Overview of Deep Learning

From a Learning perspective

Mingtian Tan, Sept 2024

Supervised Learning

Supervised learning is characterized by training models using labeled data

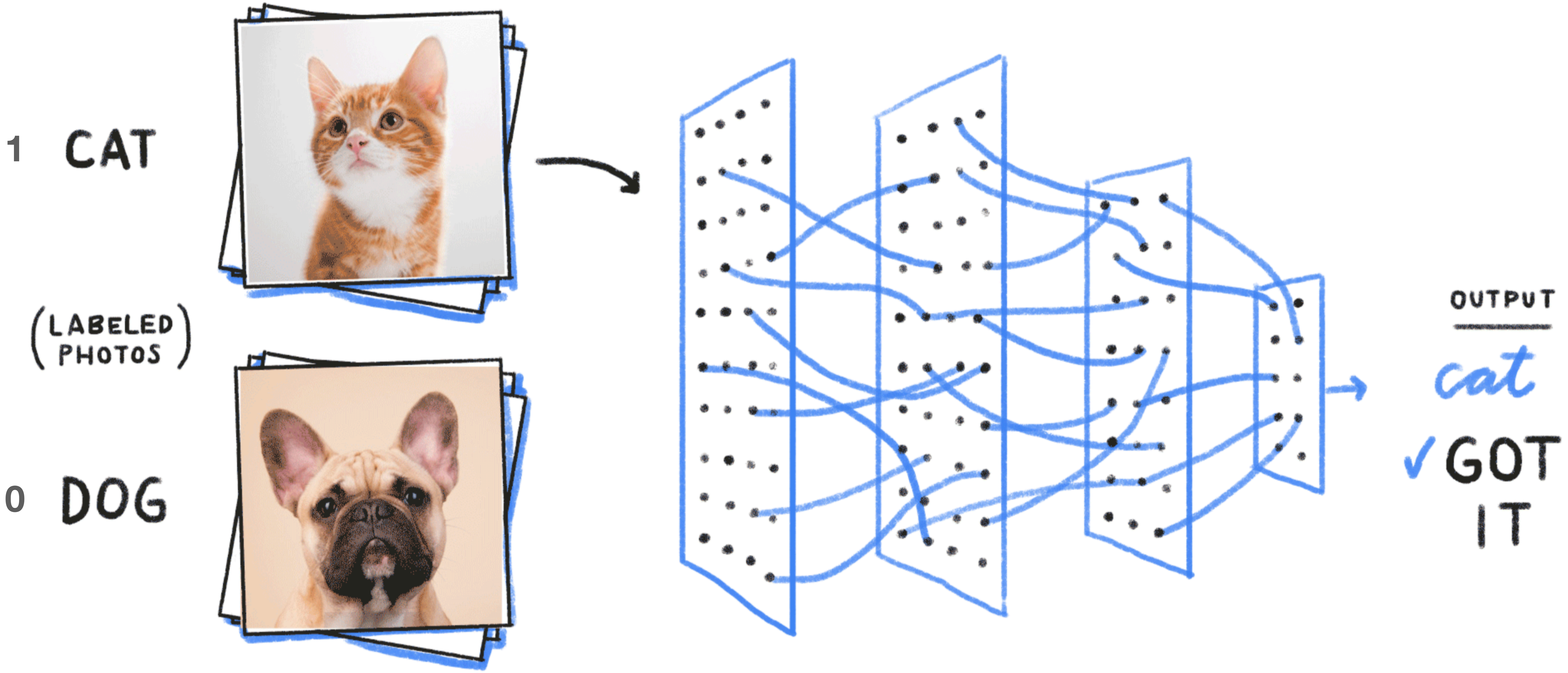
a.BackPropagation

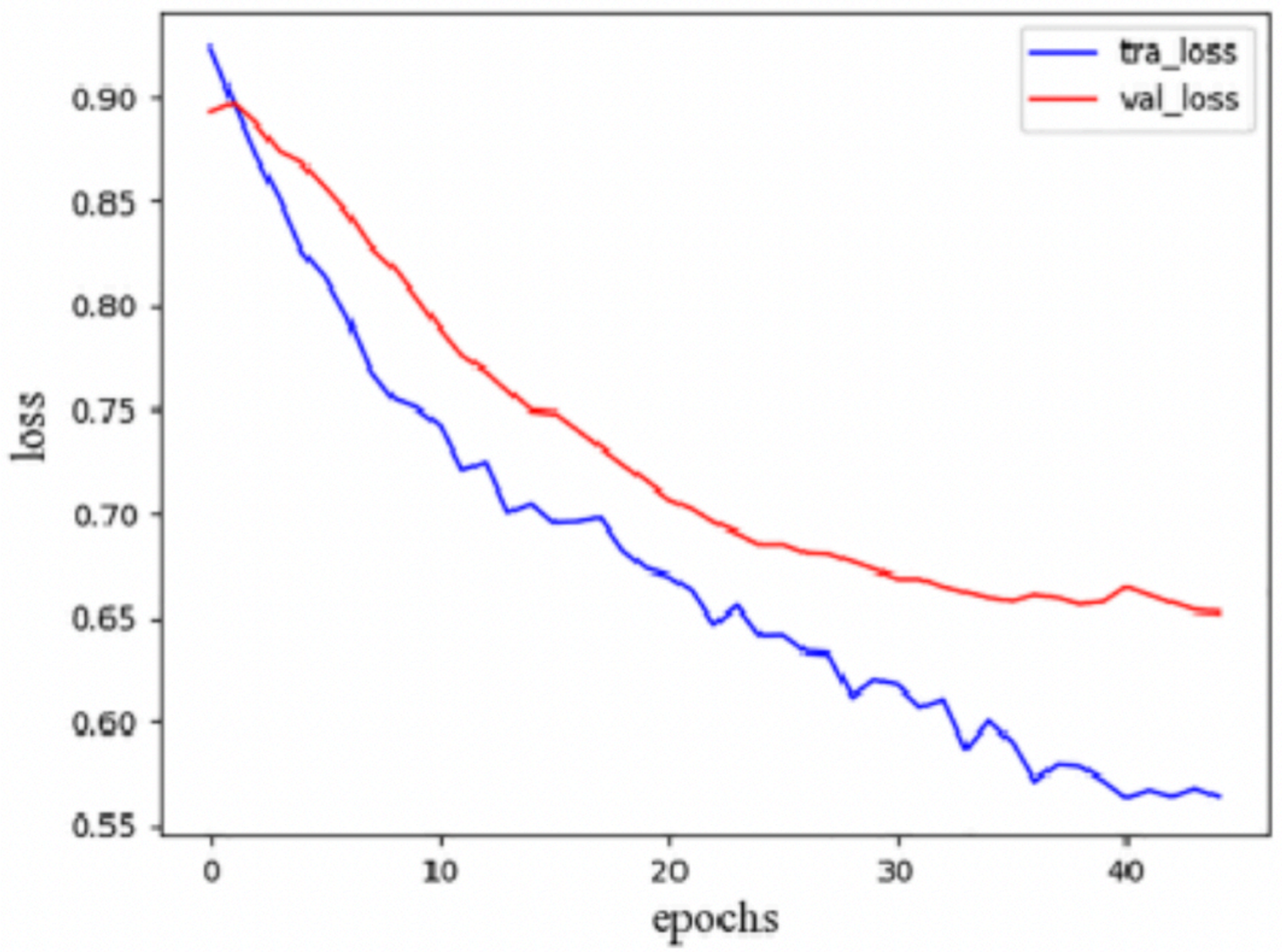
- Backpropagation is the process that helps neural networks learn by adjusting their weights. Here's how it works:
    - when we give the model some input, it makes a prediction. We then compare that prediction to the **actual result** and calculate the **error**.
    - Backpropagation takes this error and works backward through the network, adjusting the weights in each layer so the model gets closer to the correct answer next time.

Learning Representations by Back-Propagating Errors by Rumelhart et al.

Efficient BackProp by LeCun et al.

Source





- It’s like fine-tuning the model with every mistake, so over time, it becomes more accurate

b.Transfer Learning

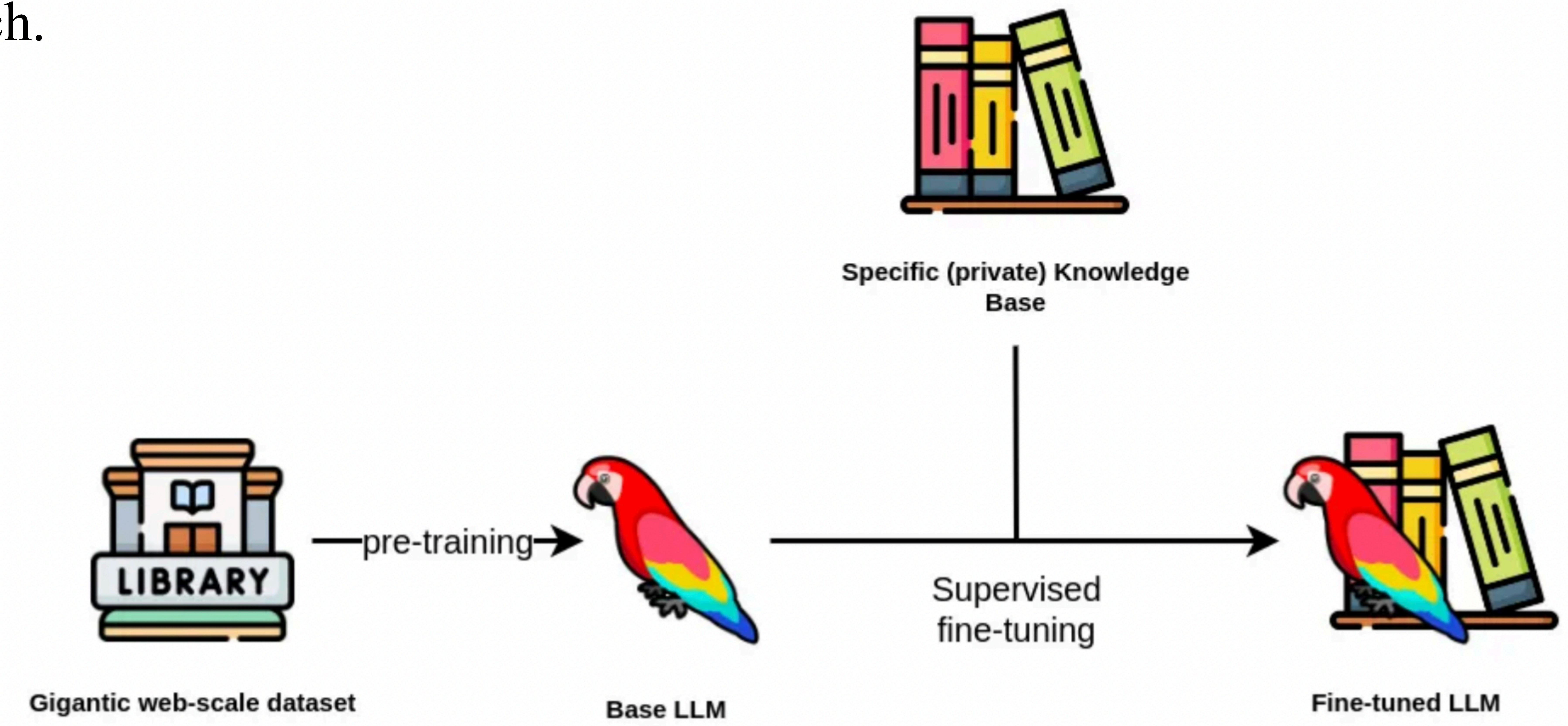
- Transfer learning is a technique where we take a model that's already been trained on a large dataset and fine-tune it for a specific task. Think of it as **borrowing the knowledge** the model has already gained and applying it to something new.

How transferable are features in deep neural networks? by Yosinski et al.

A Comprehensive Review on Transfer Learning by Pan and Yang.

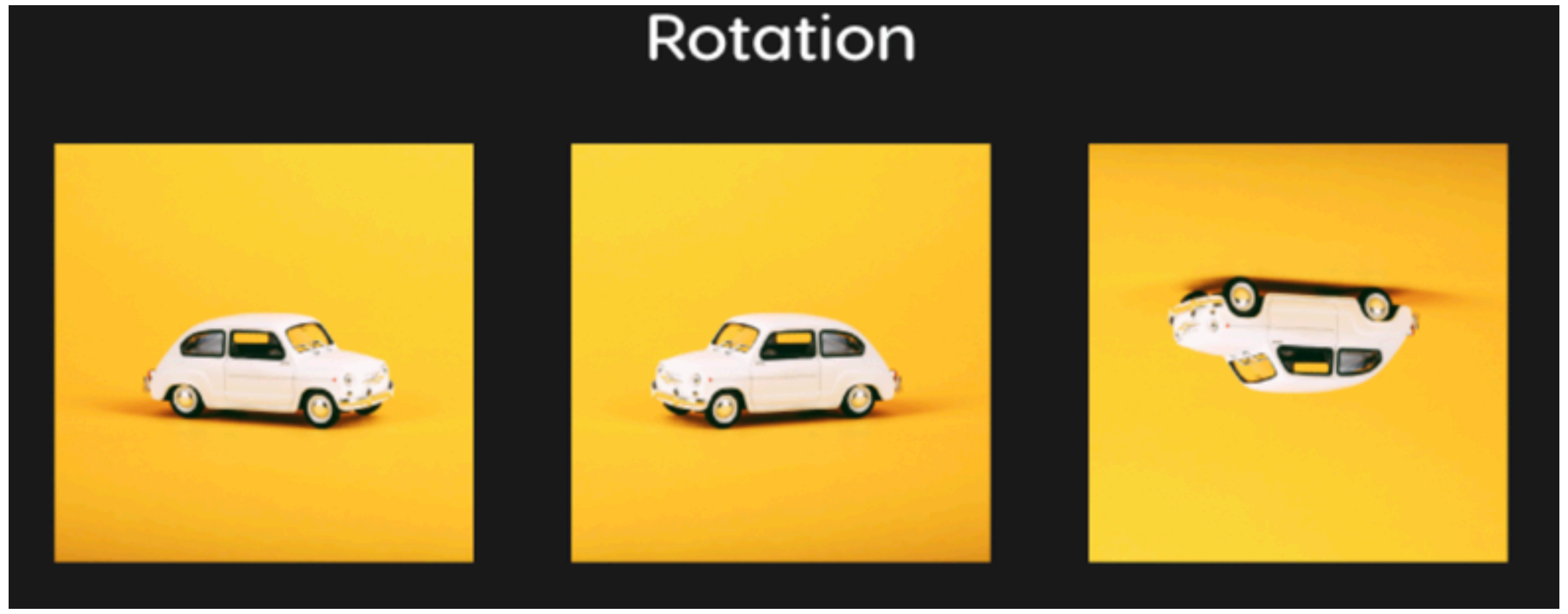
b.Transfer Learning

- For example, with large language models, the model is first trained on massive amounts of text from the internet. It learns general **<u>language patterns</u>**, **<u>grammar</u>**, and **<u>context</u>**. Then, we can take that model and fine-tune it on a smaller, specific dataset, like medical text. This process is much faster and more efficient than training a new model from scratch.



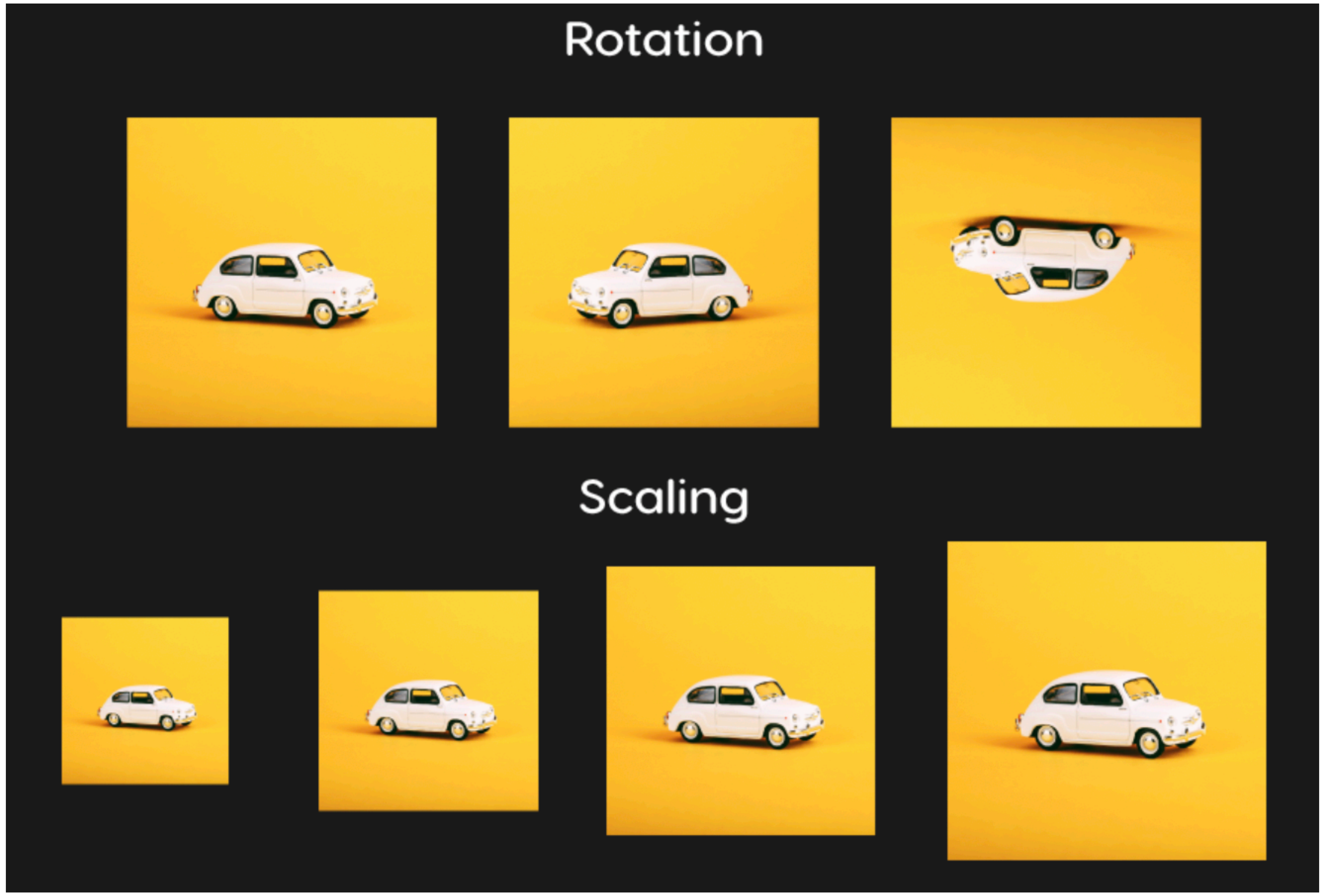
c.Data Augmentation

- Data augmentation is to improve the performance of models by artificially increasing the size of the training dataset. In image classification, for example, instead of just using the original images, we can make slight changes to them—like flipping, rotating, or cropping—so the model sees more samples.



Improved Regularization of Convolutional Neural Networks with Cutout by DeVries and Taylor.

Mixup: Beyond Empirical Risk Minimization by Zhang et al.


Rotation
Scaling

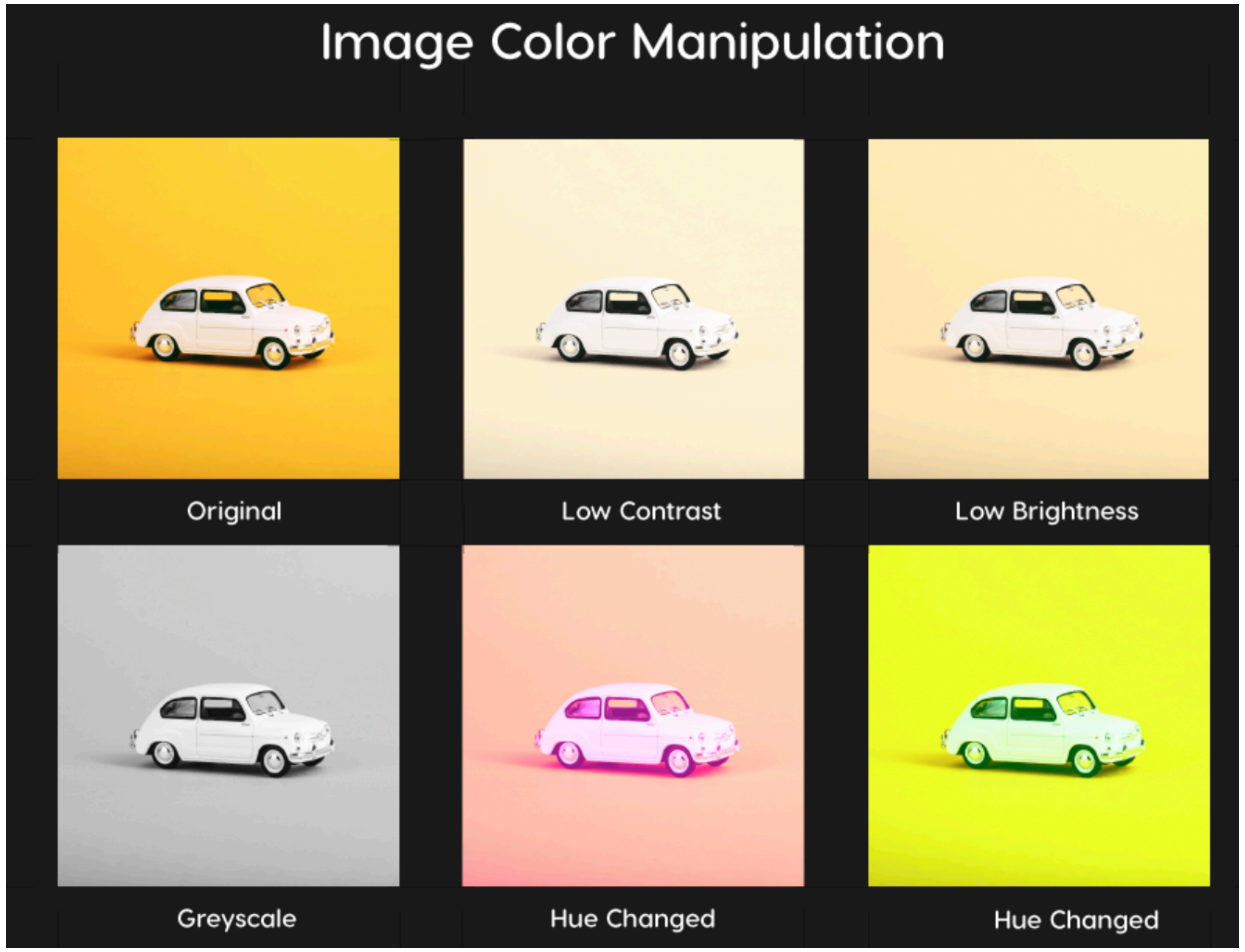
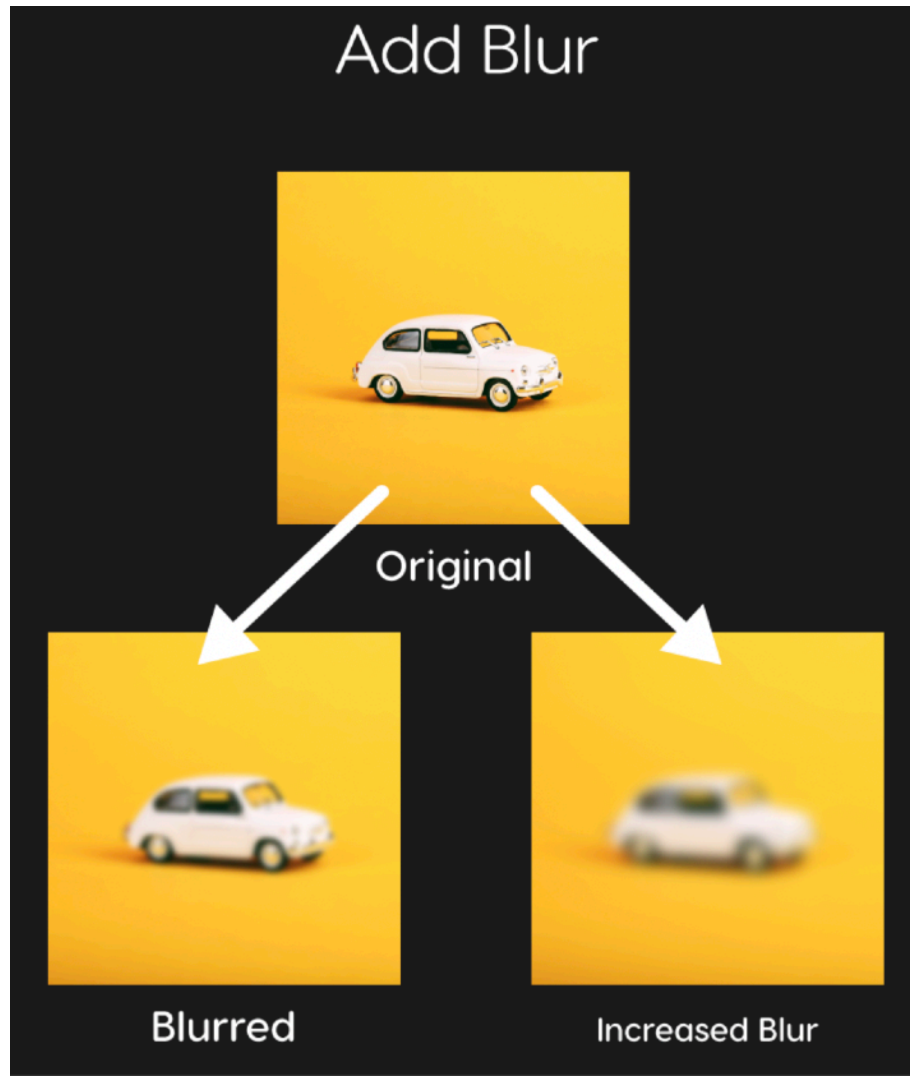
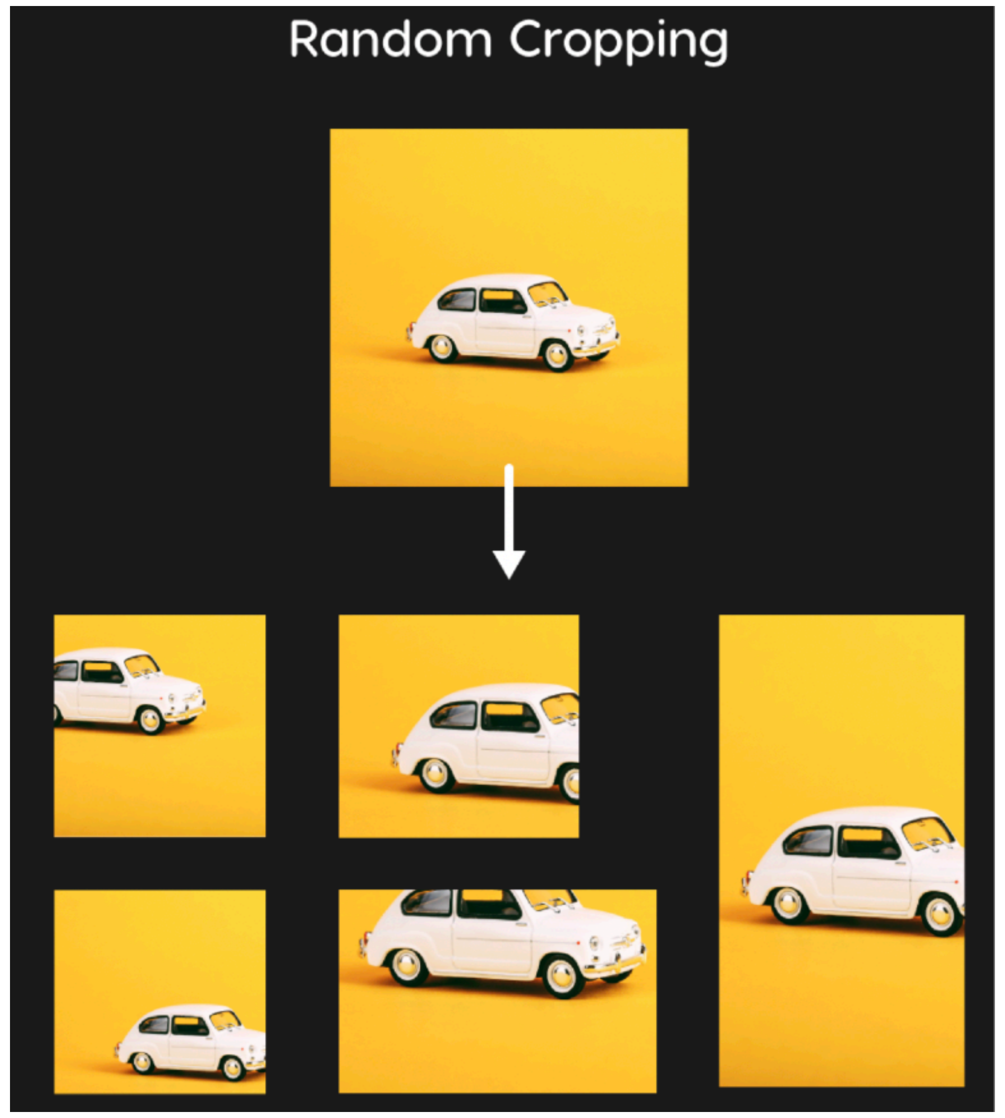


Image Color Manipulation
Original
Low Contrast
Low Brightness
Greyscale
Hue Changed
Hue Changed

Source


Add Blur
Original
Blurred
Increased Blur


Random Cropping

1 ===> 18

c.Data Augmentation

- Motivation: This helps the model become more robust and better at recognizing objects, even when they appear differently in real-world situations.

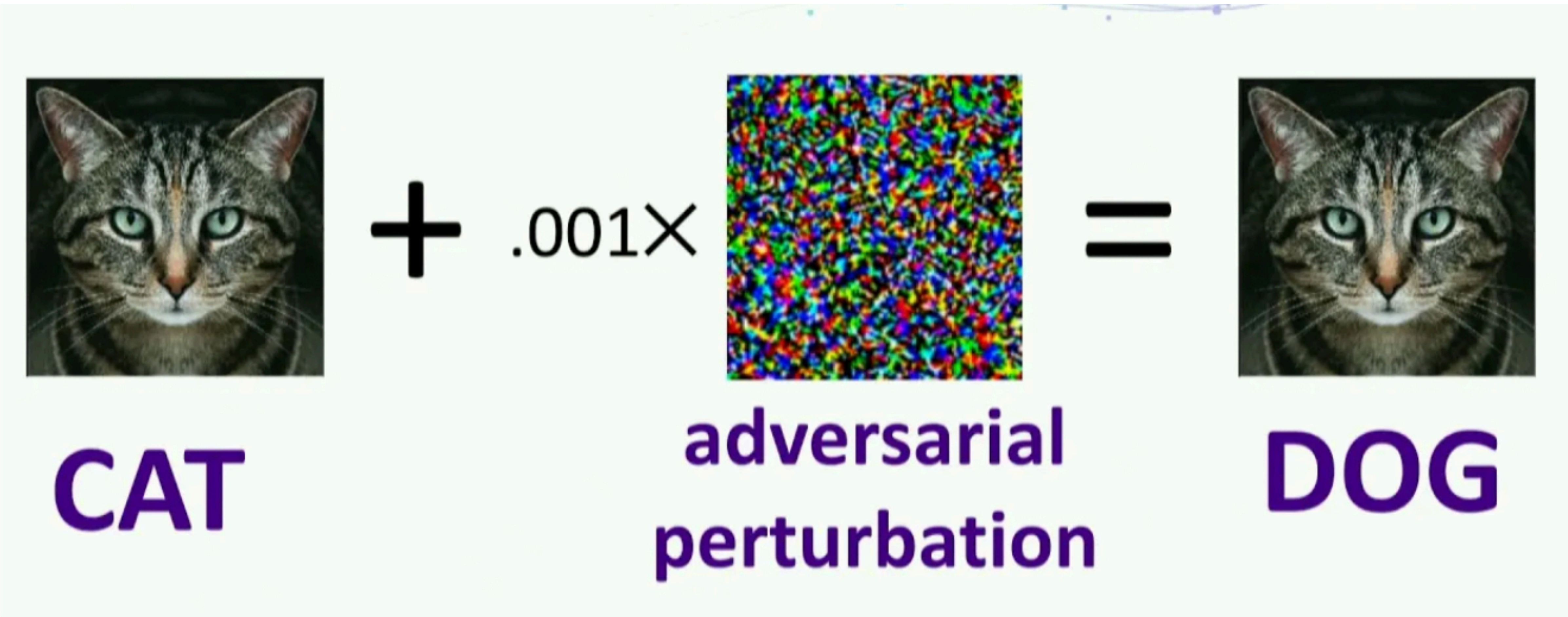
1 ===> 18

Improved Regularization of Convolutional Neural Networks with Cutout by DeVries and Taylor.

Mixup: Beyond Empirical Risk Minimization by Zhang et al.

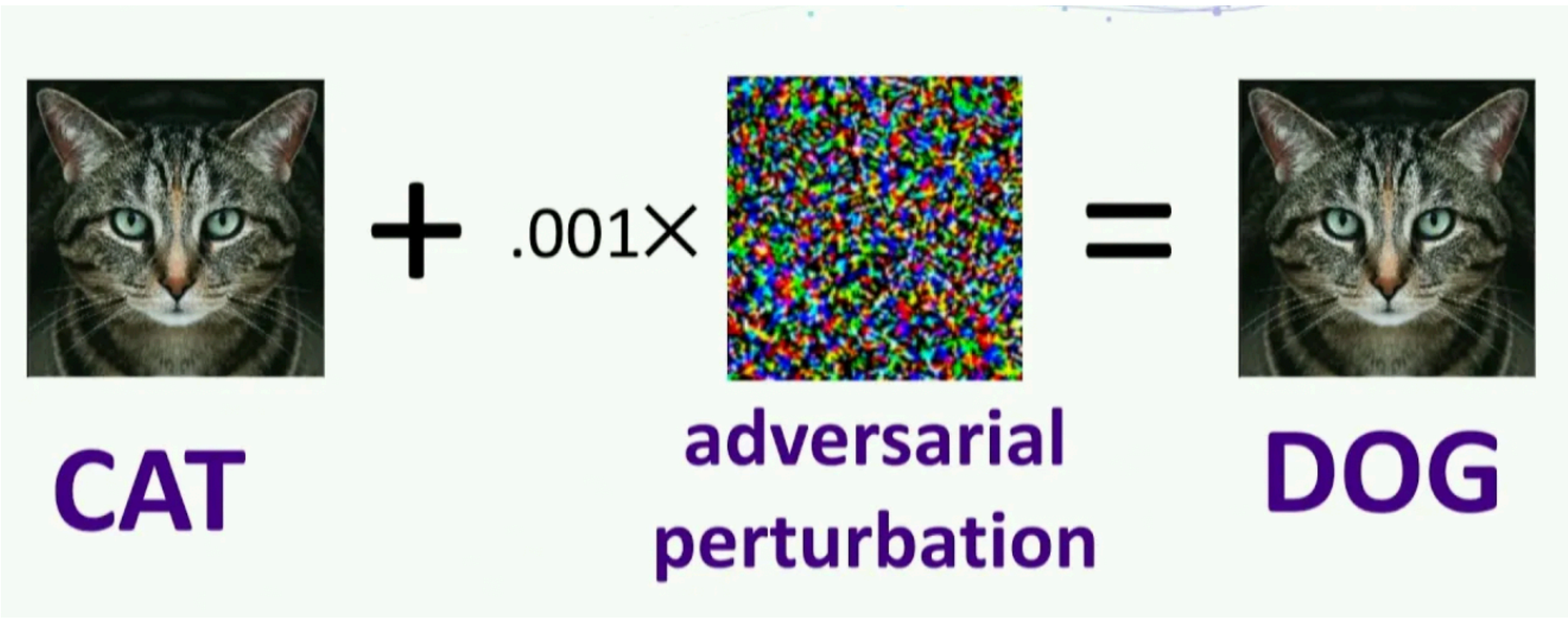
d. Adversarial Learning (Attack)

- Adversarial attack in a classification model is when someone tries to fool the model by slightly altering the input.
- For example, imagine we have a model that can classify cat and dog. In an adversarial attack, we might slightly adjust the pixels in a picture of a cat—so slightly that a human wouldn't even notice—but the model gets confused and wrongly classifies the cat as a dog.



d. Adversarial Learning (Attack)

- Adversarial attack in a classification model is when someone tries to fool the model by slightly altering the input.
- For example, imagine we have a model that can classify cat and dog. In an adversarial attack, we might slightly adjust the pixels in a picture of a cat—so slightly that a human wouldn't even notice—but the model gets confused and wrongly classifies the cat as a dog.



Want to know how to build this perturbation?
Check this out.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets.

Examples of Adversarial Attack in real world : Self-Driving

Figure 5. The attacker puts some stickers on a road sign to confuse an autonomous vehicle's road sign recognizer from any viewpoint. (Image Credit: (Eykholt et al., 2017))

https://arxiv.org/pdf/1909.08072

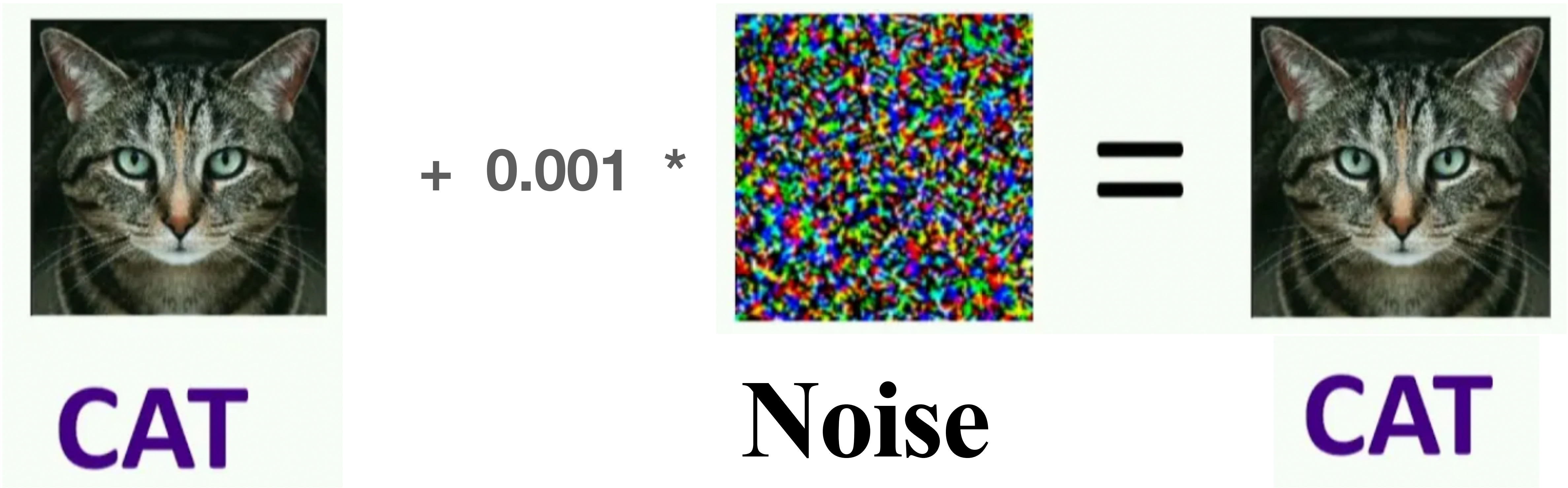
Figure 2: Illustration of our novel and domain-specific attack vector: Dirty Road Patch (DRP).

Dirty Road Attack

https://www.usenix.org/system/files/sec21-sato.pdf

d. Adversarial Learning (Training)

- Adversarial training is a technique used to make models more robust against adversarial attacks. In a cat-and-dog classifier, we **intentionally add** small, tricky changes to the training images—so the model learns to handle these attacks.



Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets.

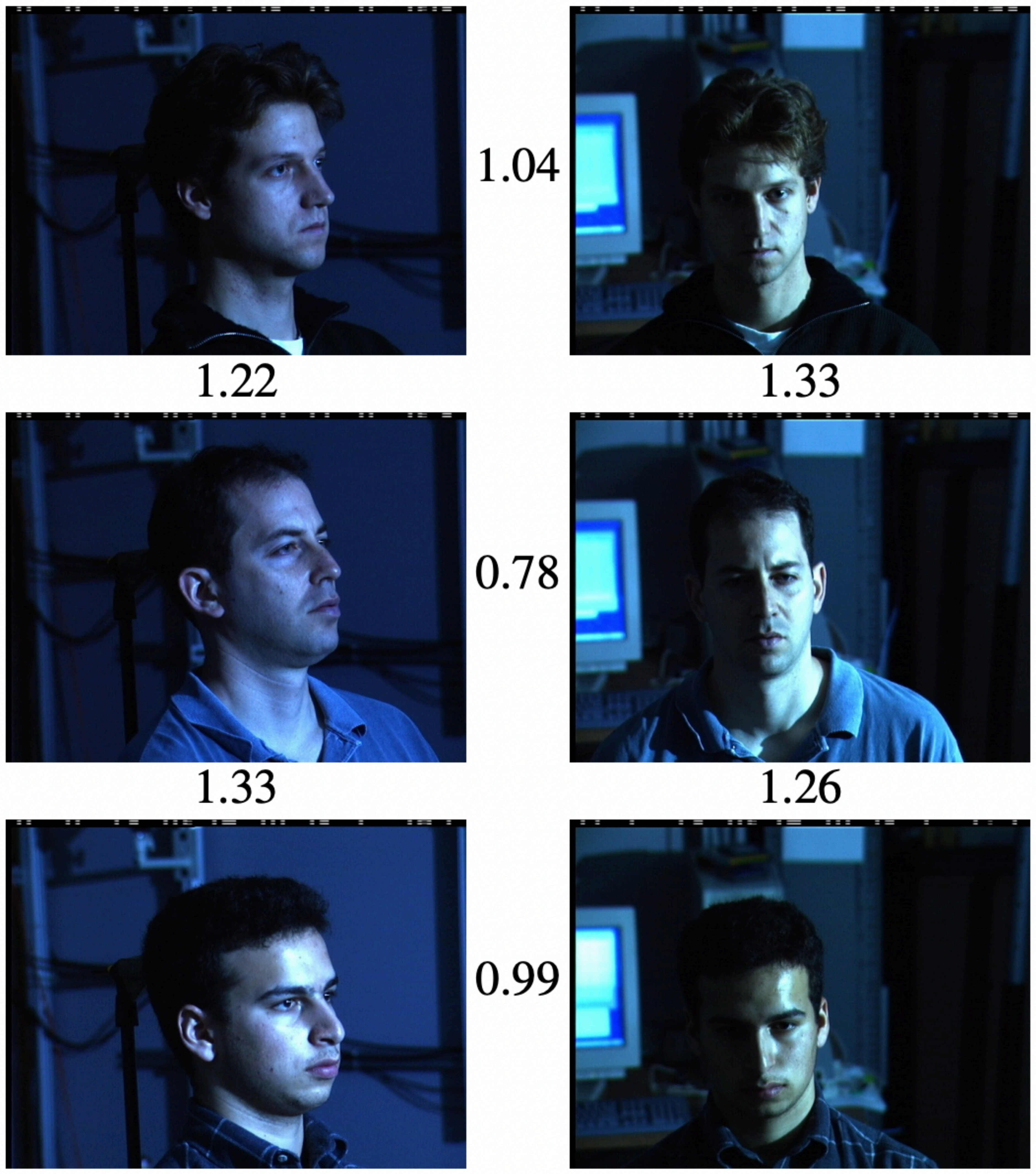
Certified Adversarial Robustness via Randomized Smoothing

c.Contrastive Learning

- Contrastive learning is a way to train models by teaching them to tell apart similar and different things.
  - *Example : FaceNet: A Unified Embedding for Face Recognition and Clustering*
- Contrastive learning can also be used to connect text and images….
  - *CLIP (Contrastive Language-Image Pretraining)*

$$loss = \|F(Cat_1) - F(Cat_2)\| - \|F(Cat_1) - F(Dog_1)\|$$

A Simple Framework for Contrastive Learning of Visual Representations (SimCLR) by Chen et al.
Momentum Contrast for Unsupervised Visual Representation Learning (MoCo) by He et al.



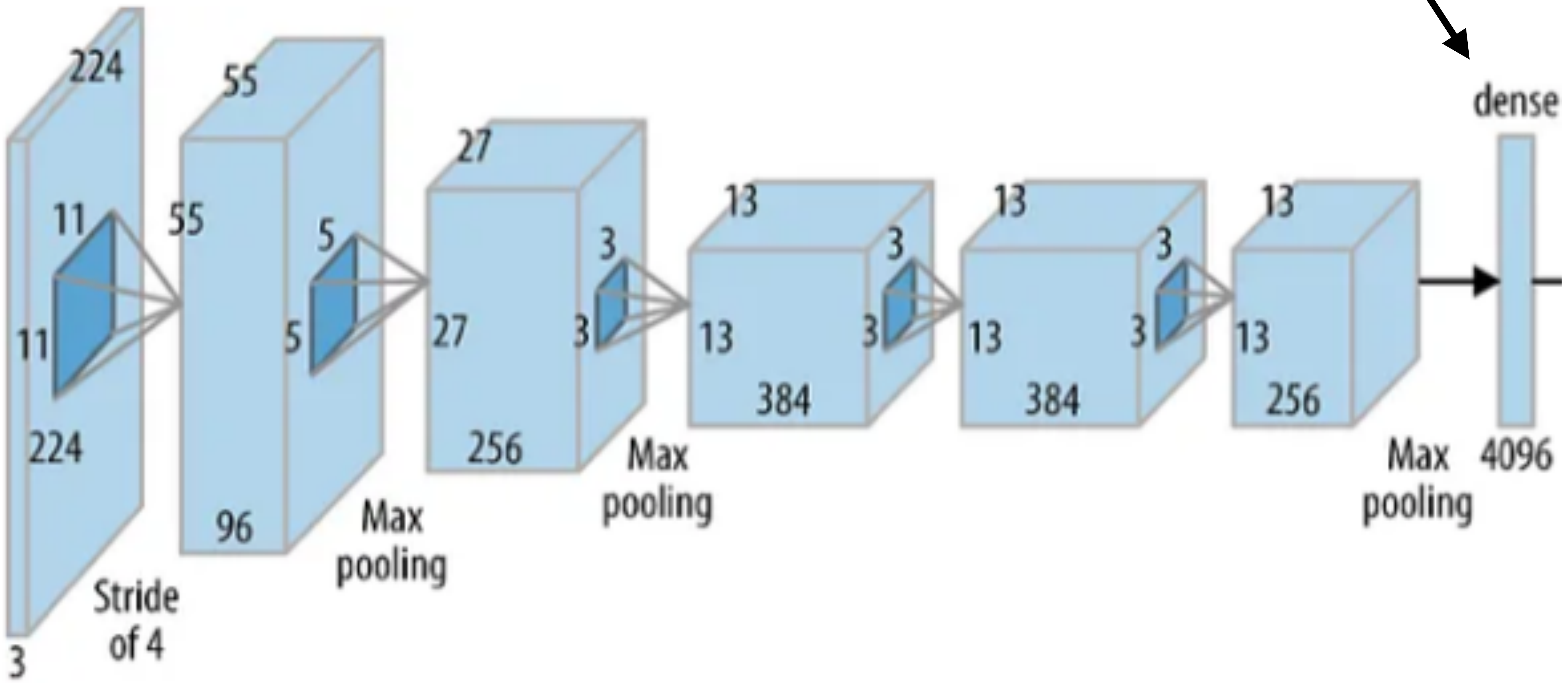
Jack

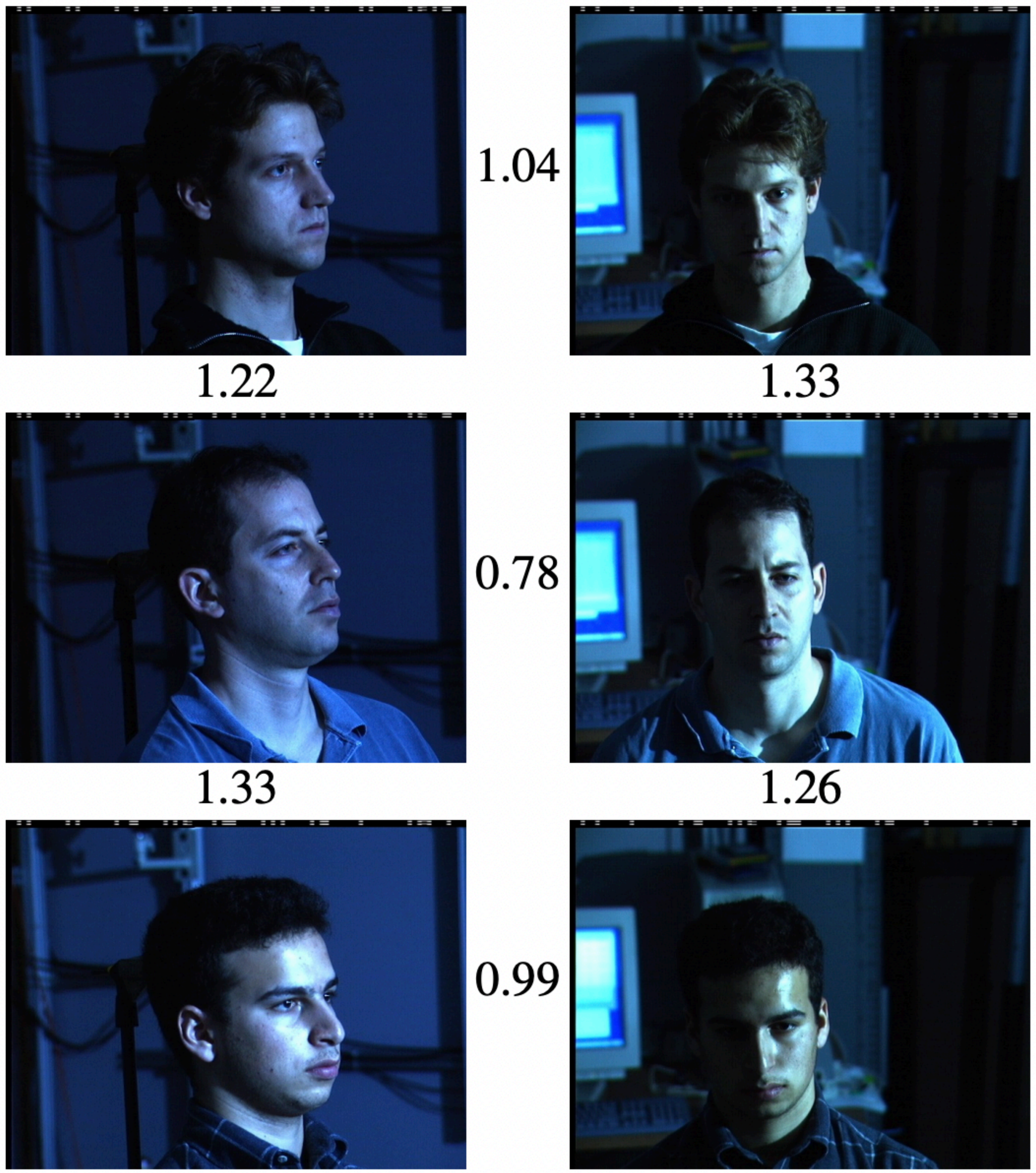
Bob

Tim

https://arxiv.org/pdf/1503.03832

FaceNet aims to keep the same people close in **latent space** and different people far apart.





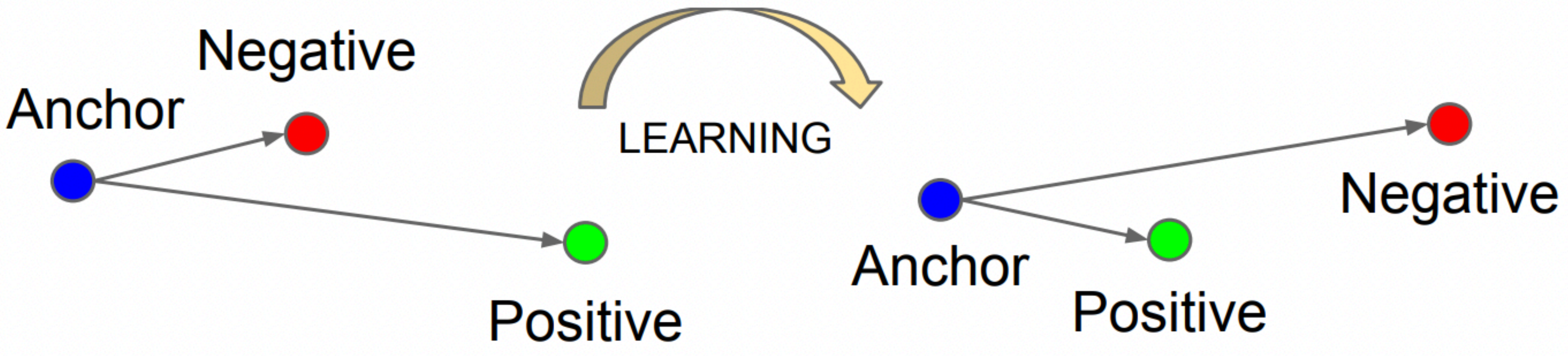


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

$$\|F(Jack_1) - F(Jack_2)\| - \|F(Jack_1) - F(Bob_1)\|$$

https://arxiv.org/pdf/1503.03832

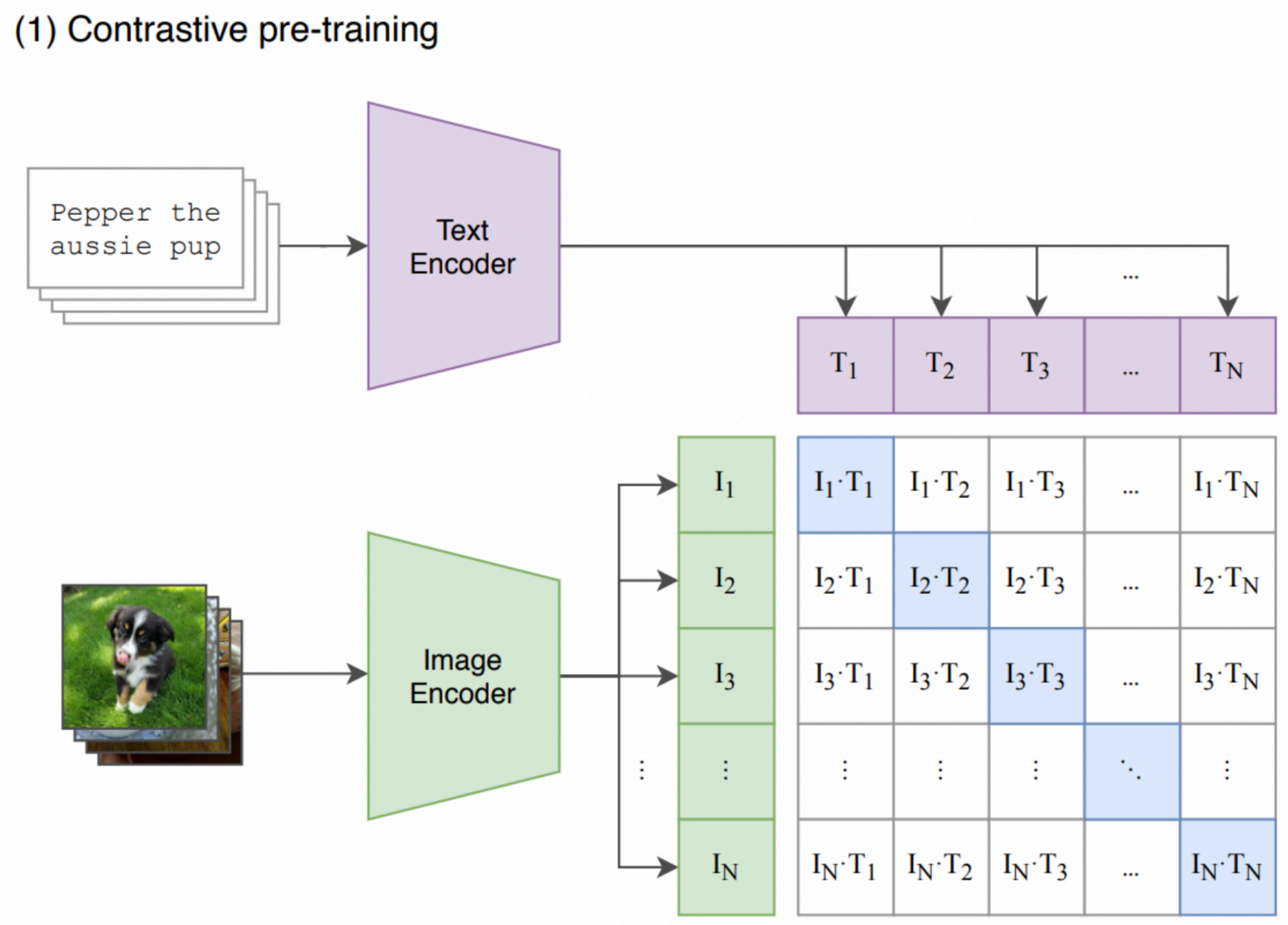
c.Contrastive Learning

- Contrastive learning can also be used to connect text and images. For example, we can show the model the word 'cat' along with a picture of a cat, and the word 'dog' with a picture of a dog. The model learns that the word 'cat' matches the image of a cat, and 'dog' matches the image of a dog. If we later show it a new picture of a dog, it should correctly link it to the word 'dog.' This way, the model gets better at connecting words and images that go together.

A Simple Framework for Contrastive Learning of Visual Representations (SimCLR) by Chen et al.

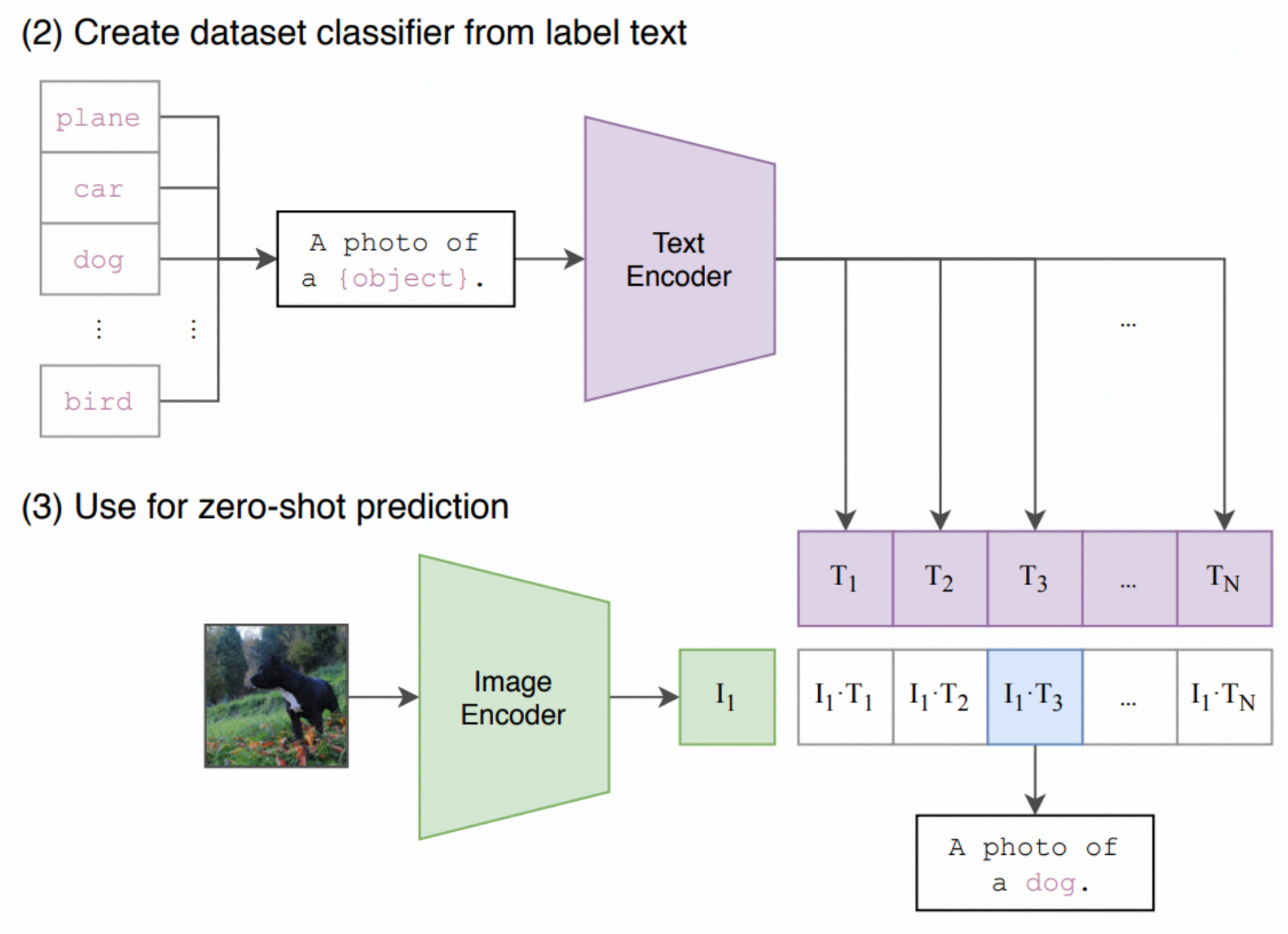
Momentum Contrast for Unsupervised Visual Representation Learning (MoCo) by He et al.

CLIP (Contrastive Language-Image Pretraining) 2021-2022



Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *OpenAI*.

CLIP (Contrastive Language-Image Pretraining) 2021-2022



Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *OpenAI*.

Unsupervised Learning

Unsupervised learning focuses on discovering patterns in data without the need for labeled datasets.

a.AutoEncoders

- Autoencoders are unsupervised models that learn to encode data into a lower-dimensional representation and then reconstruct it. This has been used for tasks like dimensionality reduction, anomaly detection, and feature learning.

$$loss = \|D(E(X_i)) - X_i\|$$

Auto-Encoding Variational Bayes by Kingma and Welling.

An Introduction to Autoencoders for Deep Learning by Goodfellow et al.

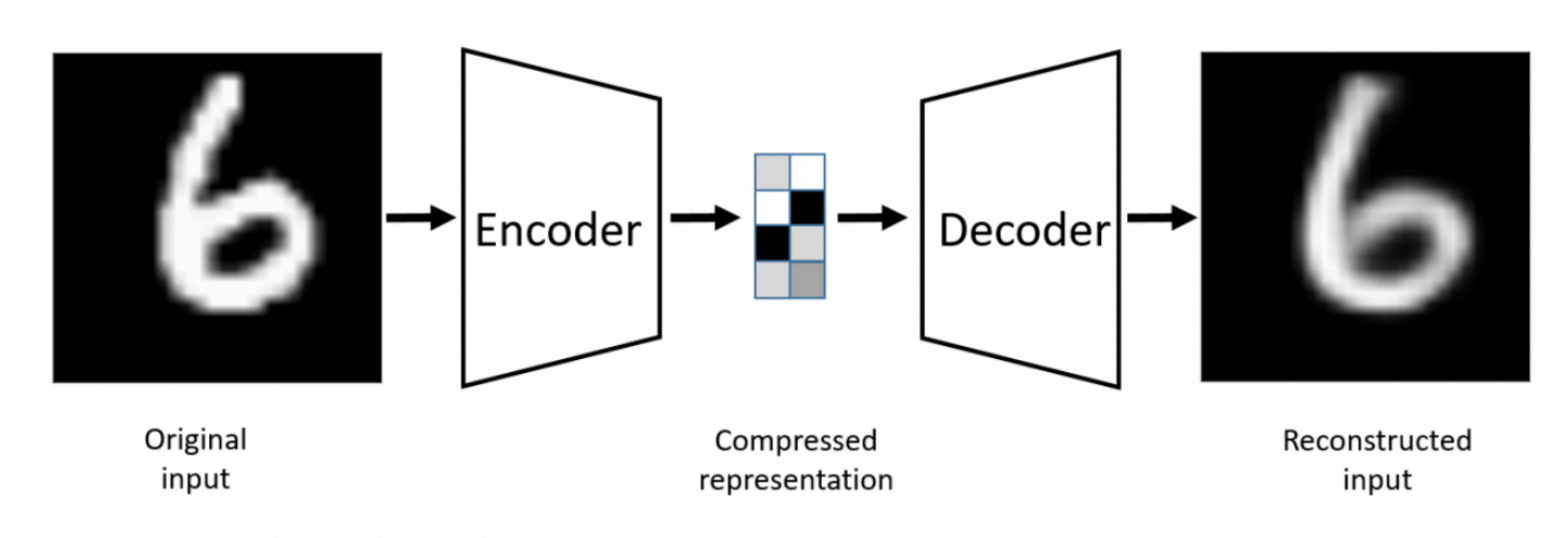


Image Source

$$loss = \|F_{AE}(Image6_i) - Image6_i\|$$

b.Generative Adversarial Networks (GANs)

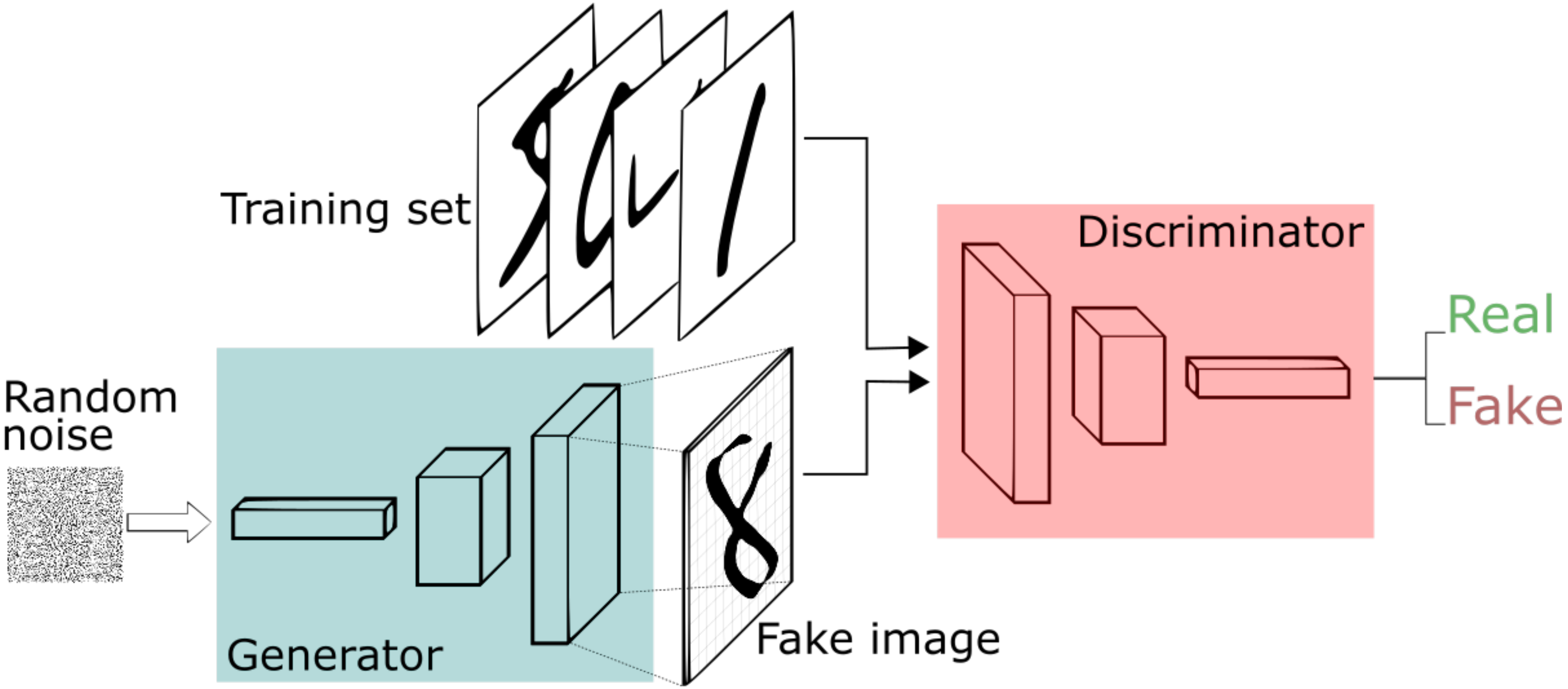
- GANs represent one of the most successful unsupervised learning approaches. Introduced by Ian Goodfellow in 2014, GANs use two networks: a generator and a discriminator, which compete against each other to produce realistic data.

- GANs have been widely adopted for tasks like image generation, style transfer, and data augmentation, leading to improved generative modeling techniques.

Generative Adversarial Nets by Ian Goodfellow et al.

Improved Techniques for Training GANs by Salimans et al.

Image Source



$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$$

d.Self-Supervised Learning

- Self-supervised learning is a method where the model learns from the **data itself** without needing labeled examples.
    - Example.a - in large language model (LLM) training, the model learns by predicting missing words in a sentence. It sees part of a sentence, like 'The cat ___ on the mat,' and learns to fill in the blank
    - Example.b - in MAE (Masked Autoencoder) training for images, part of an image is masked, and the model learns to reconstruct the missing part.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding by Devlin et al.

SimCLR: A Simple Framework for Contrastive Learning of Visual Representations by Chen et al.

Auto-regression (Decoder-only LLM, GPT3) -2021 Image Source

Encoder-only *Decoder-only*

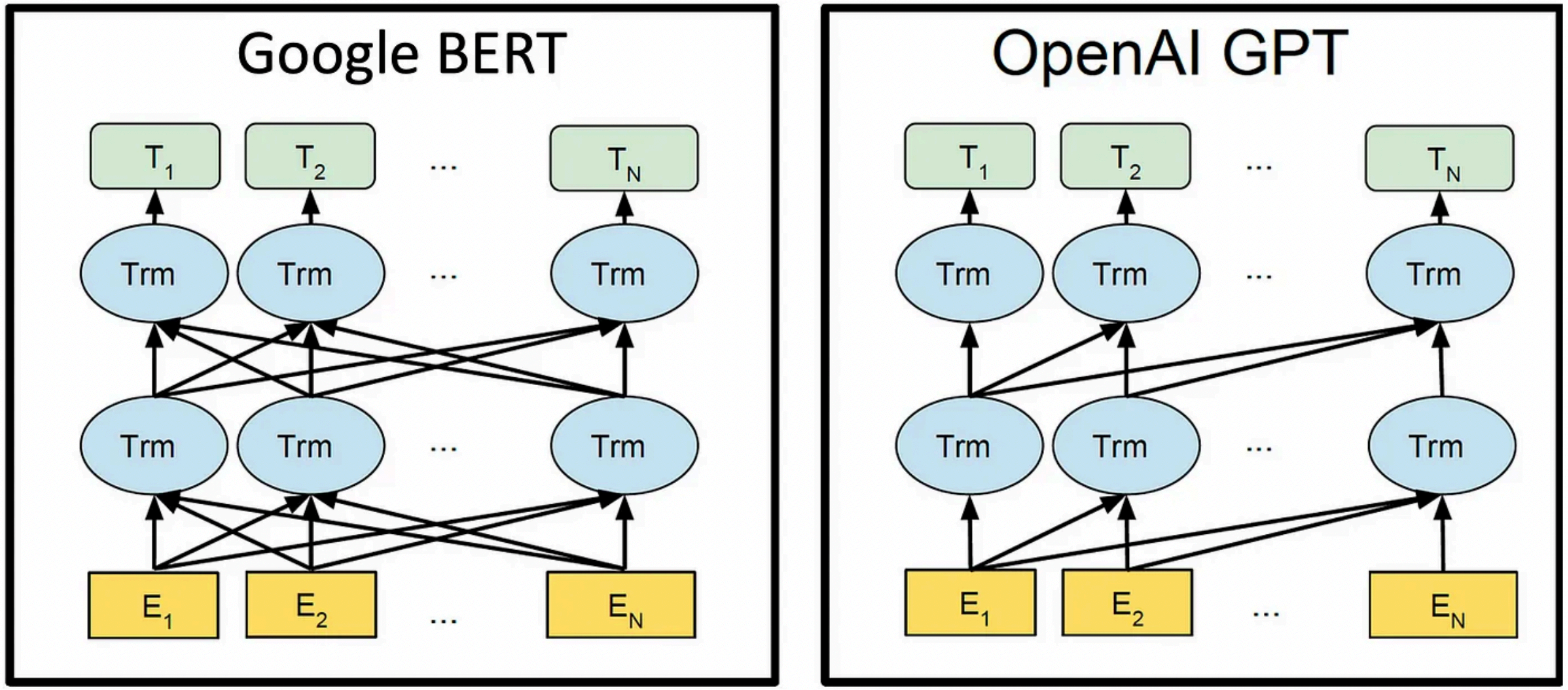
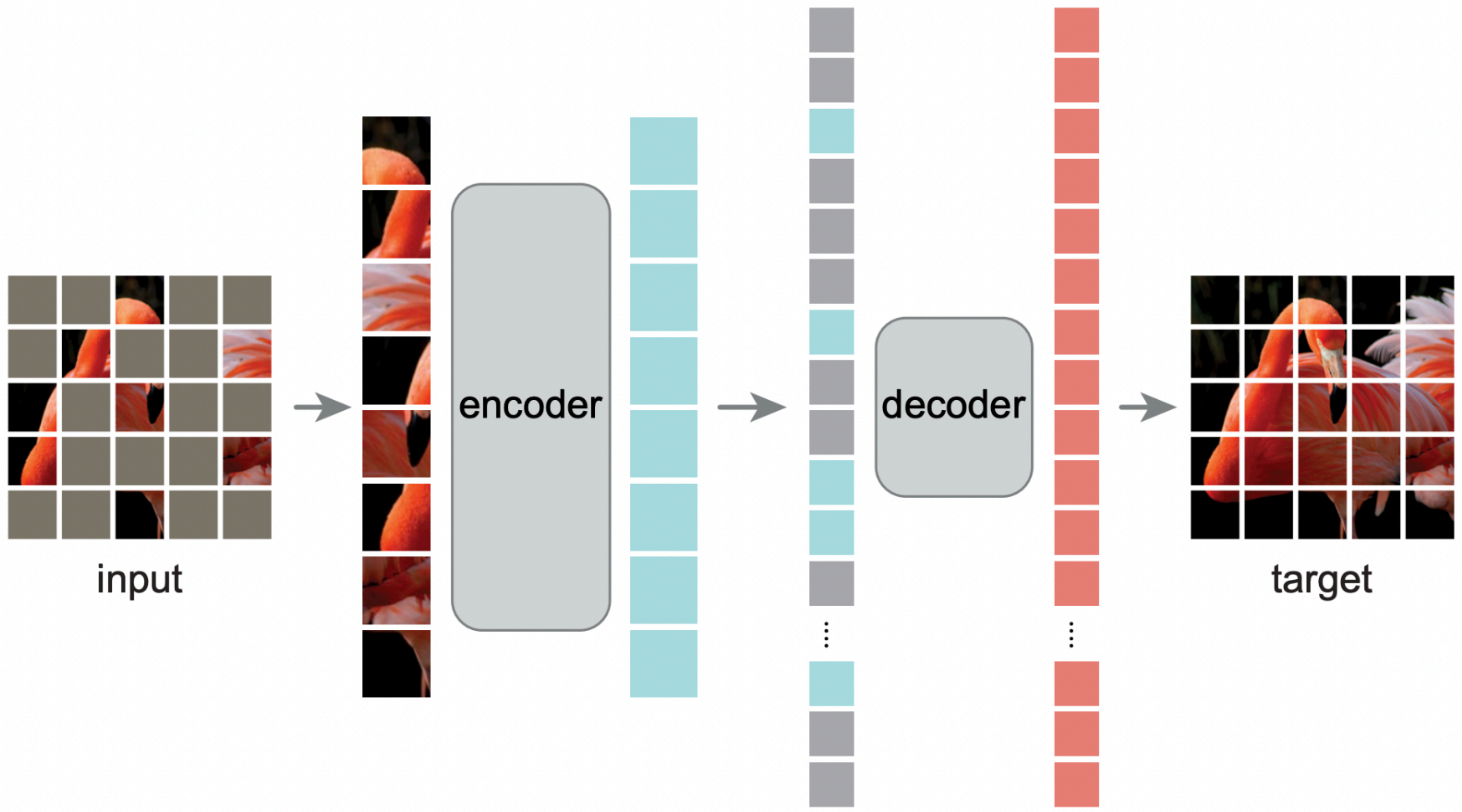


Fig. 2: BERT vs GPT — typical autoencoding (AE), encoder-only, bidirectional model vs typical autoregressive (AR), decoder-only, left-to-right model. Note that this figure is only qualitatively correct. To accurately understand transformer encoder and decoder structures, please read my blog “Step-by-Step Illustrated Explanations of Transformer”. (Image Source: Devlin, et. al., 2018)

By predicting the "next" or "middle" word, it learns the distribution of natural language.

Masked autoencoders are scalable vision learners - 2021



By learning to predict different parts of an image, it learns the image's distribution.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked autoencoders are scalable vision learners.

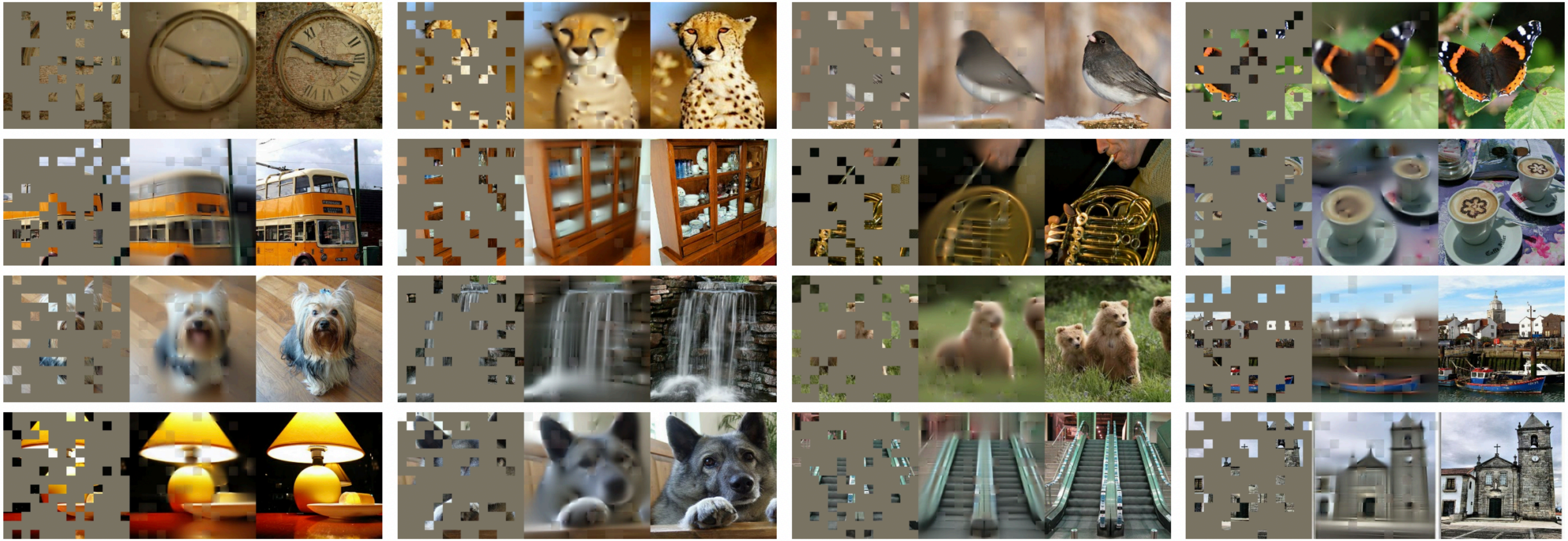


Figure 2. Example results on ImageNet *validation* images. For each triplet, we show the masked image (left), our MAE reconstruction† (middle), and the ground-truth (right). The masking ratio is 80%, leaving only 39 out of 196 patches. More examples are in the appendix.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked autoencoders are scalable vision learners.

d.Self-Supervised Learning

- **Advantage**: Both methods allow the model to understand patterns and structure in the data without needing explicit labels, making the training more scalable and flexible.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding by Devlin et al.

SimCLR: A Simple Framework for Contrastive Learning of Visual Representations by Chen et al.

Thanks