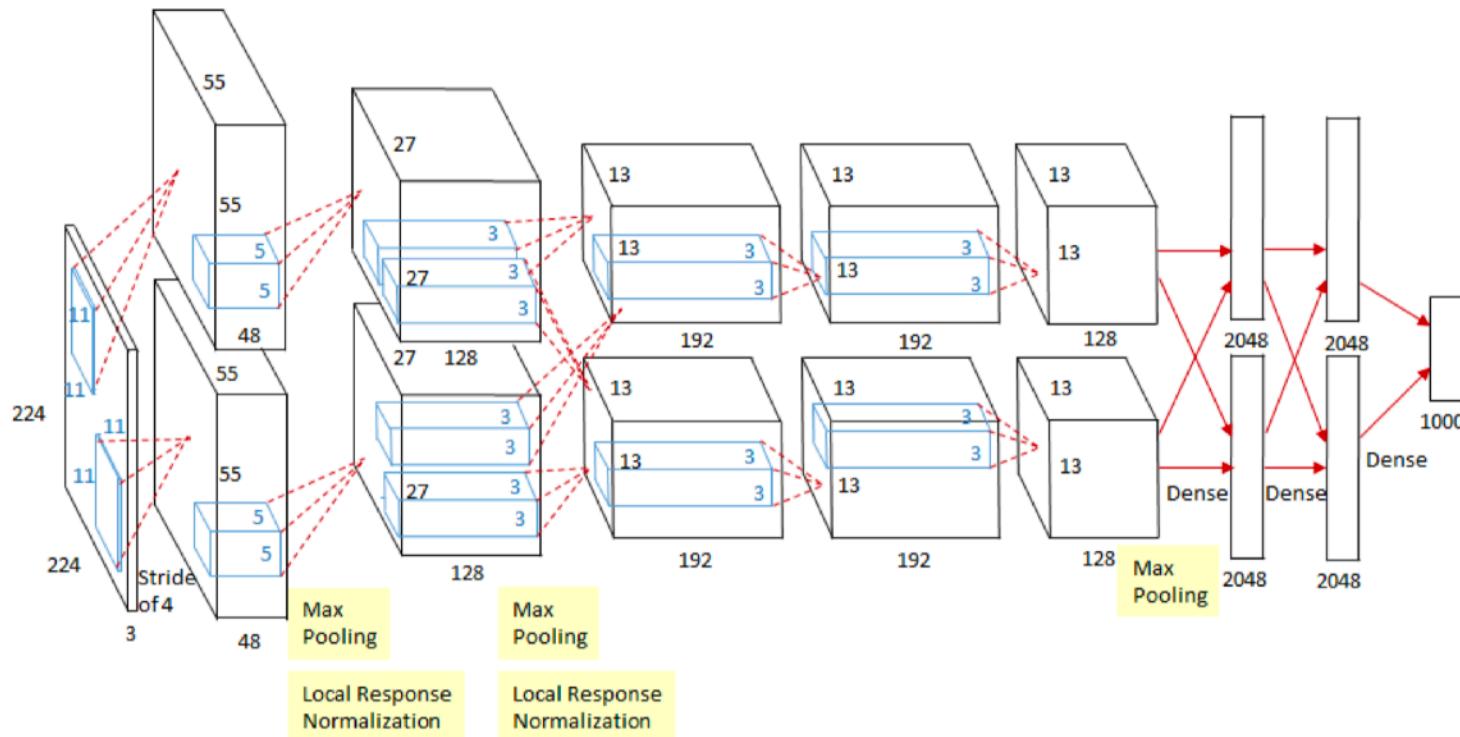
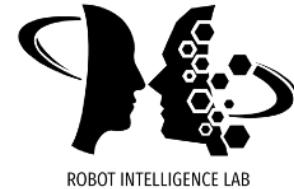
2020-07-29

Historical Review

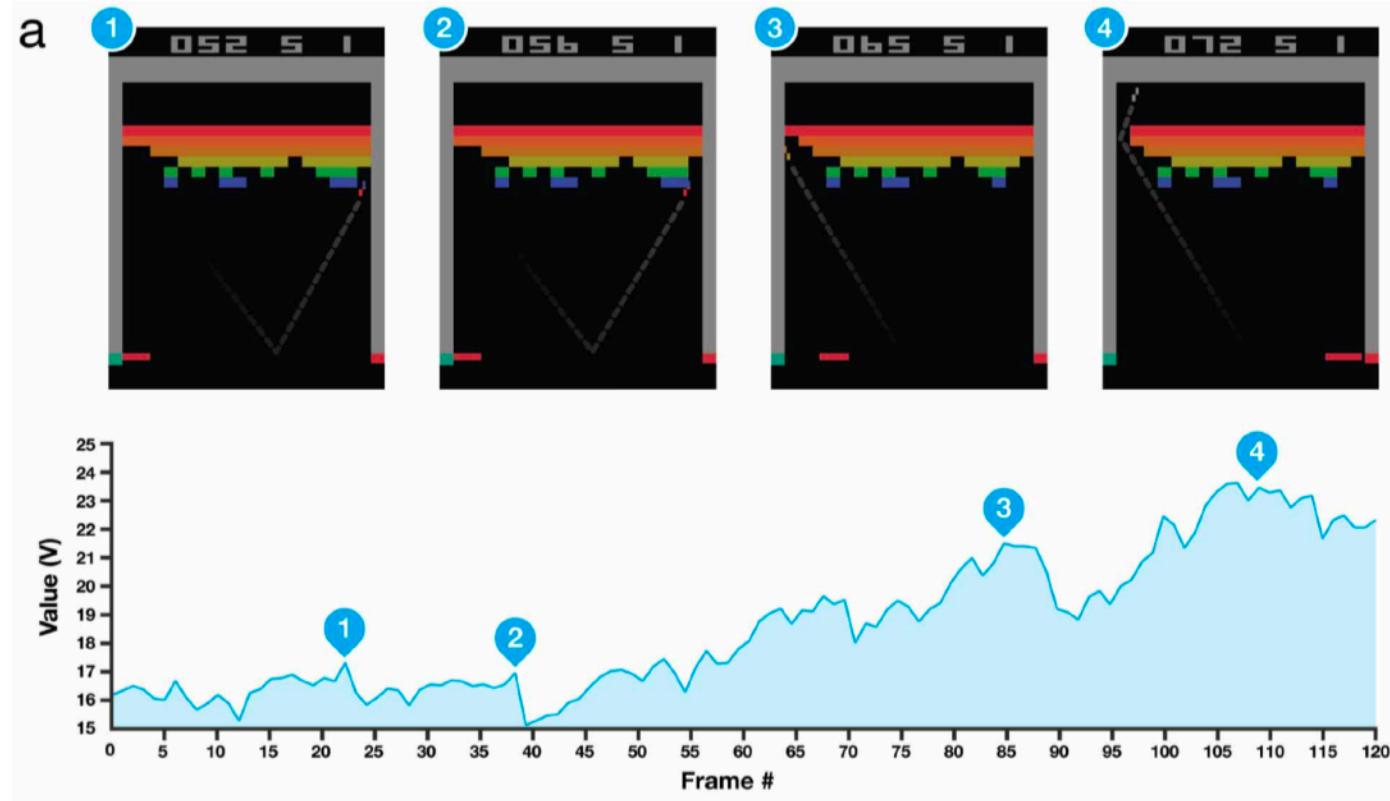
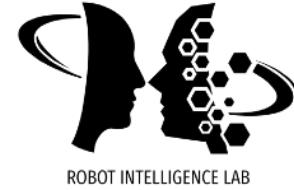


The goal of this post is to review those ideas that have stood the **test of time**.

2012 - AlexNet



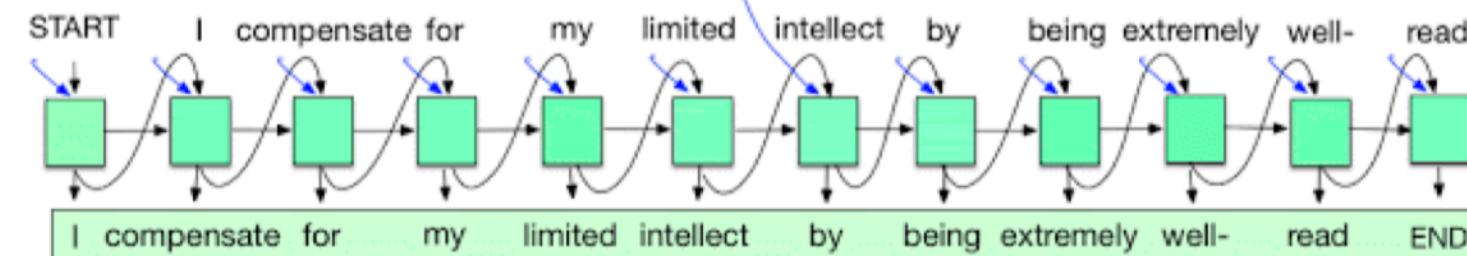
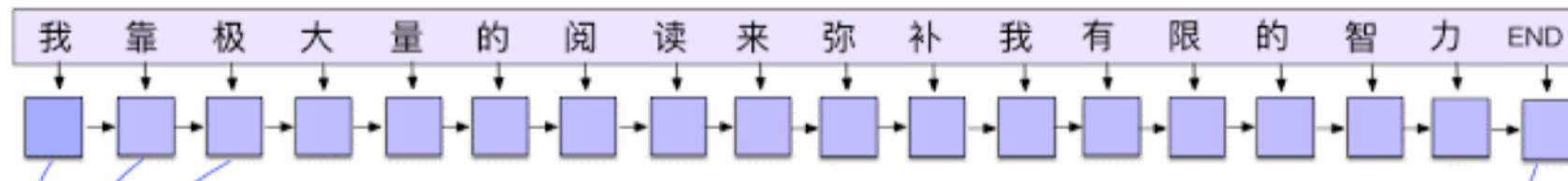
2013 - DQN



2014 - Encoder-Decoder with Attention

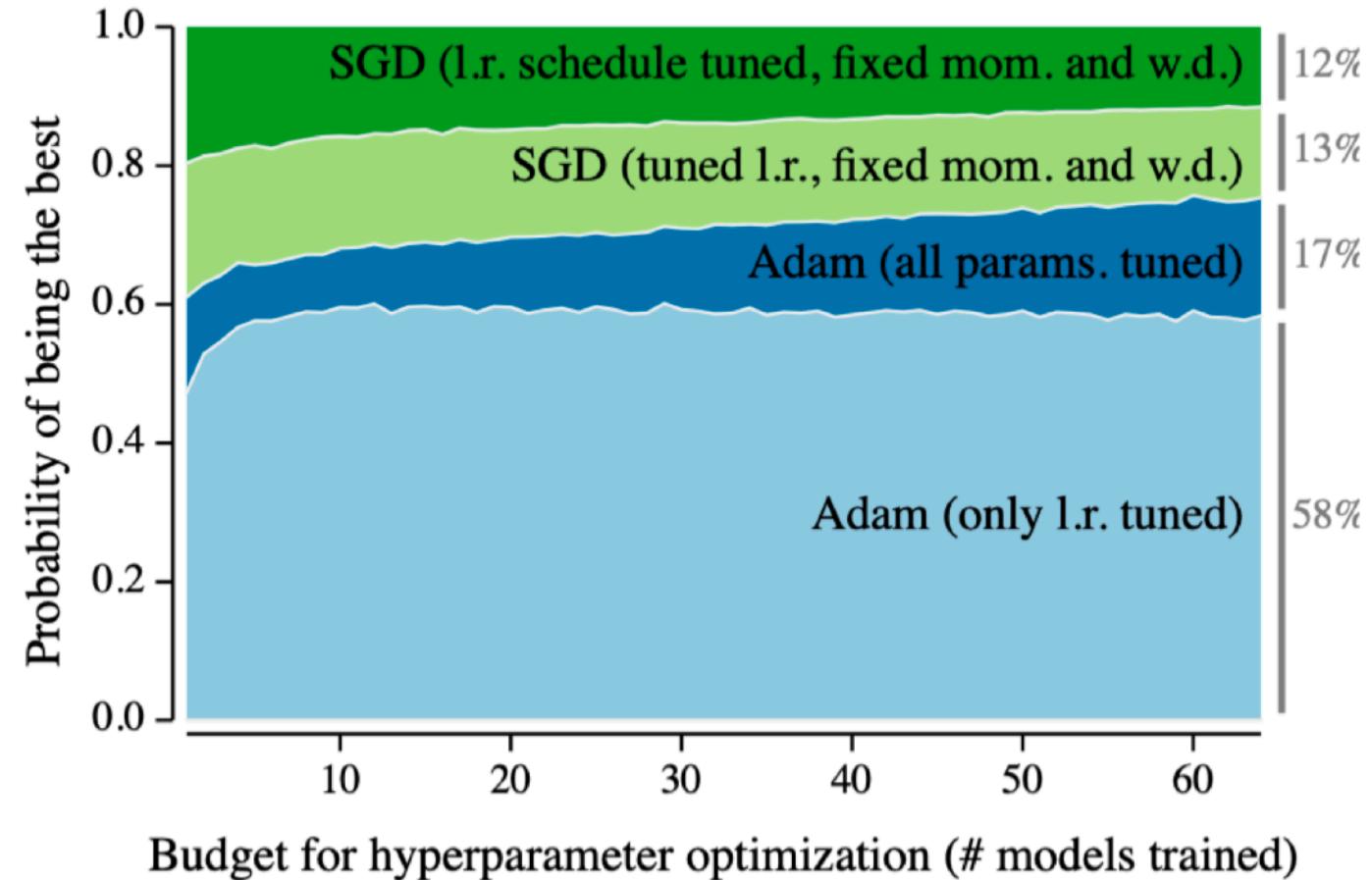


ENCODER

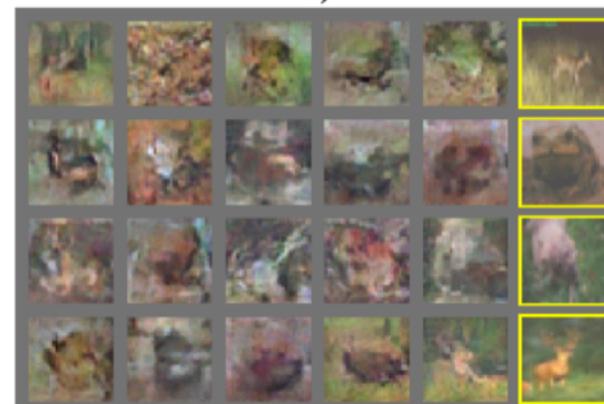
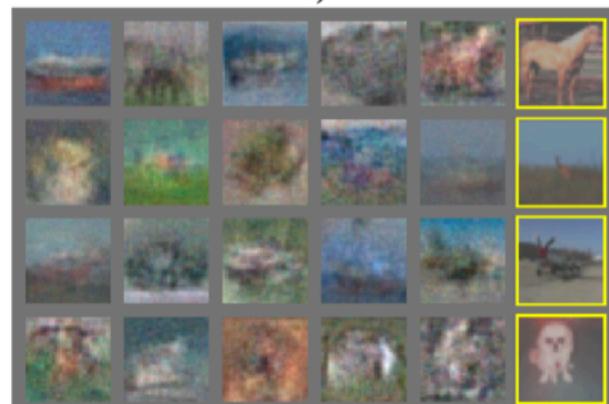
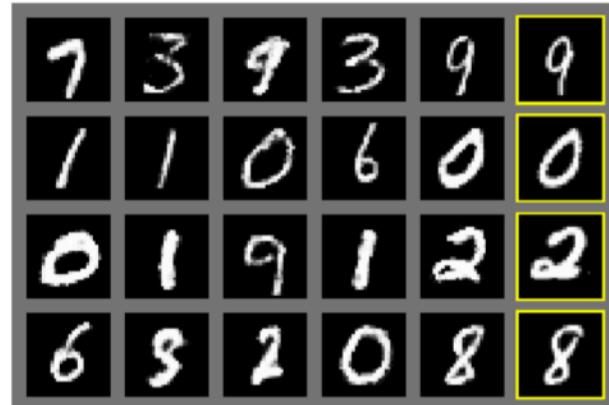


DECODER

2014 - Adam



2015 - GAN

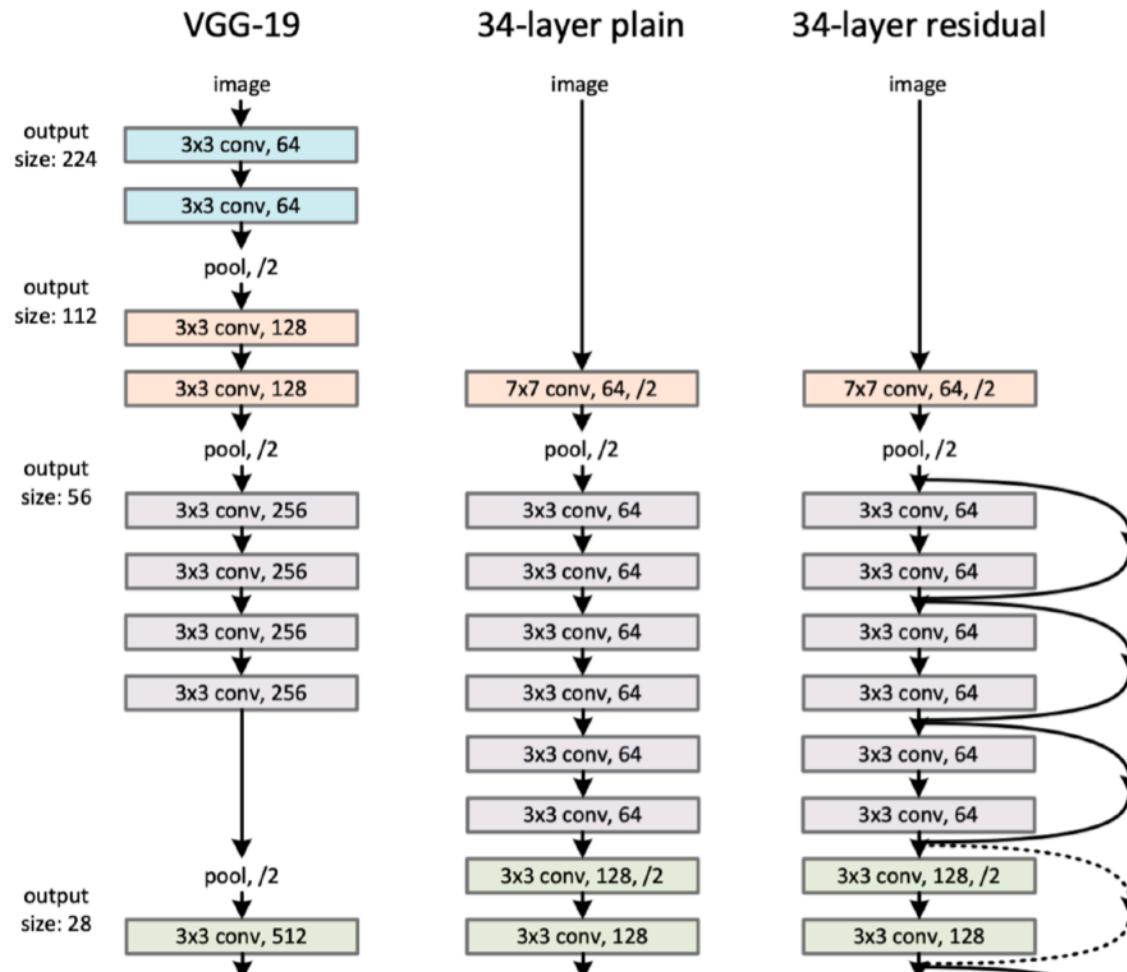
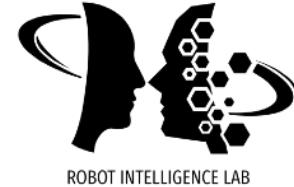


2015 - GAN



“Finally, we would like to thank Les Trois Brasseurs for stimulating our creativity.”

2015 - ResNet



2017 - Transformer



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

2017 - Transformer

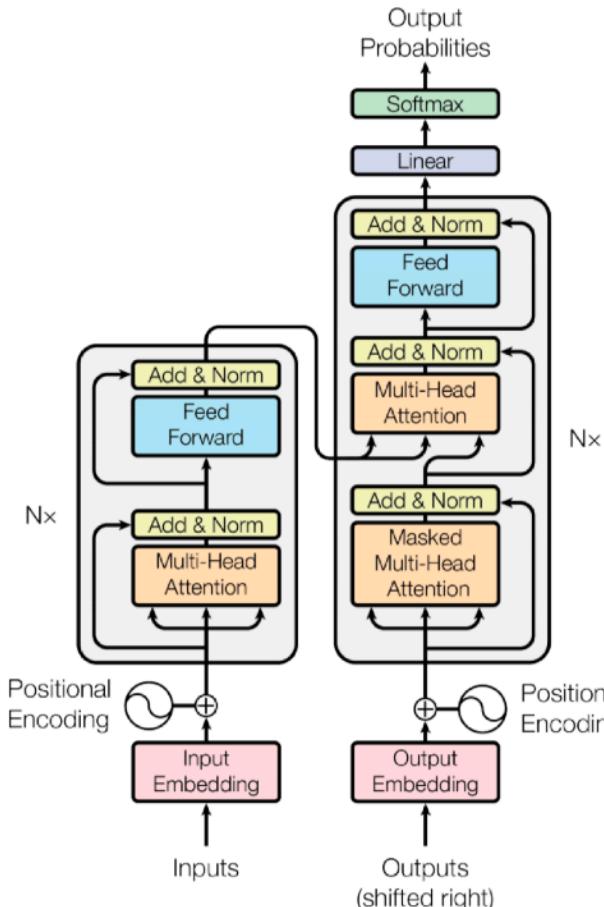


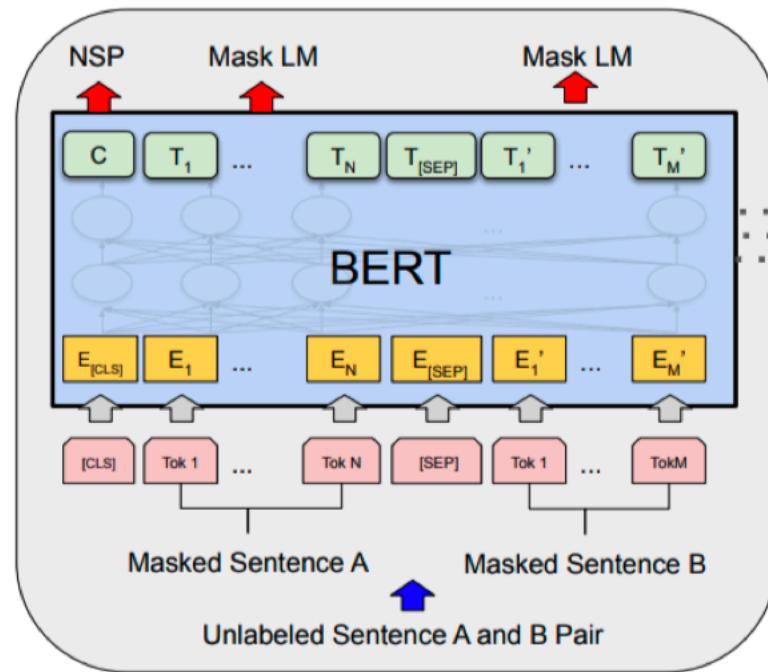
Figure 1: The Transformer - model architecture.

Single Head Self-Attention Calculation

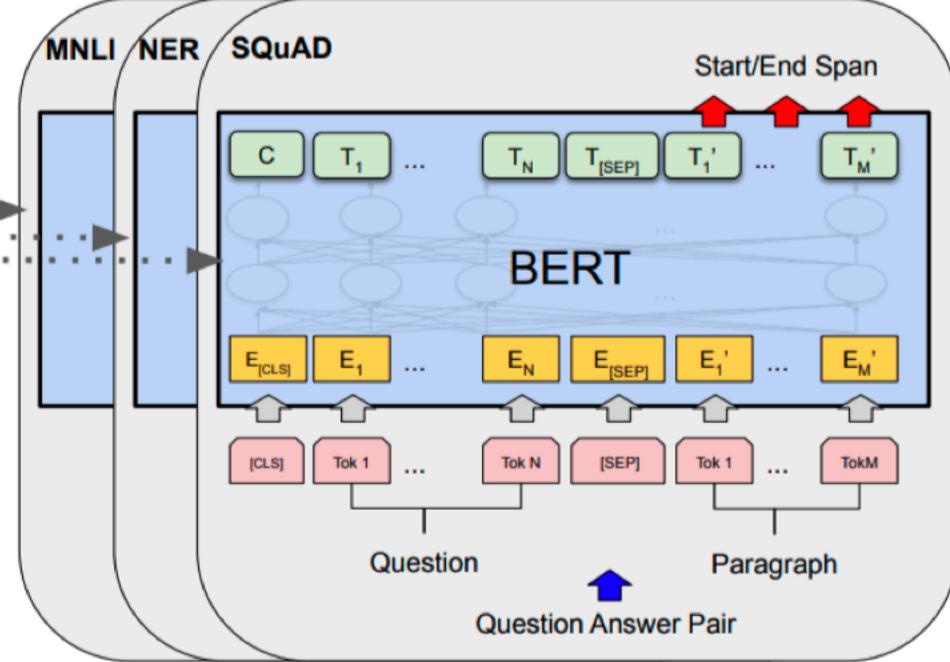
$$\text{softmax} \left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_k}} \right) = \mathbf{Z}$$

The diagram shows the calculation of a single head of self-attention. It consists of three matrices: \mathbf{Q} (purple 3x3 grid), \mathbf{K}^T (orange 3x3 grid), and \mathbf{V} (blue 3x3 grid). The \mathbf{Q} and \mathbf{K}^T matrices are multiplied together, and the result is divided by the square root of the dimension d_k . The final result is a pink 3x3 grid labeled \mathbf{Z} .

2018 - BERT

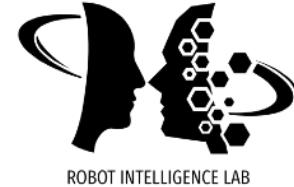


Pre-training



Fine-Tuning

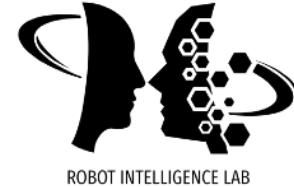
2019 - Pre-trained Language Model



OpenAI

GPT-3, an autoregressive language model with 175 billion parameters

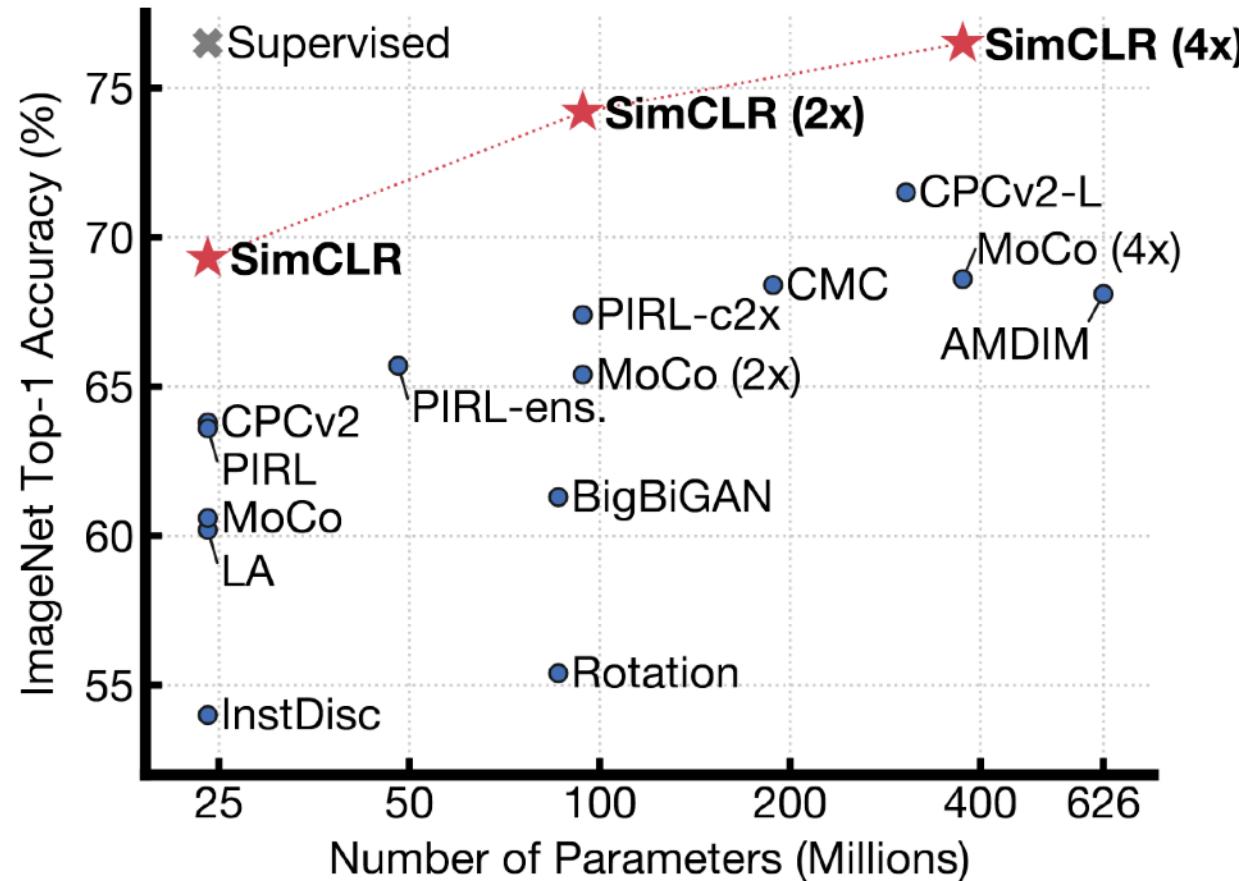
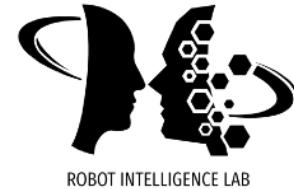
2020~ - Self Supervised Learning



A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen¹ Simon Kornblith¹ Mohammad Norouzi¹ Geoffrey Hinton¹

2020~ - Self Supervised Learning



Thank You



ROBOT INTELLIGENCE LAB