Self-Supervised Learning for Vision

BIODS 271 / CS 277

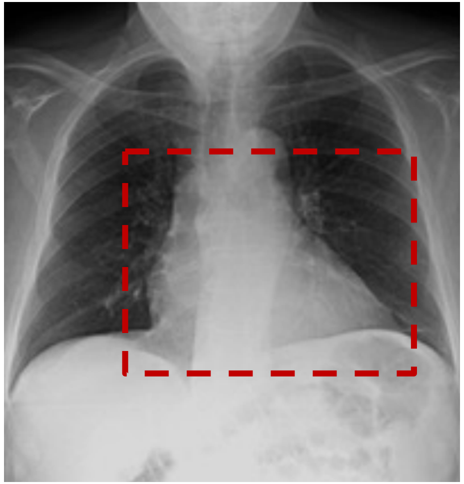
Akshay Chaudhari

Data Enables Solving Medical Problems

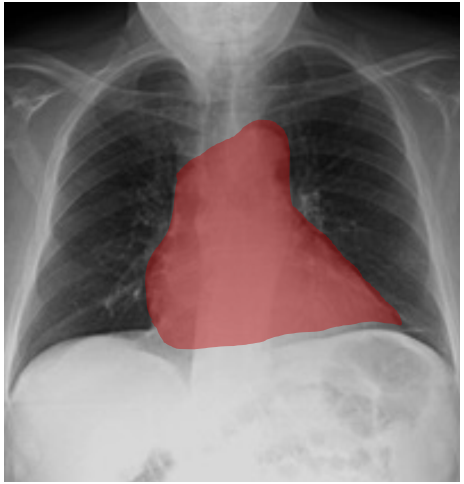
Classification

✓: Cardiomegaly

Detection



Segmentation



Regression

Ejection Fraction: **49%**

Supervised Learning

IM = Image
L = Label

IM1	IM2	IM3
IM4	IM5	IM6
IM7	IM8	IM9

+

L1	L2	L3
L4	L5	L6
L7	L8	L9

= Model!

( , Dog) ( , Cat) ( , Hot Dog)

DATA

Inputs

and so on…

Labels

Dog

Cat

Hot Dog

…

Label-Starved Learning

IM = Image
L = Label

IM1	IM2	IM3
IM4	IM5	IM6
IM7	IM8	IM9

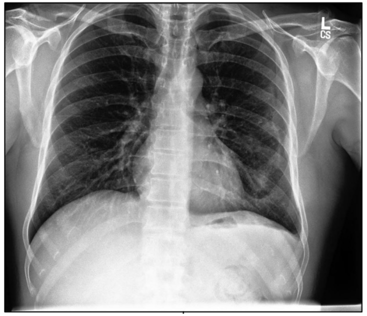
+

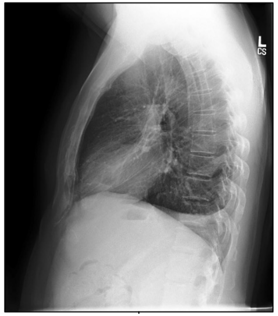
~~L1~~	~~L2~~	~~L3~~
~~L4~~	~~L5~~	~~L6~~
~~L7~~	~~L8~~	~~L9~~

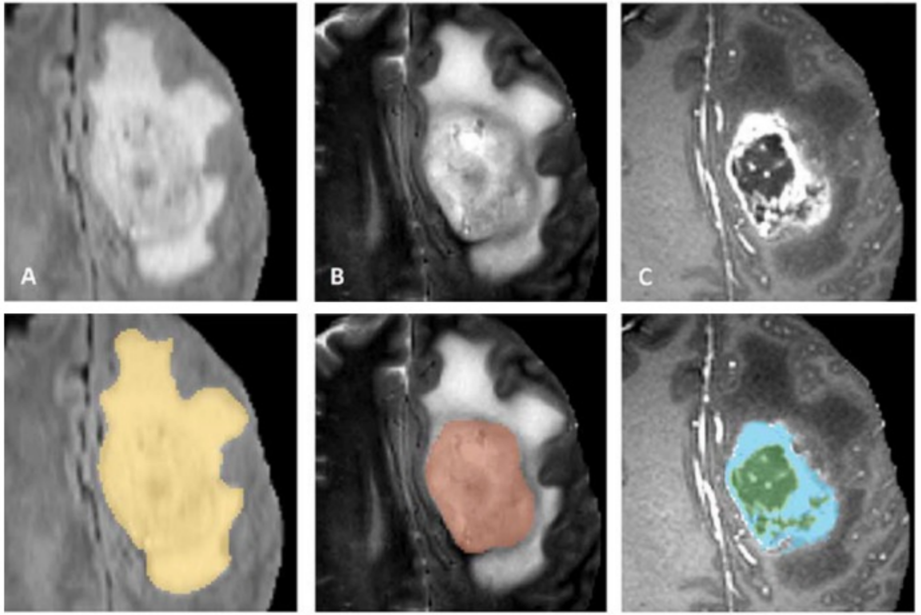
= **No Model!**

( , **?**) ( , **?**) ( , **?**)

Example Labels in Medical Imaging

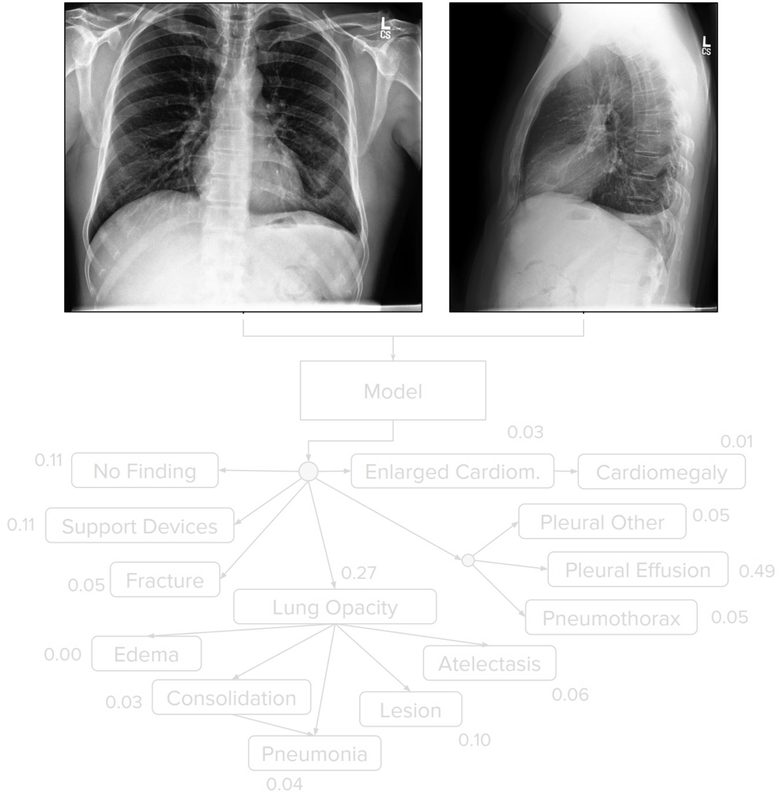


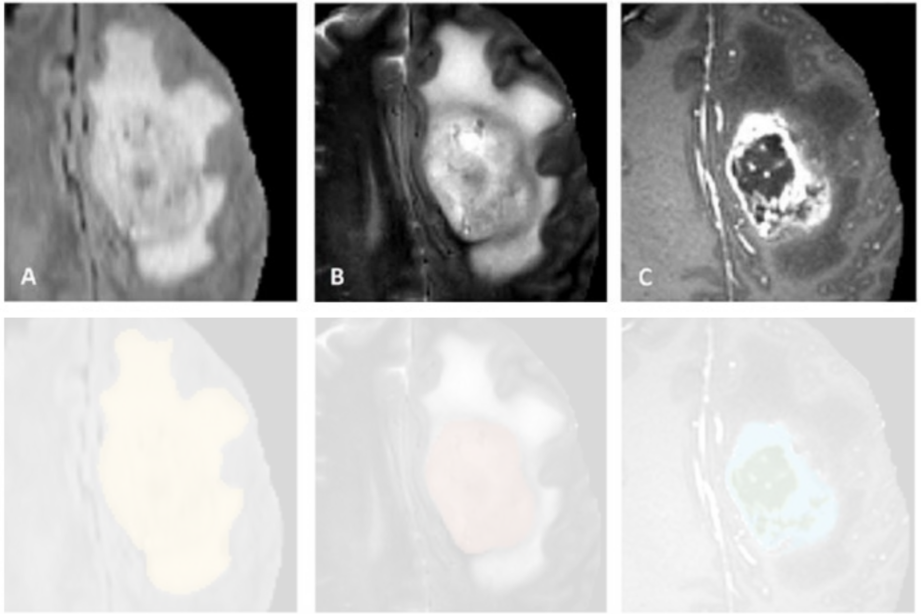




[1] Irwin et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. AAAI. 2019
[2] Menze et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE TMI. 2015

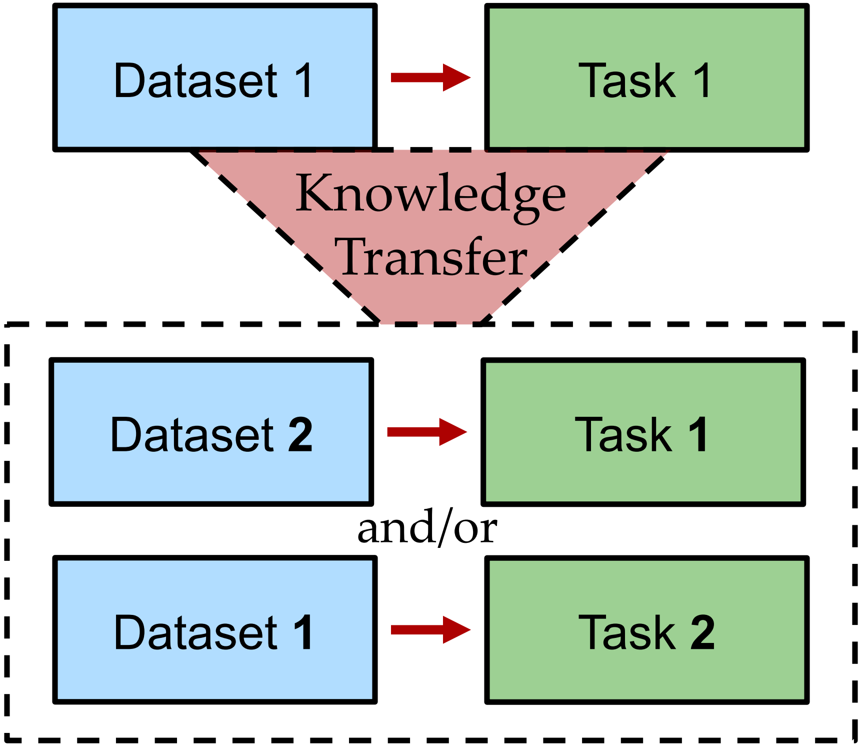
Example Labels in Medical Imaging





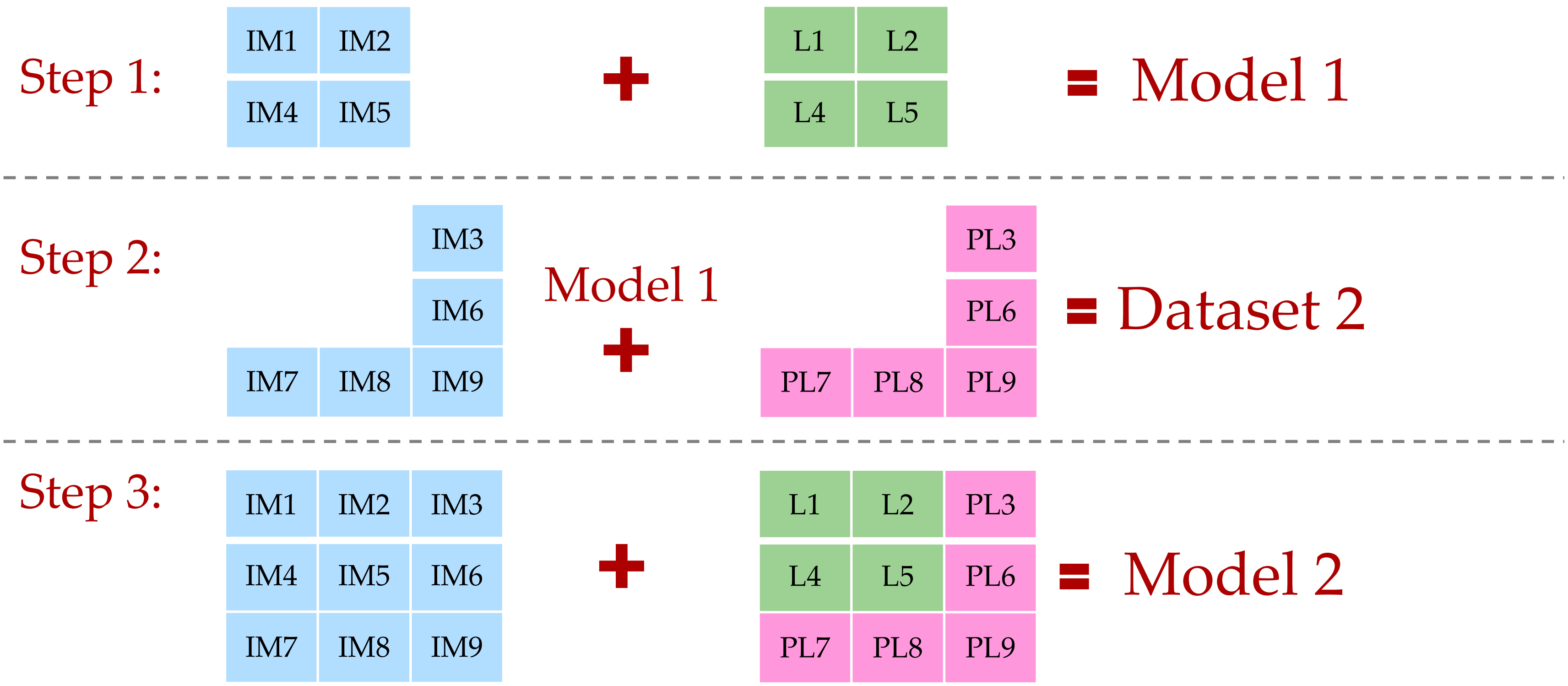
[1] Irwin et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. AAAI. 2019
[2] Menze et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE TMI. 2015

Transfer Learning



Beyond Supervised Learning

- Semi-Supervised Learning



Semi-Supervised Learning: Self-Training

[1] Xie et al. Self-training with Noisy Student improves ImageNet classification. CVPR 2020.

Self-Training Terminology

- **Teacher Network:** Network that creates pseudo labels
- **Student Network:** Network that learns using pseudo labels

Terminology

- **Self-Training:** When student network is same/larger sized than teacher network

- **Knowledge Distillation:** When student network is smaller than teacher network

Noisy Student Self Training

ImageNet (**14M** examples)

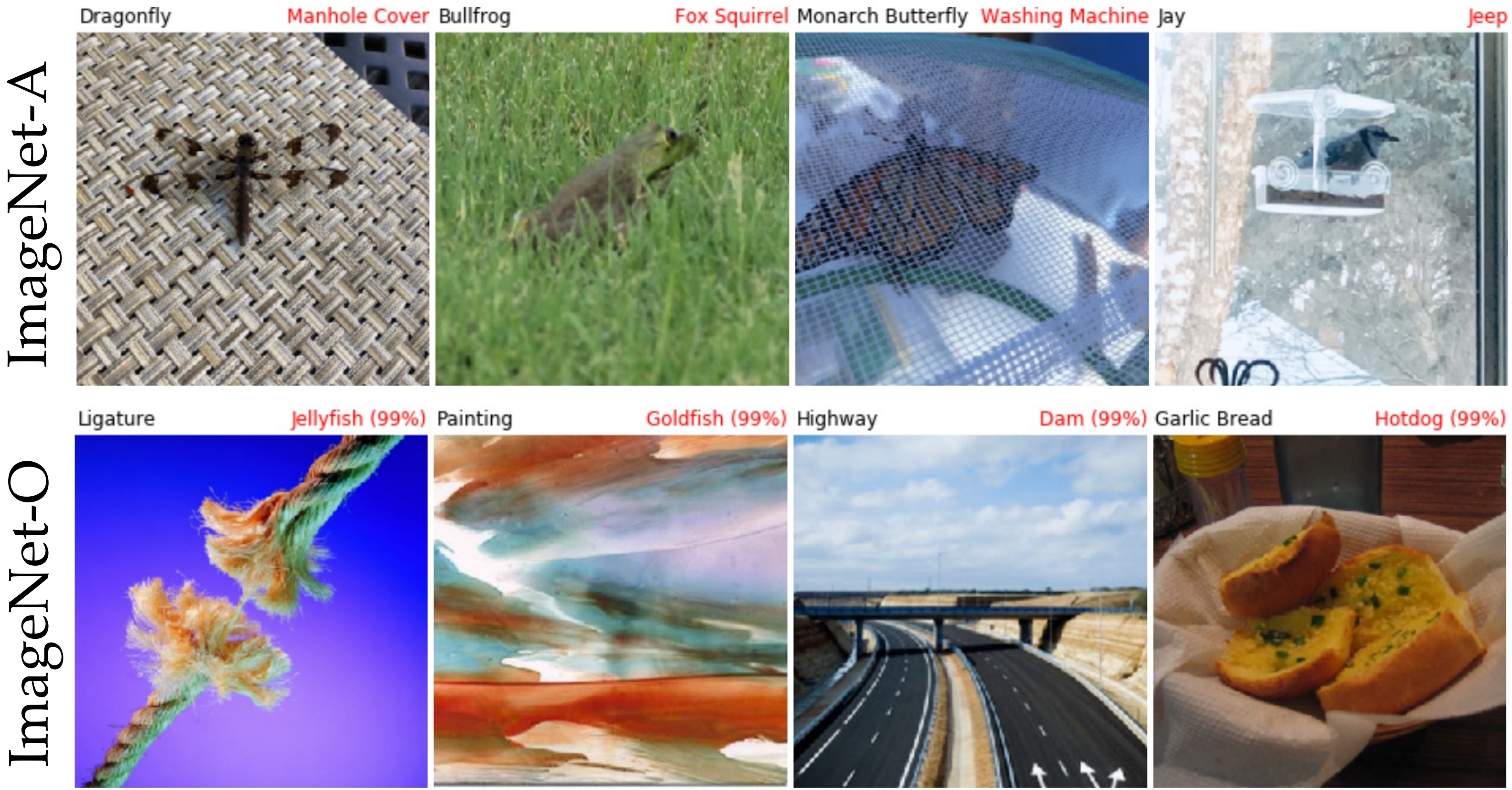
JFT (**300M** examples)

	ImageNet top-1 acc.	ImageNet-A top-1 acc.
Prev. SOTA	86.4%	61.0%
Ours	**88.4%**	**83.7%**

Large accuracy gain for **adversarial** examples

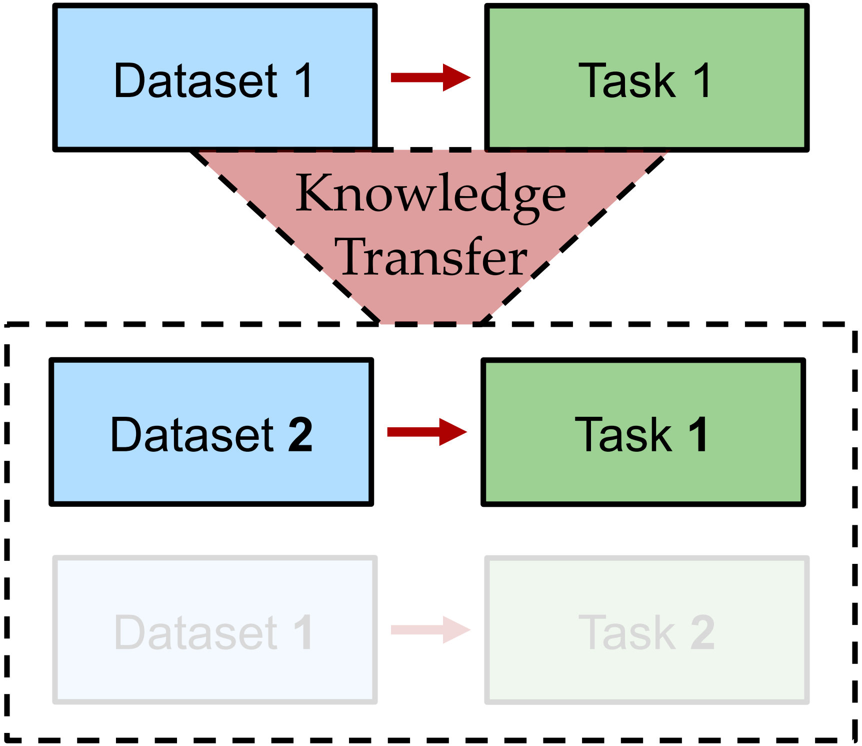
[1] Xie et al. Self-training with Noisy Student improves ImageNet classification. CVPR 2020.

Adversarial Examples



[1] Hendrycks et al. Natural Adversarial Examples. CVPR 2021.

Semi-Supervised Transfer Learning



Self- Supervised Learning

IM = Image
PL = Pseudolabel
L = Label

Step 1:

IM1	IM2	IM3
IM4	IM5	IM6
IM7	IM8	IM9

+

PL1	PL2	PL3
PL4	PL5	PL6
PL7	PL8	PL9

= Pretrained Model

Step 2:

IM1	IM2
IM3	IM4

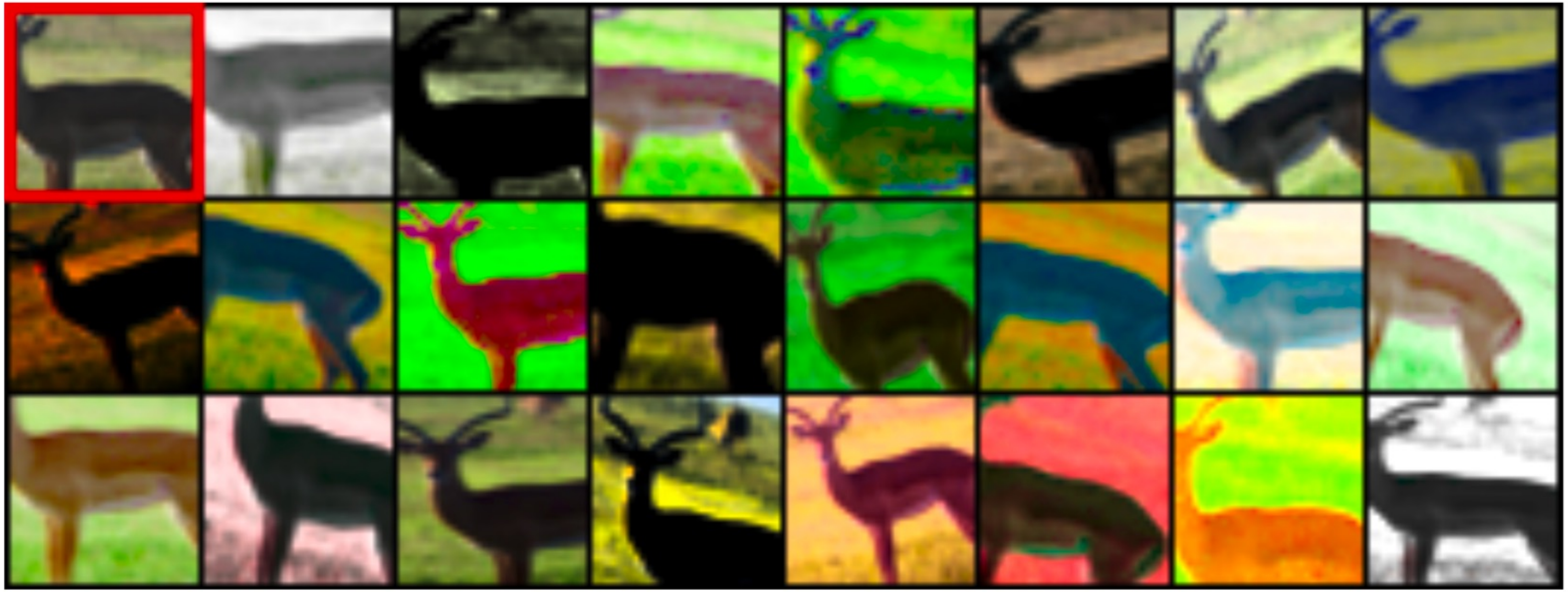
+

L1	L2
L3	L4

= Model!

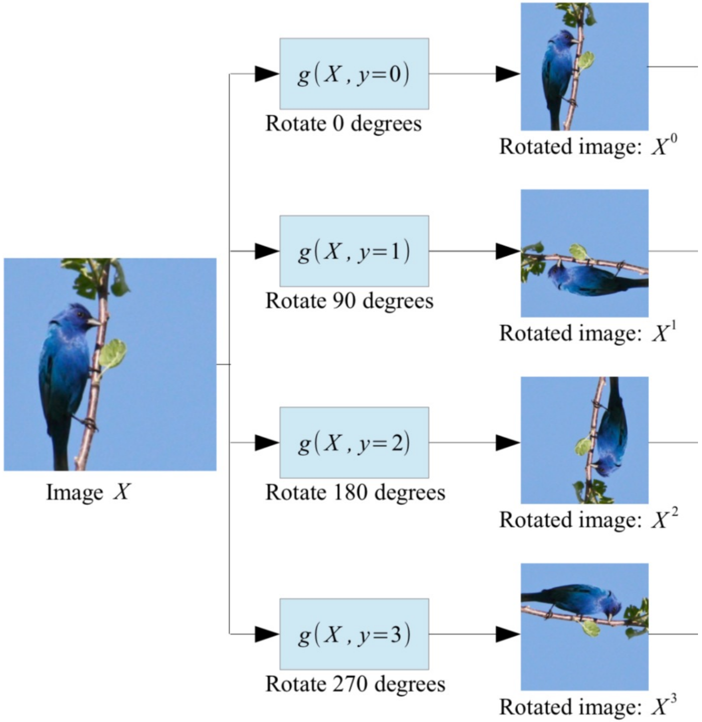
Pretext Tasks

- Exemplar images



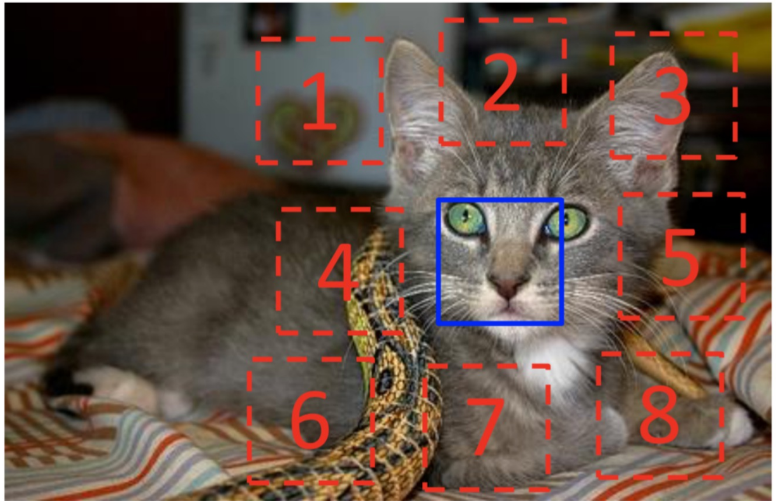
[1] Dosovitskiy et al. Discriminative Unsupervised Feature Learning with Convolutional Neural Networks. NeuRIPS 2014.

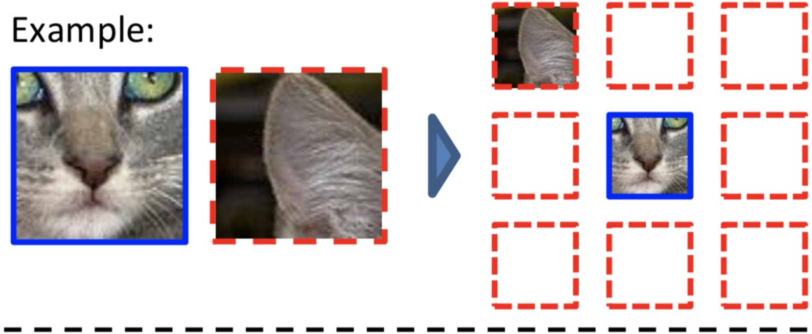
Pretext Tasks - Rotation



[1] Gidaris et al. Unsupervised Representation Learning by Predicting Image Rotations. ICLR 2018.

Pretext Tasks - Patching





[1] Doersch et al. Unsupervised Visual Representation Learning by Context Prediction. ICCV 2015.

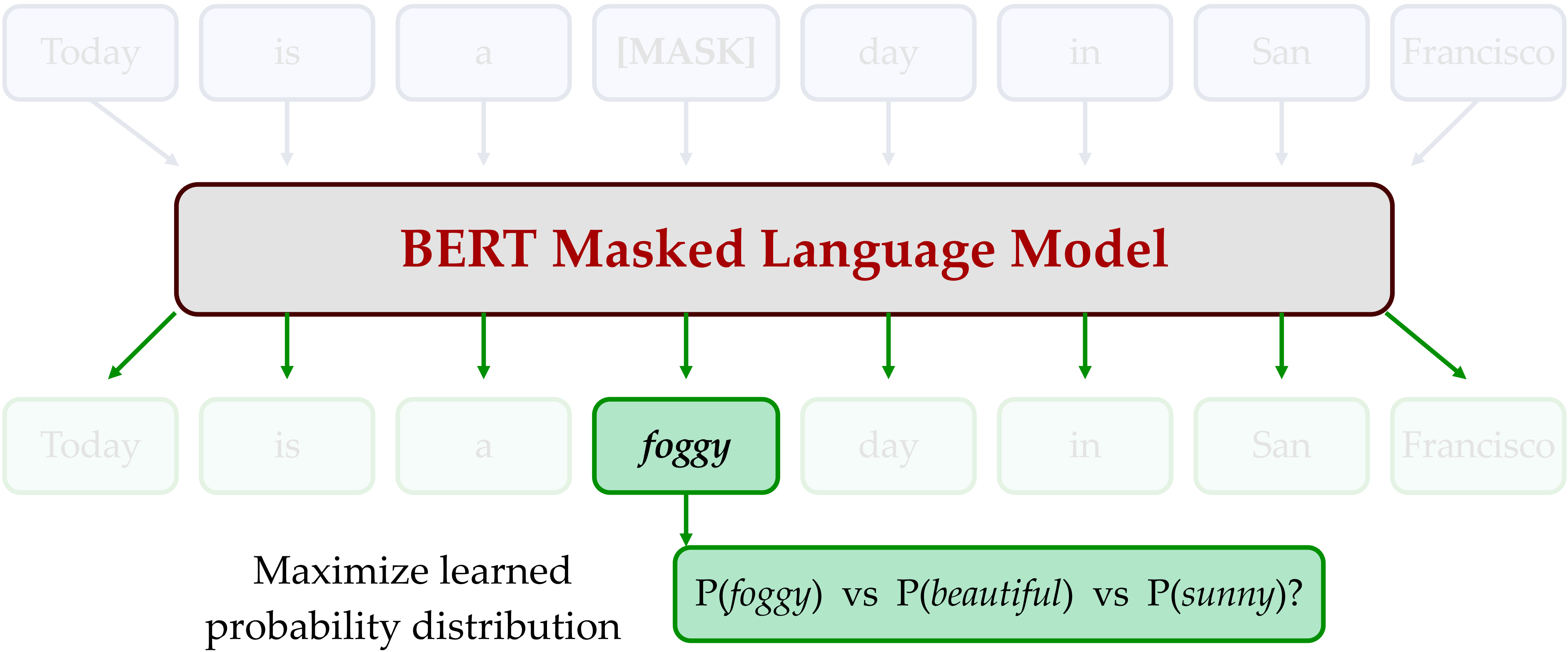
What Position Does the Blue Square Occupy?

1	2	3
4	5	6
7	8	9

Image Inpainting

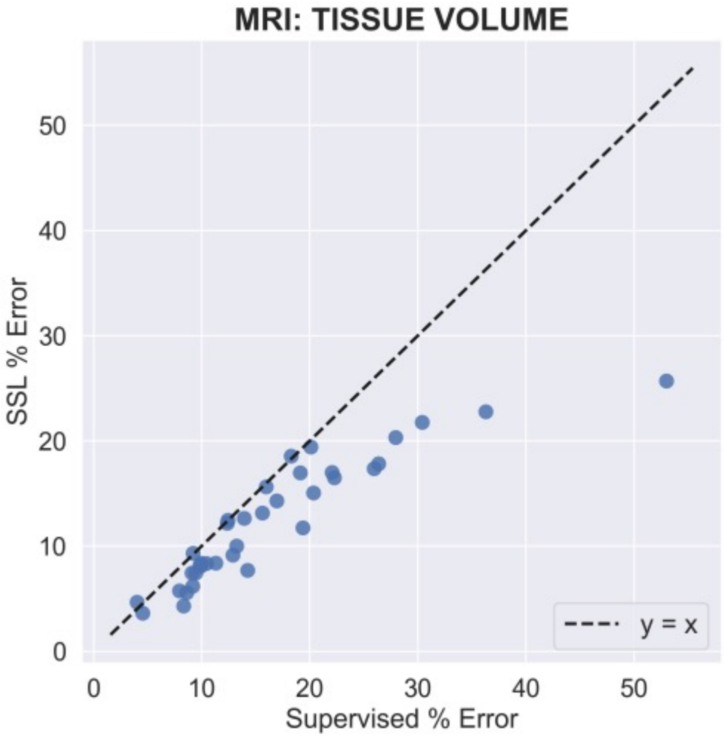
[1] Dominic et al. Improving Data-Efficiency and Robustness of Medical Imaging Segmentation Using Inpainting-Based Self-Supervised Learning. Bioengineering 2021.

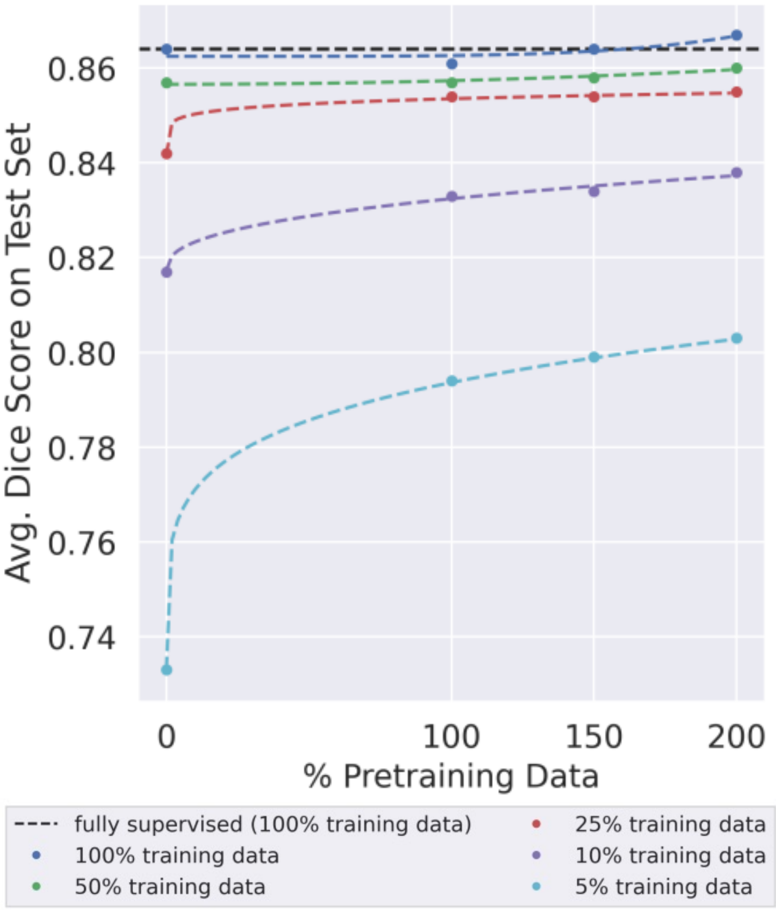
BERT Pretraining



Maximize learned
probability distribution

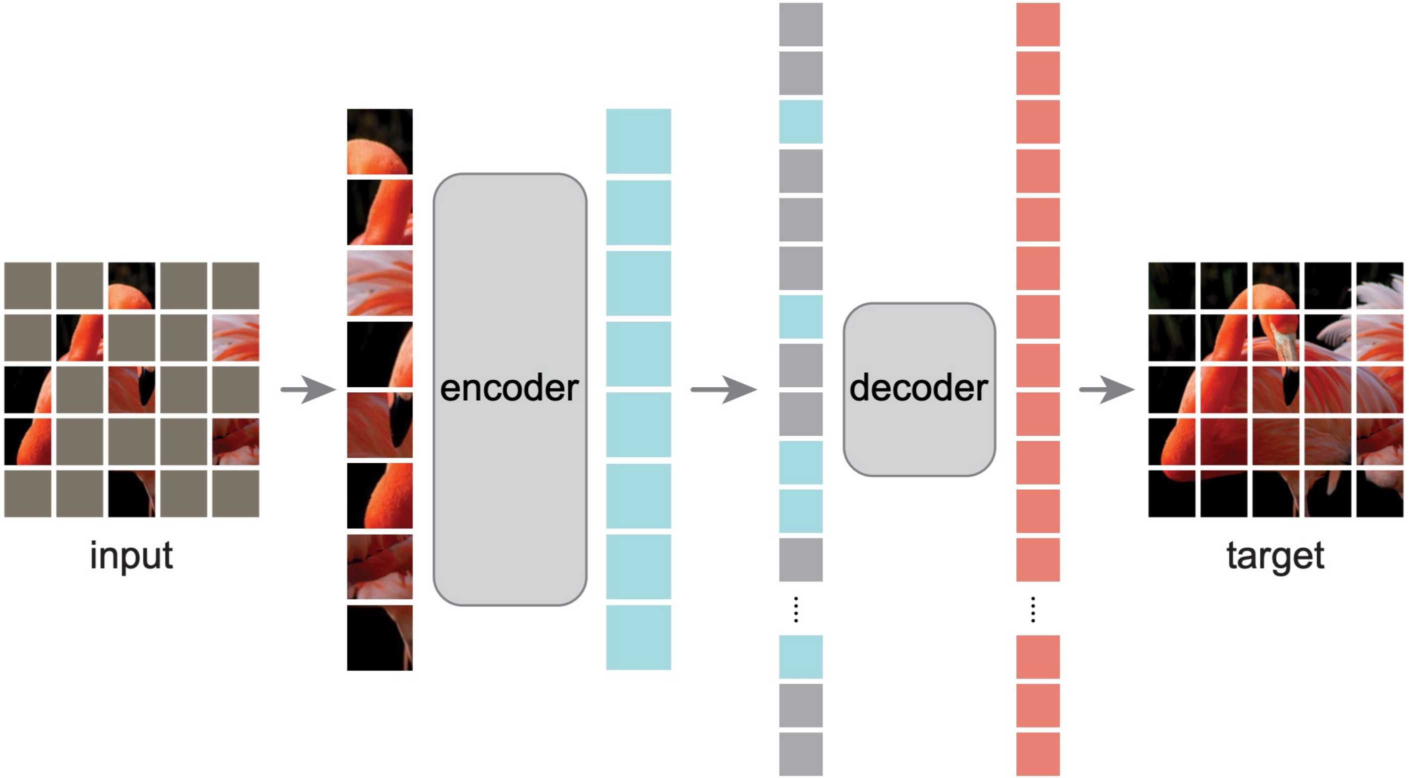
Image SSL Benefits





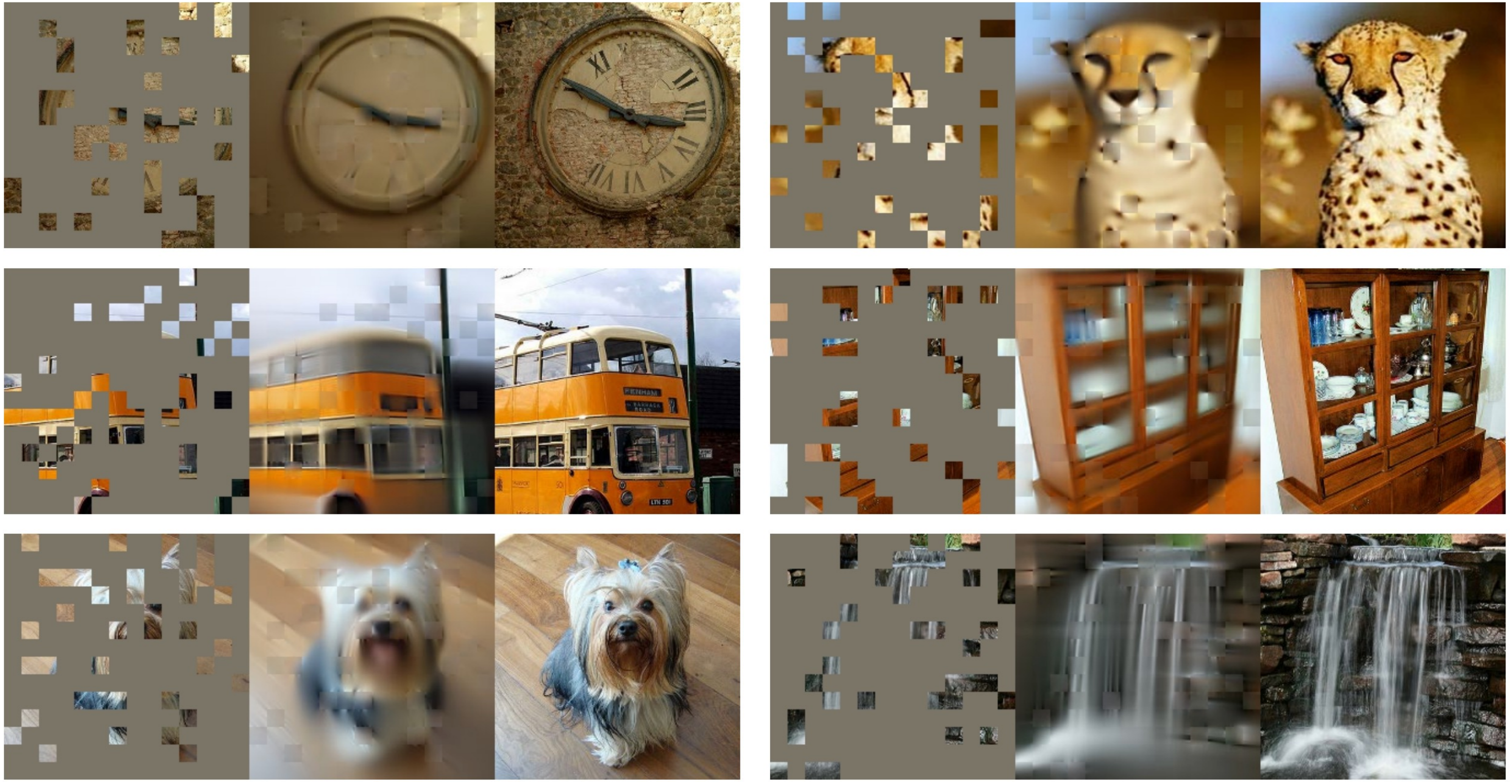
[1] Dominic et al. Improving Data-Efficiency and Robustness of Medical Imaging Segmentation Using Inpainting-Based Self-Supervised Learning. Bioengineering 2021.

Latest Self-Supervised Learning



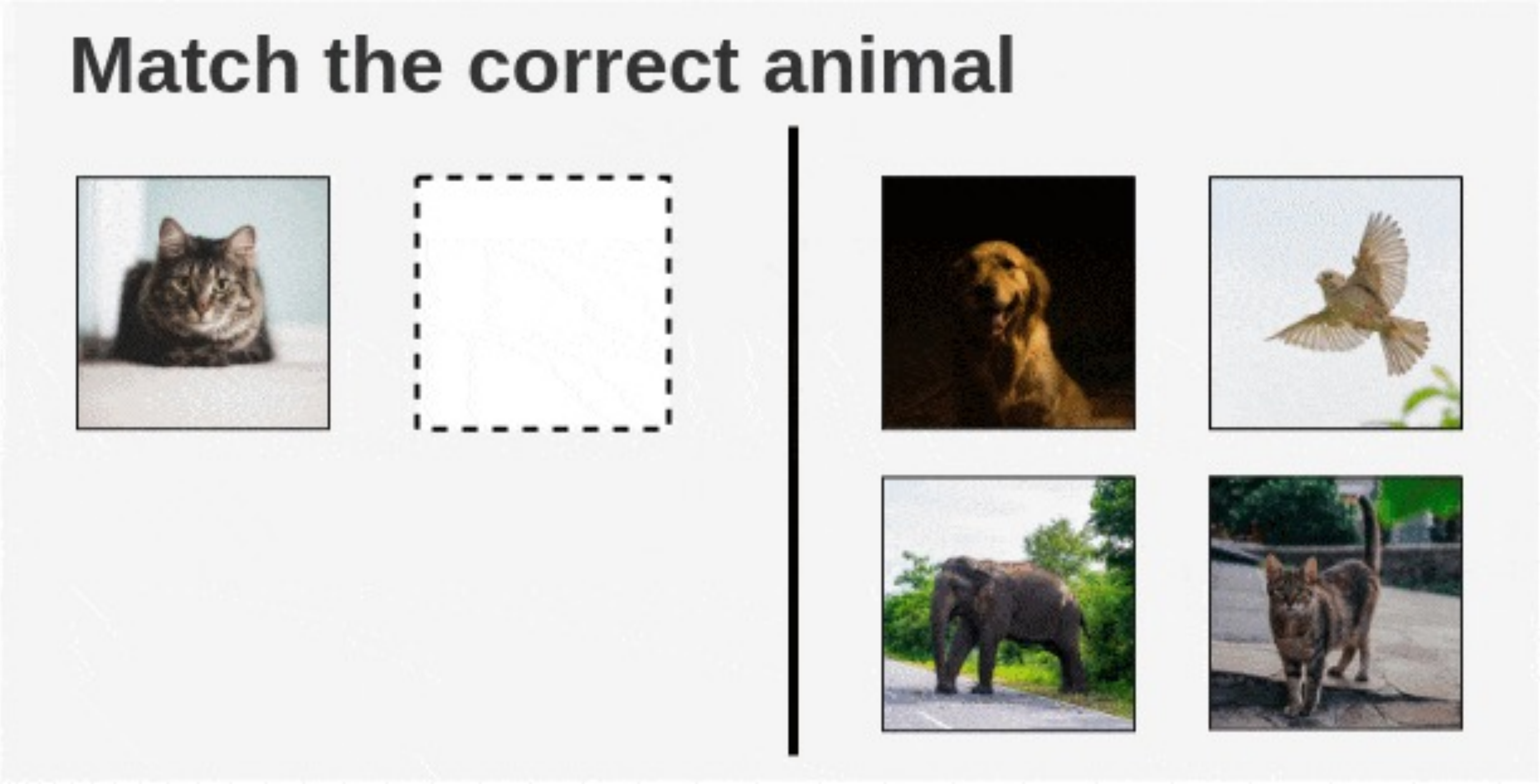
[1] He et al. Masked Autoencoders are Scaling Vision Learners. CVPR 2022.

Latest Self-Supervised Learning



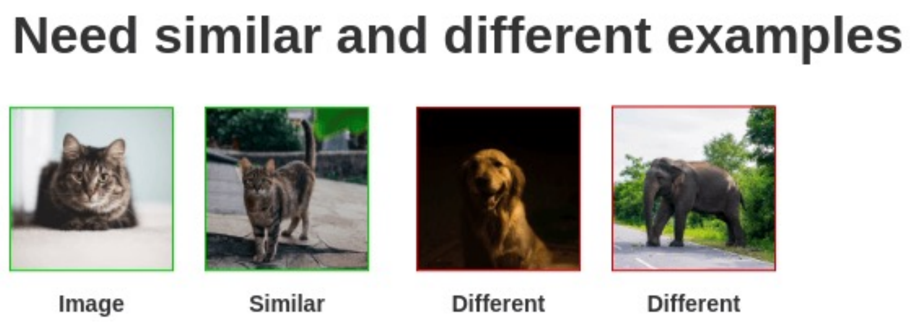
[1] He et al. Masked Autoencoders are Scaling Vision Learners. CVPR 2022.

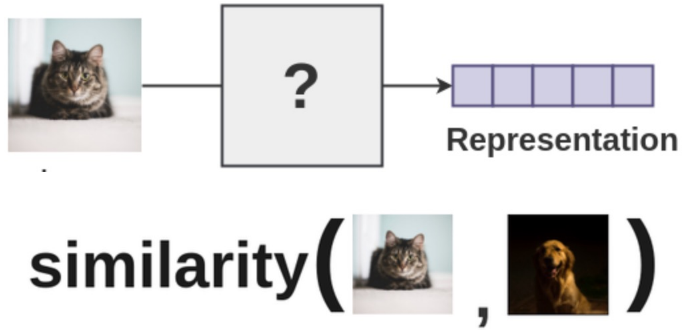
Contrastive Learning

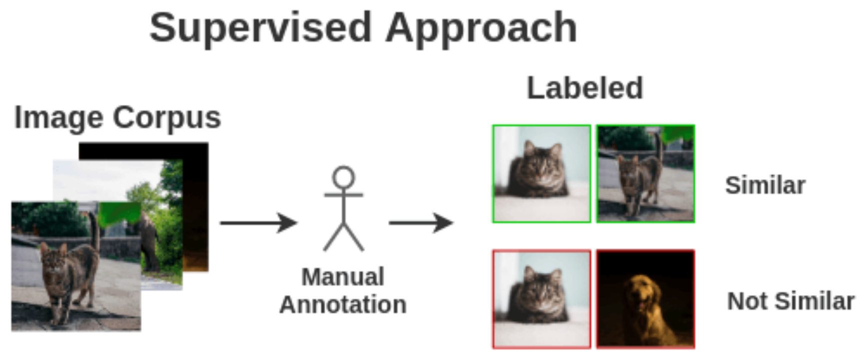


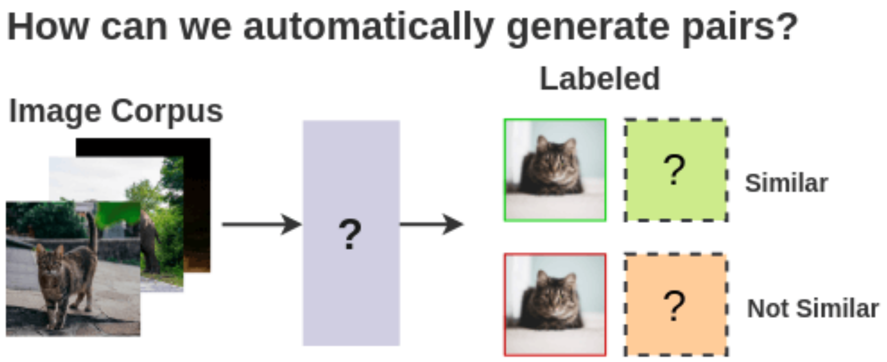
[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.

Contrastive Learning



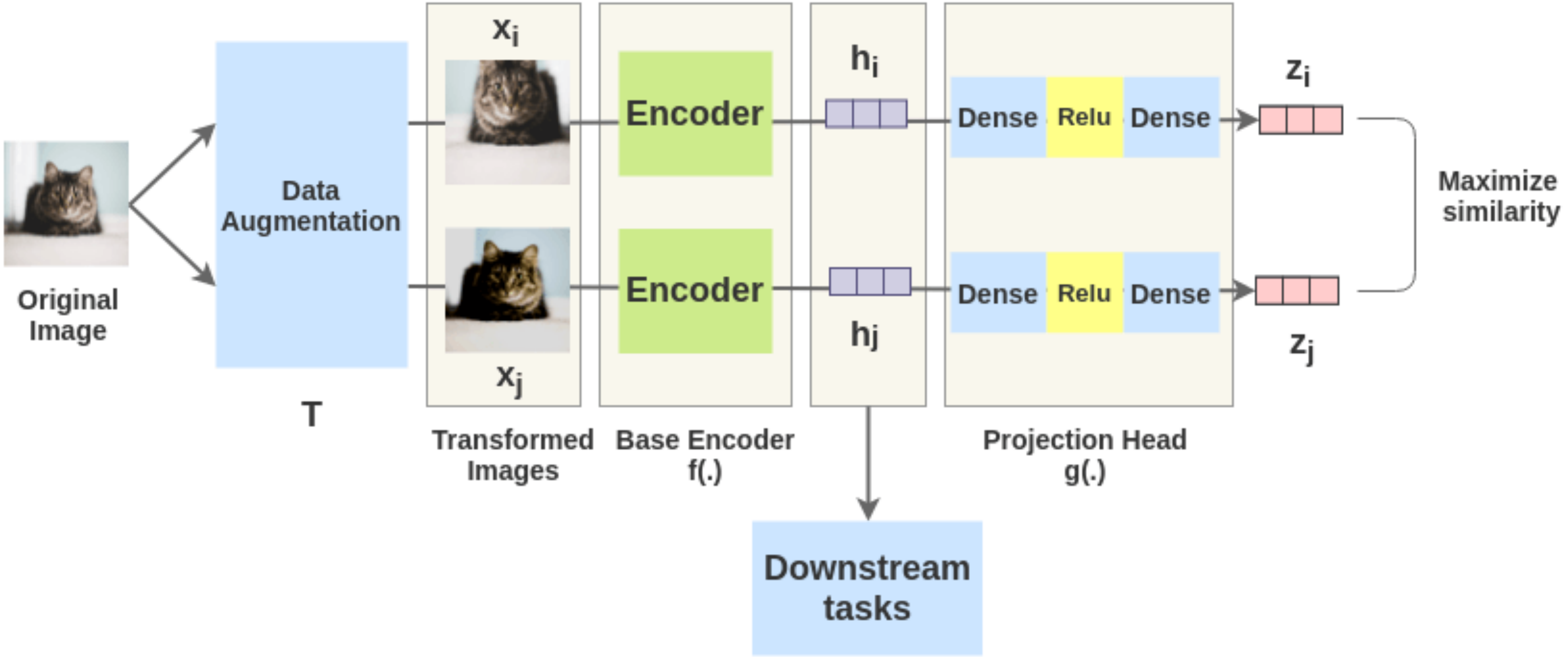






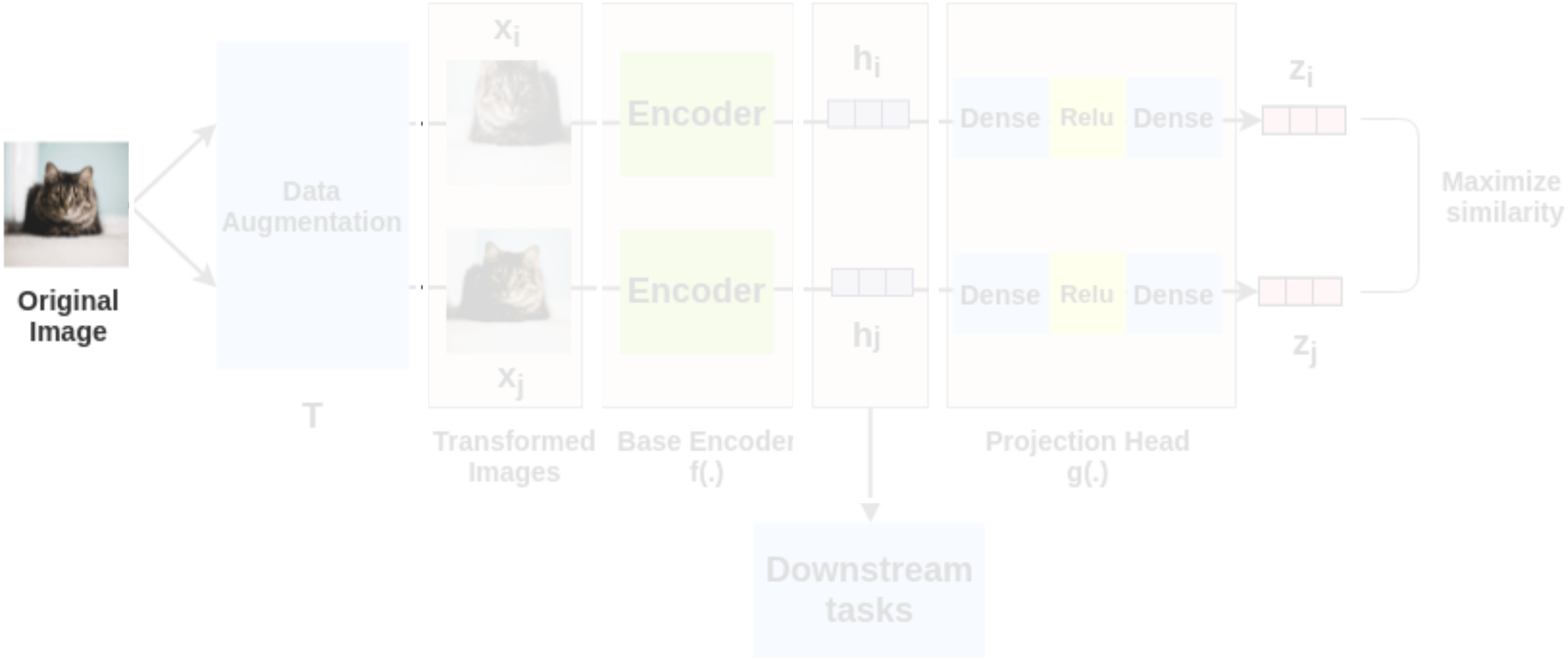
[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.

Contrastive Learning



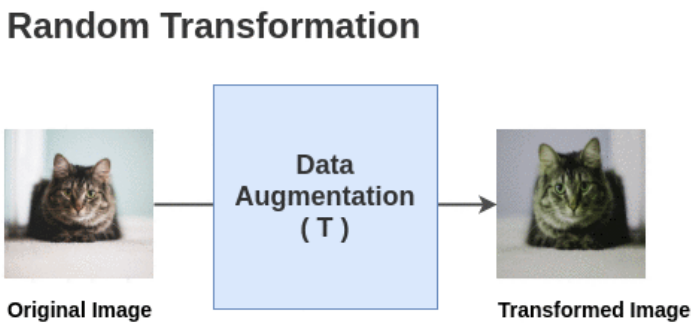
[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.
[2] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

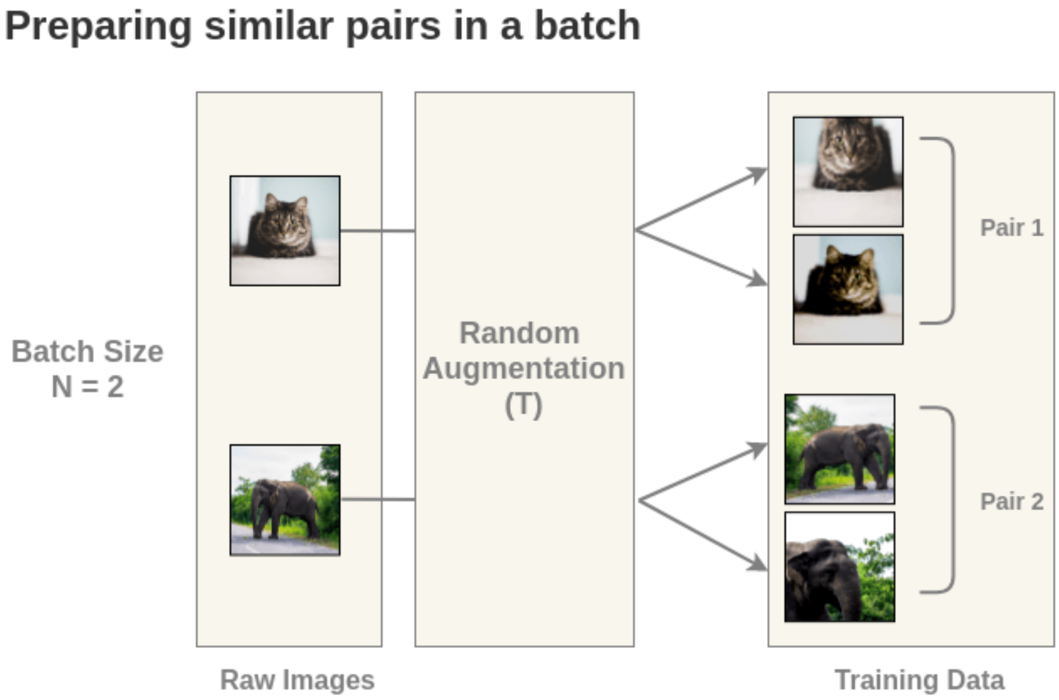
Contrastive Learning



[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.
[2] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

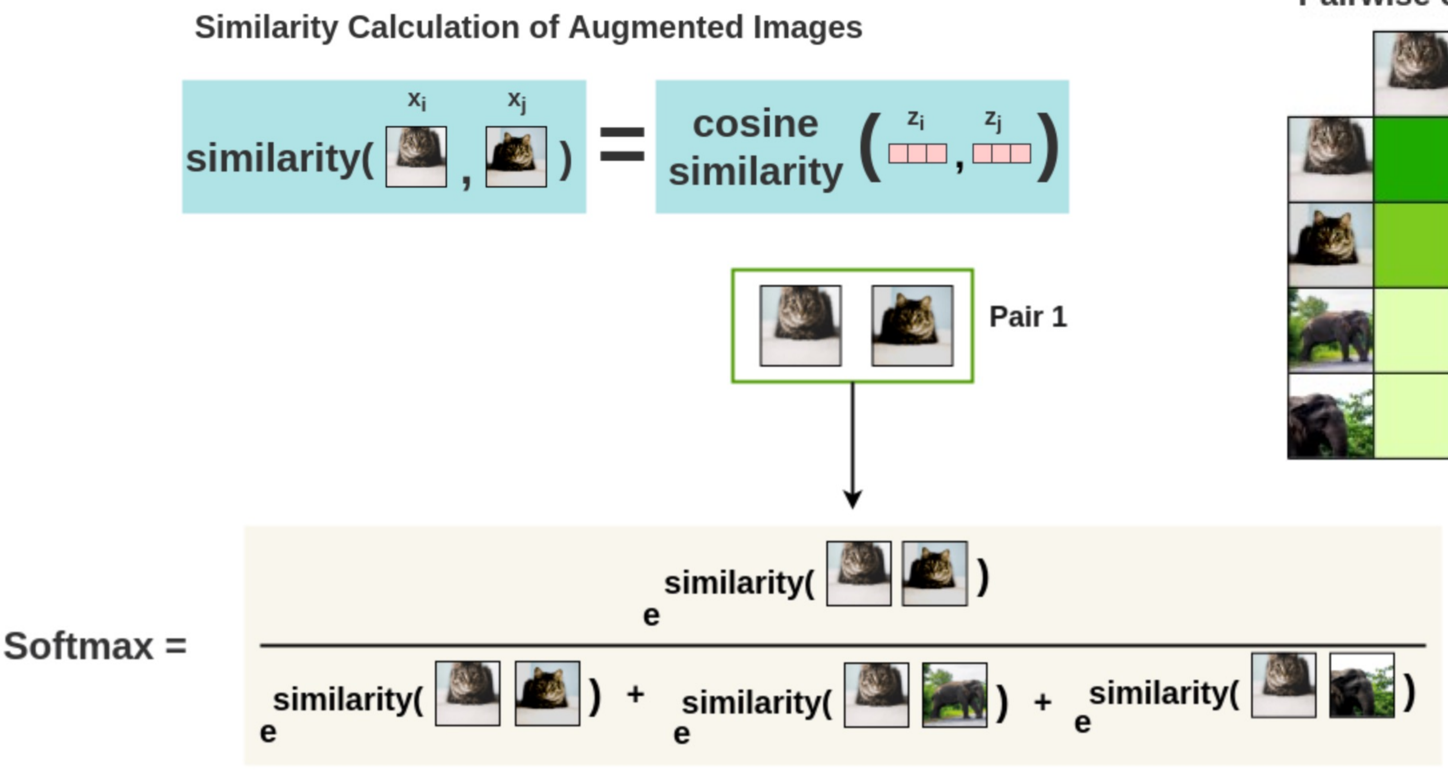
Contrastive Learning

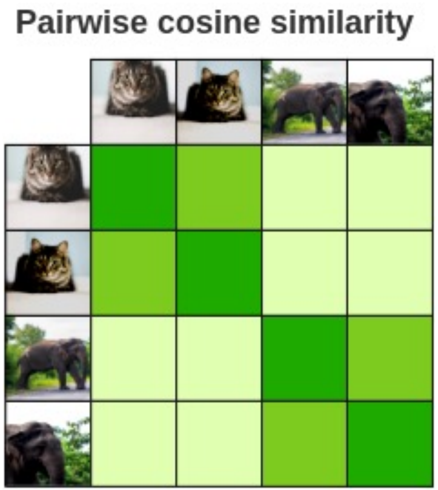




[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.
[2] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

Contrastive Learning



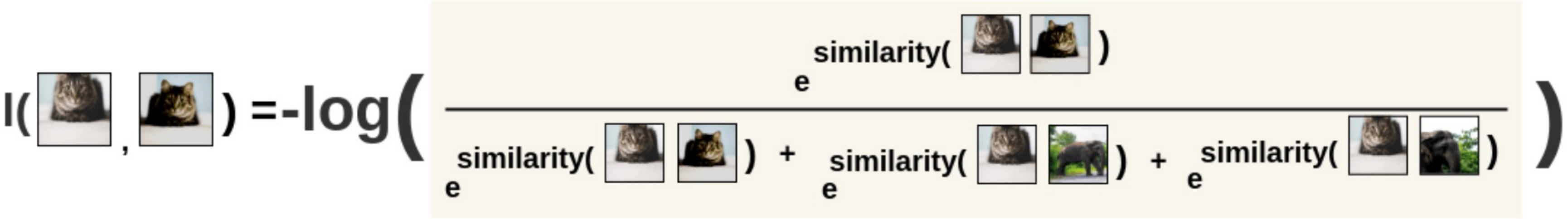


[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.
[2] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

Noise Contrastive Estimation Loss

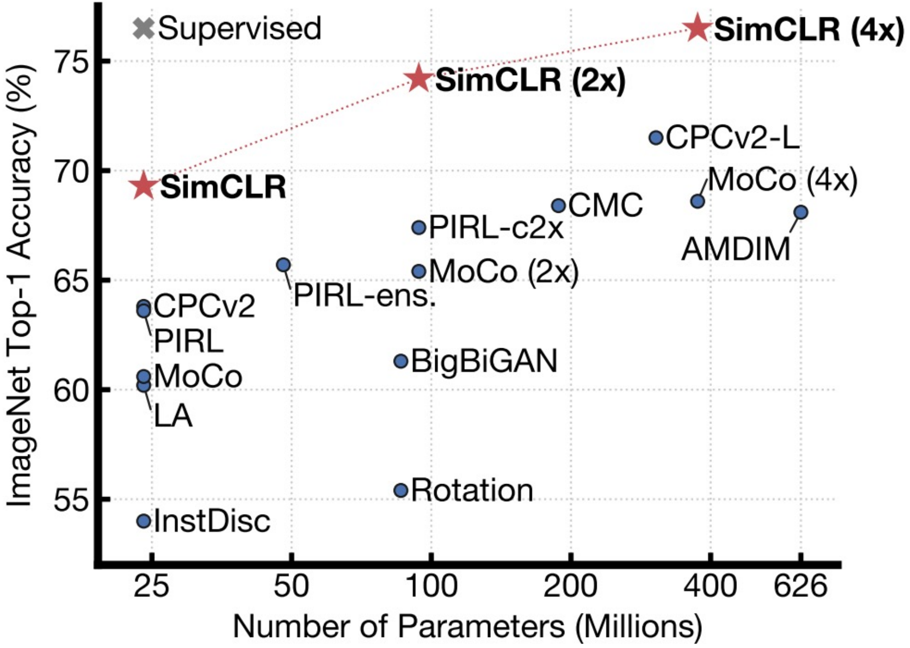
- Compute over both pairs to account for asymmetry

$$l(i,j) = -log\frac{exp(s_{i,j})}{\sum_{k=1}^{2N} l_{[k!=i]}exp(s_{i,k})}$$



[1] Chaudhary A. The Illustrated SimCLR Framework. https://amitness.com/2020/03/illustrated-simclr/ 2020.
[2] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

SimCLR Technique



- Large batch size requirements
- Long training times
- Heuristic data augmentations

[1] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

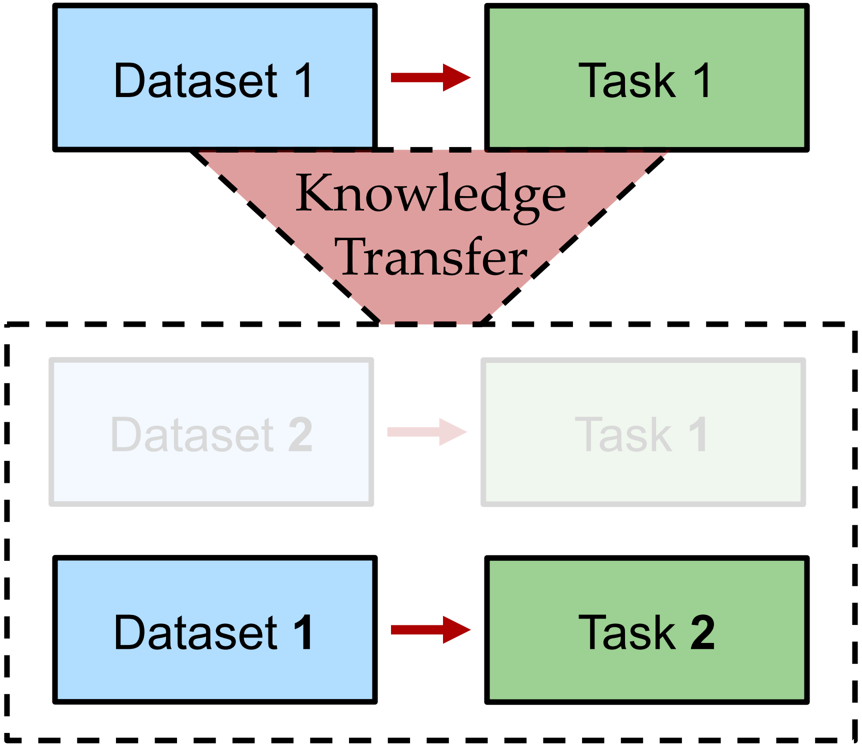
SimCLR Augmentations

(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

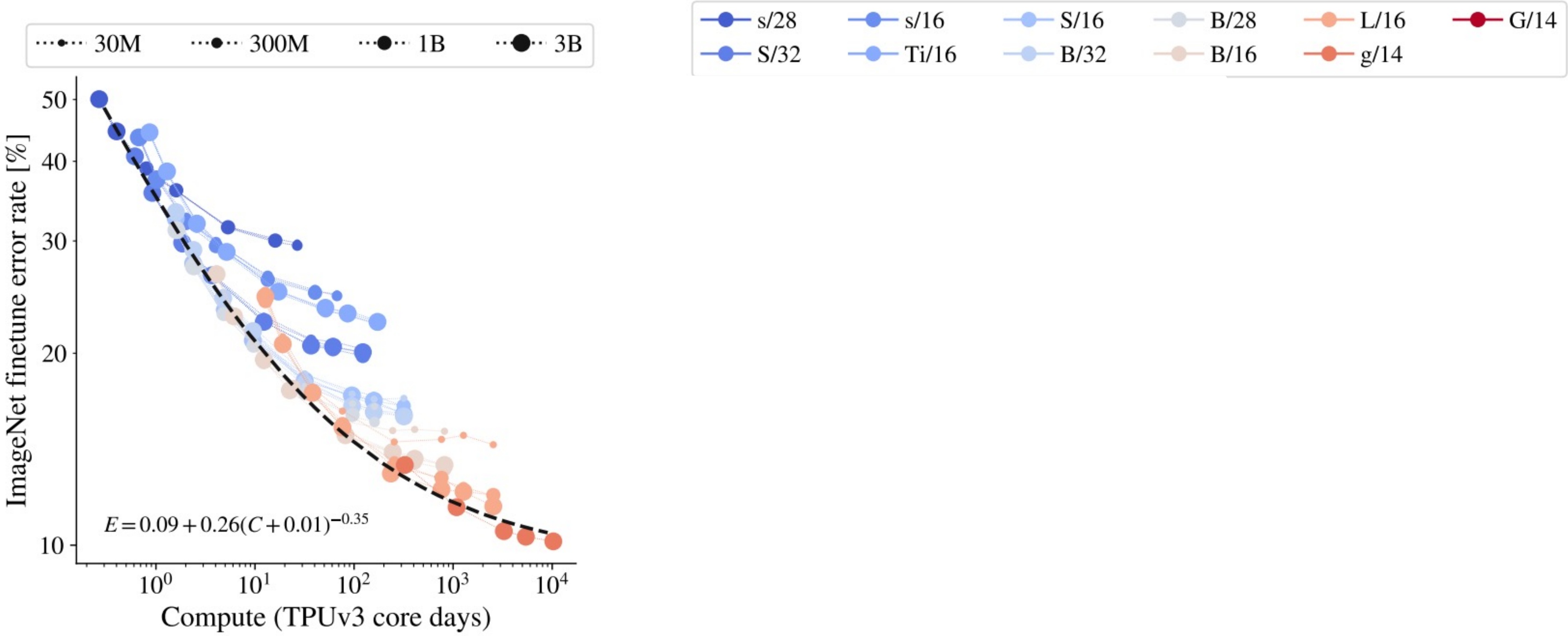
(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

[1] Chen et al. A Simple Framework for Contrastive Learning of Visual Representations. ICML 2020.

Transfer Learning

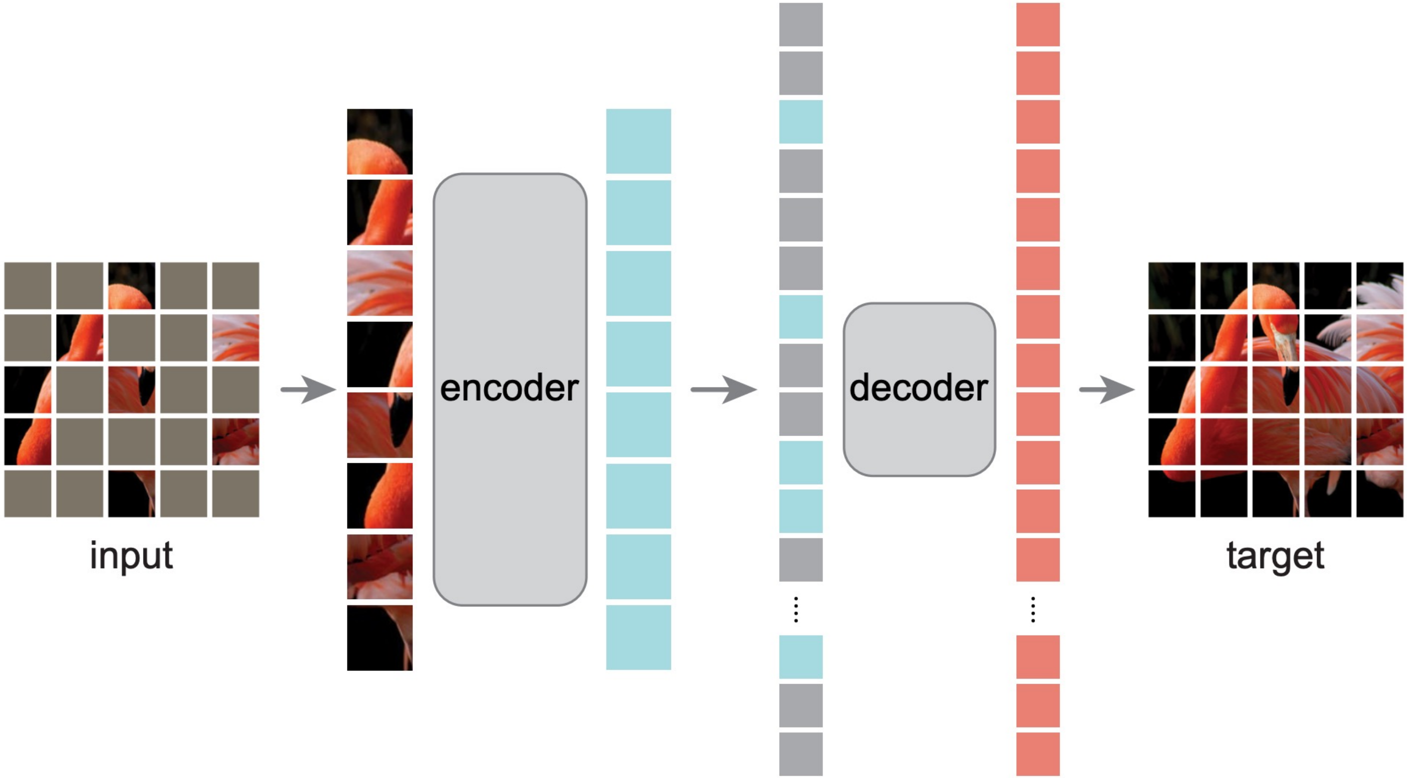


Benefits of Scaling



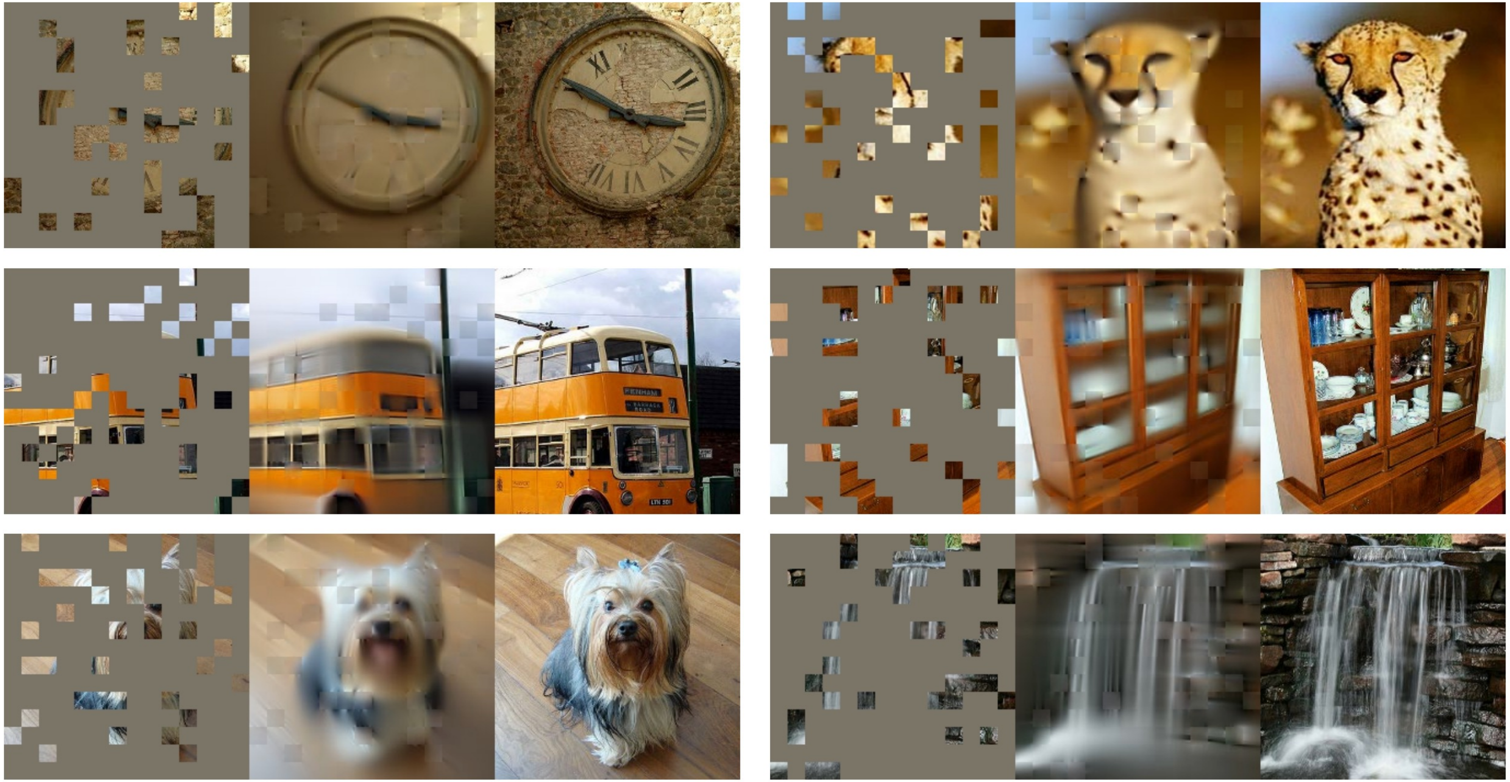
[1] Zhai et al. Scaling Vision Transformers. CVPR 2022.

Latest Self-Supervised Learning



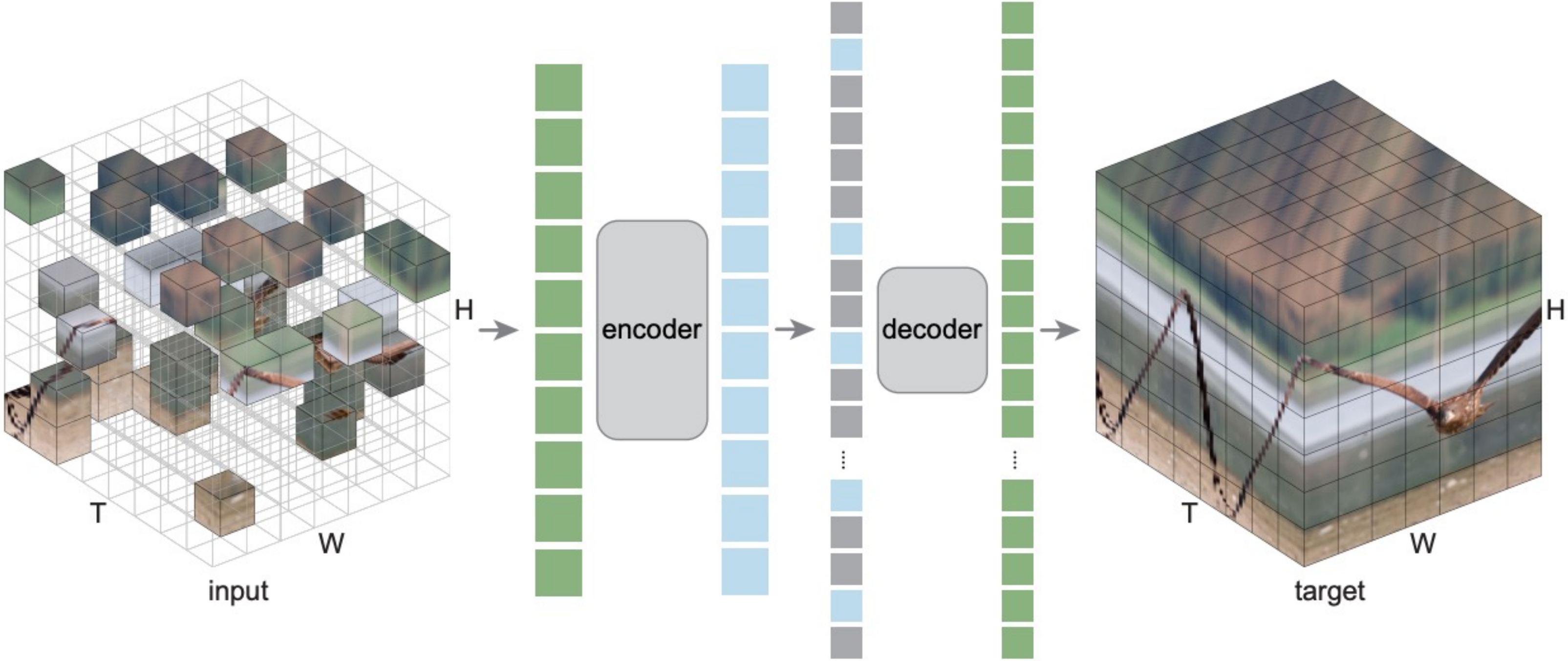
[1] He et al. Masked Autoencoders are Scaling Vision Learners. CVPR 2022.

Latest Self-Supervised Learning



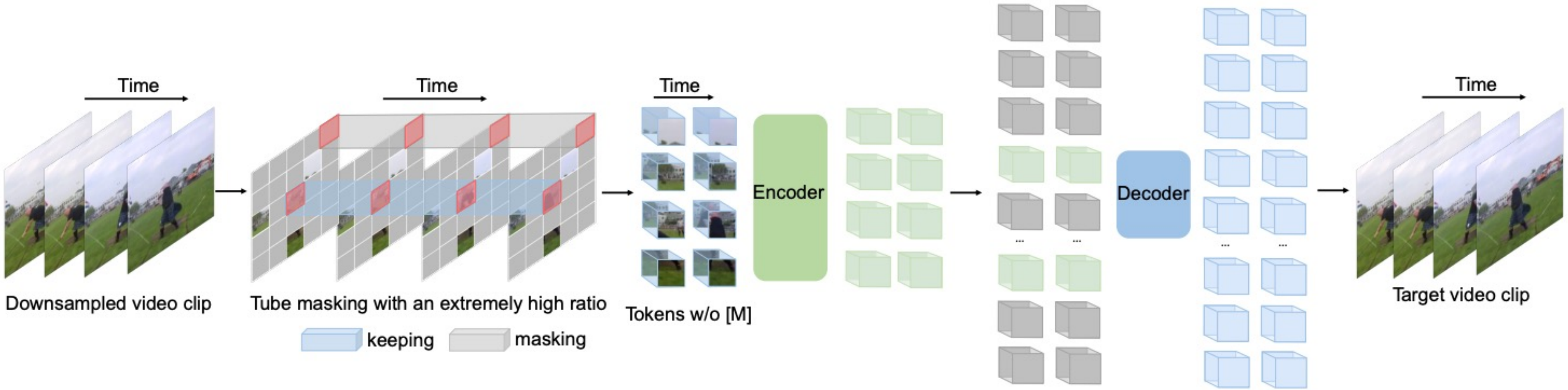
[1] He et al. Masked Autoencoders are Scaling Vision Learners. CVPR 2022.

Extensions of Prior Approaches



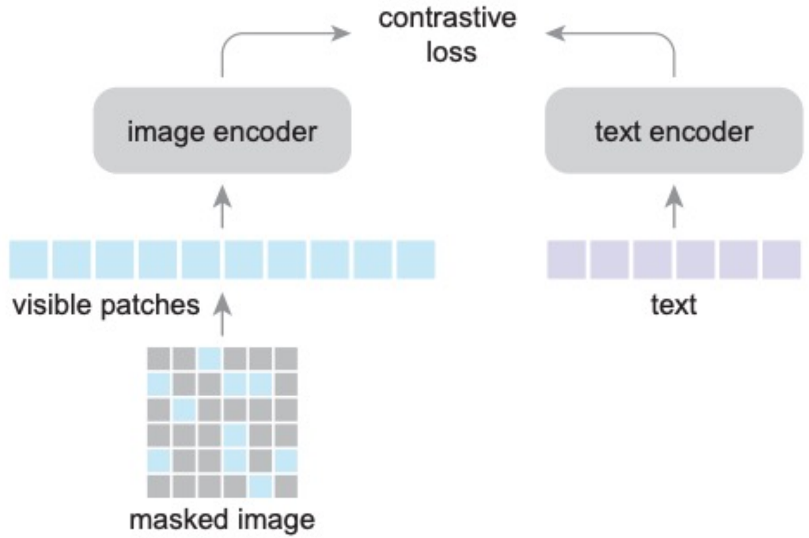
[1] Feichtenhofer et al. Masked Autoencoders As Spatiotemporal Learners. NeurIPS 2022.

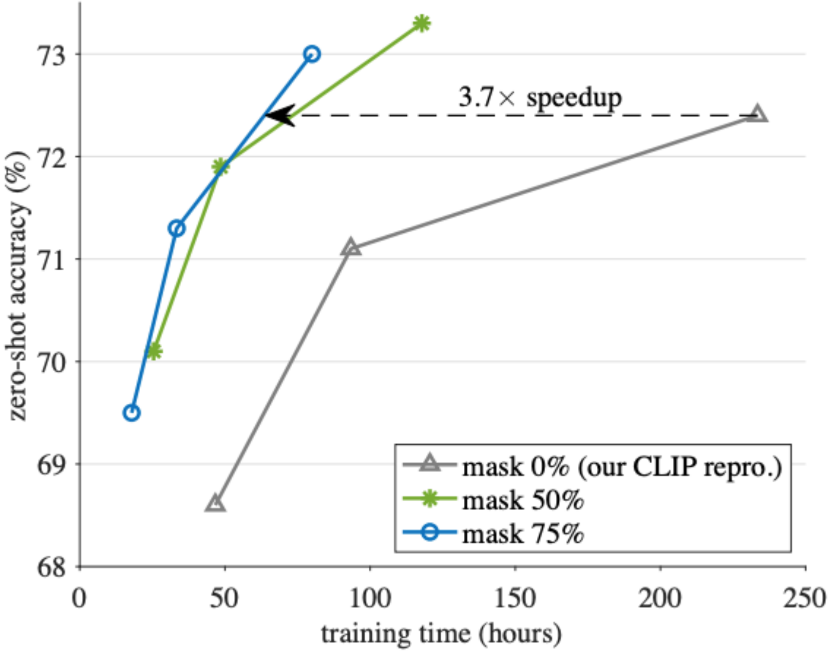
Tube Masking for Video MAE



[1] Tong et al. VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training. NeurIPS 2022.

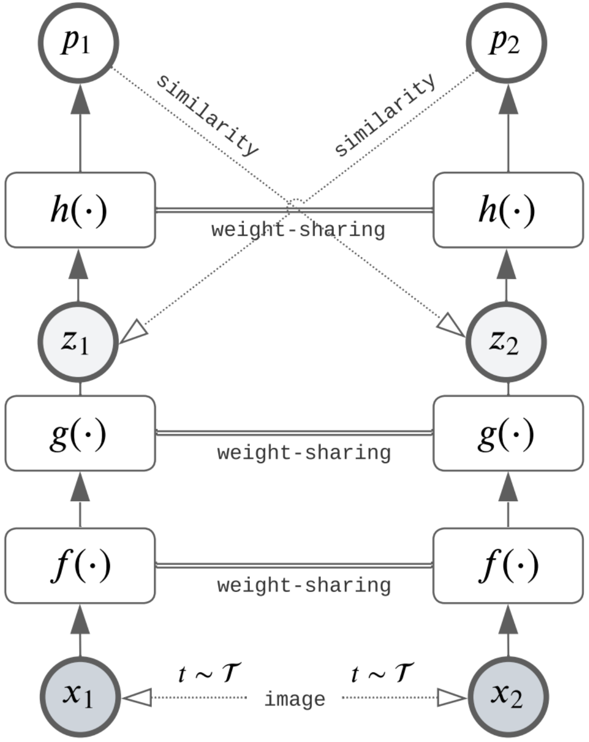
Masking for Multi-Modal Learning





[1] Li et al. Scaling Language-Image Pre-training via Masking. CVPR 2023.

Adaptations to Medical Imaging

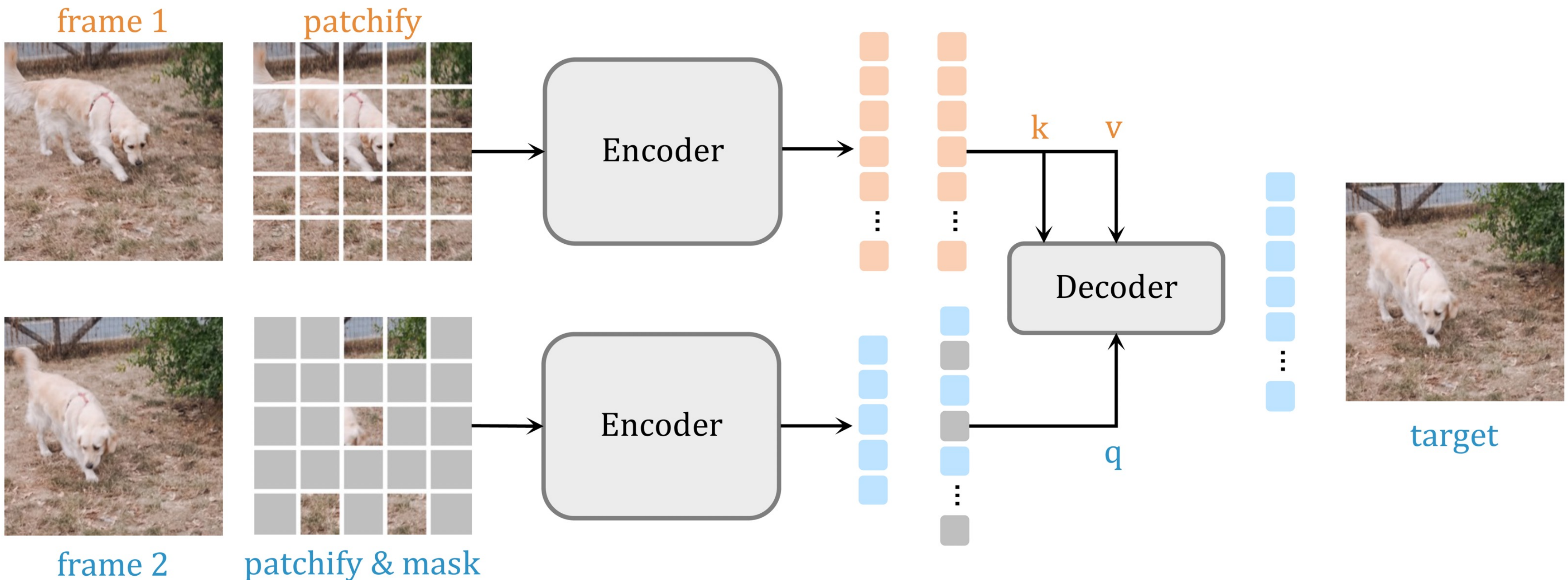


- Architectures inspired by the SimSiam method
- X = Image
- F = Image Encoder
- G, H = MLP projectors
- Z = latent representation
- P = latent representation predictors

$$\mathcal{L} = -\frac{1}{2}\left(\frac{p_1}{\|p_1\|_2} \cdot \frac{z_2}{\|z_2\|_2}\right) - \frac{1}{2}\left(\frac{p_2}{\|p_2\|_2} \cdot \frac{z_1}{\|z_1\|_2}\right)$$

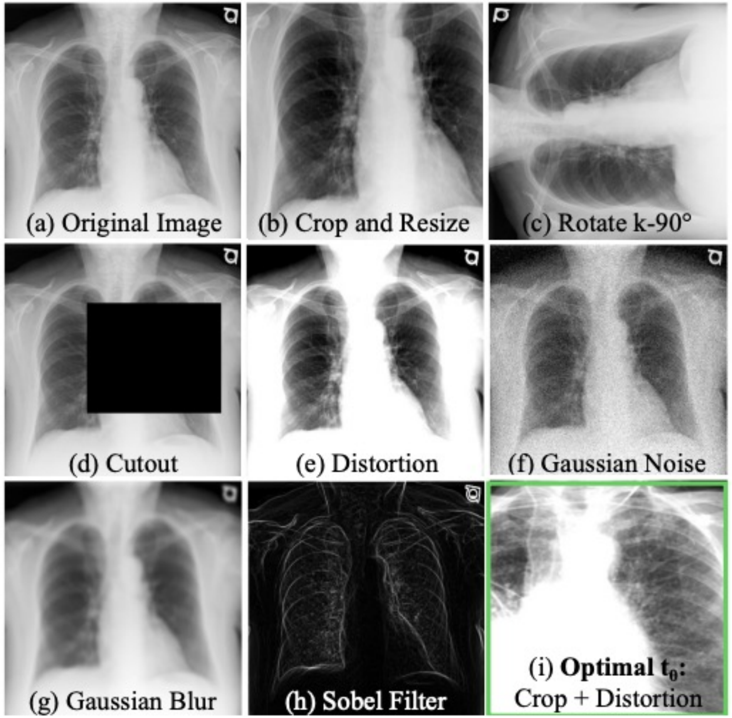
[1] van der Sluijs et al. Exploring Image Augmentations for Siamese Representation Learning with Chest X-Rays. MIDL 2023
[2] Chen et al. Exploring Simple Siamese Representation Learning. CVPR 2021

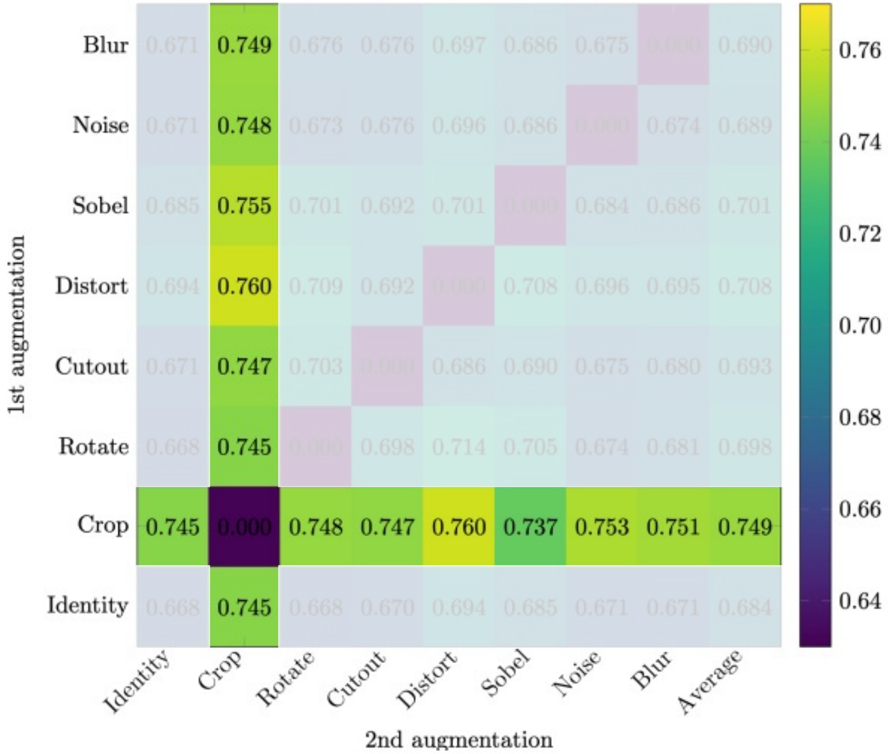
Siamese Masked Autoencoders



[1] Gupta et al. Siamese Masked Autoencoders. NeurIPS 2023.

Augmentations for Medical Imaging





[1] van der Sluijs et al. Exploring Image Augmentations for Siamese Representation Learning with Chest X-Rays. MIDL 2023.

Suggested Reading

A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen [1] **Simon Kornblith** [1] **Mohammad Norouzi** [1] **Geoffrey Hinton** [1]

Masked Autoencoders Are Scalable Vision Learners

Kaiming He[*,†] Xinlei Chen[*] Saining Xie Yanghao Li Piotr Dollár Ross Girshick

[*]equal technical contribution [†]project lead

Facebook AI Research (FAIR)

Questions?

akshaysc@stanford.edu