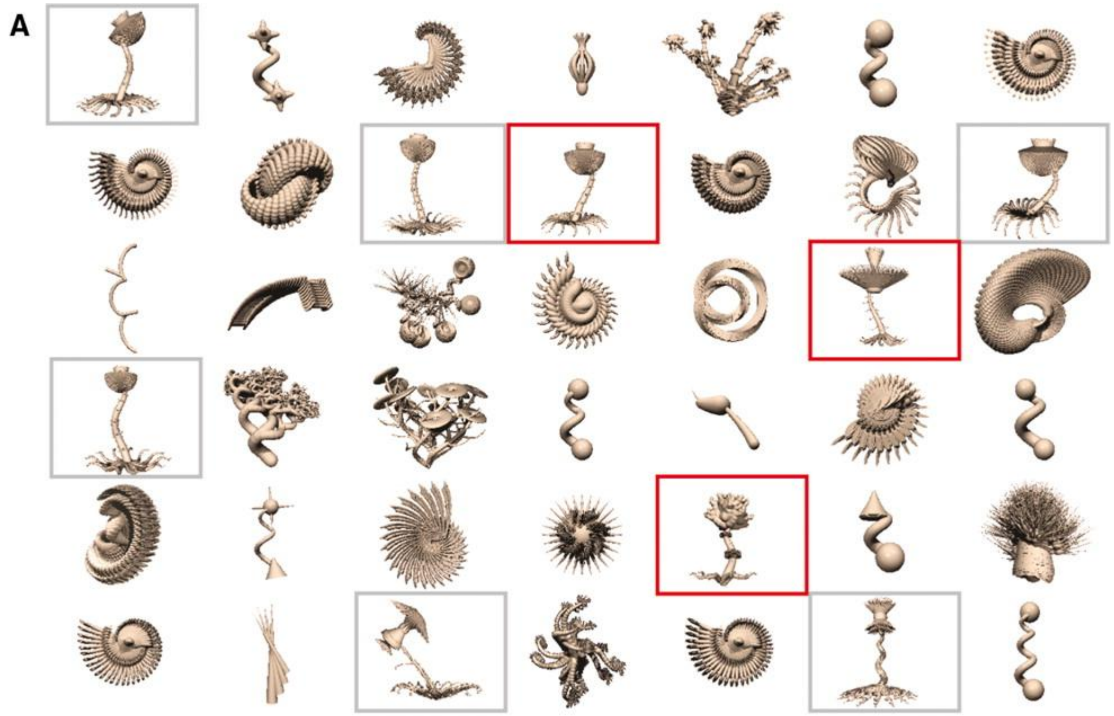
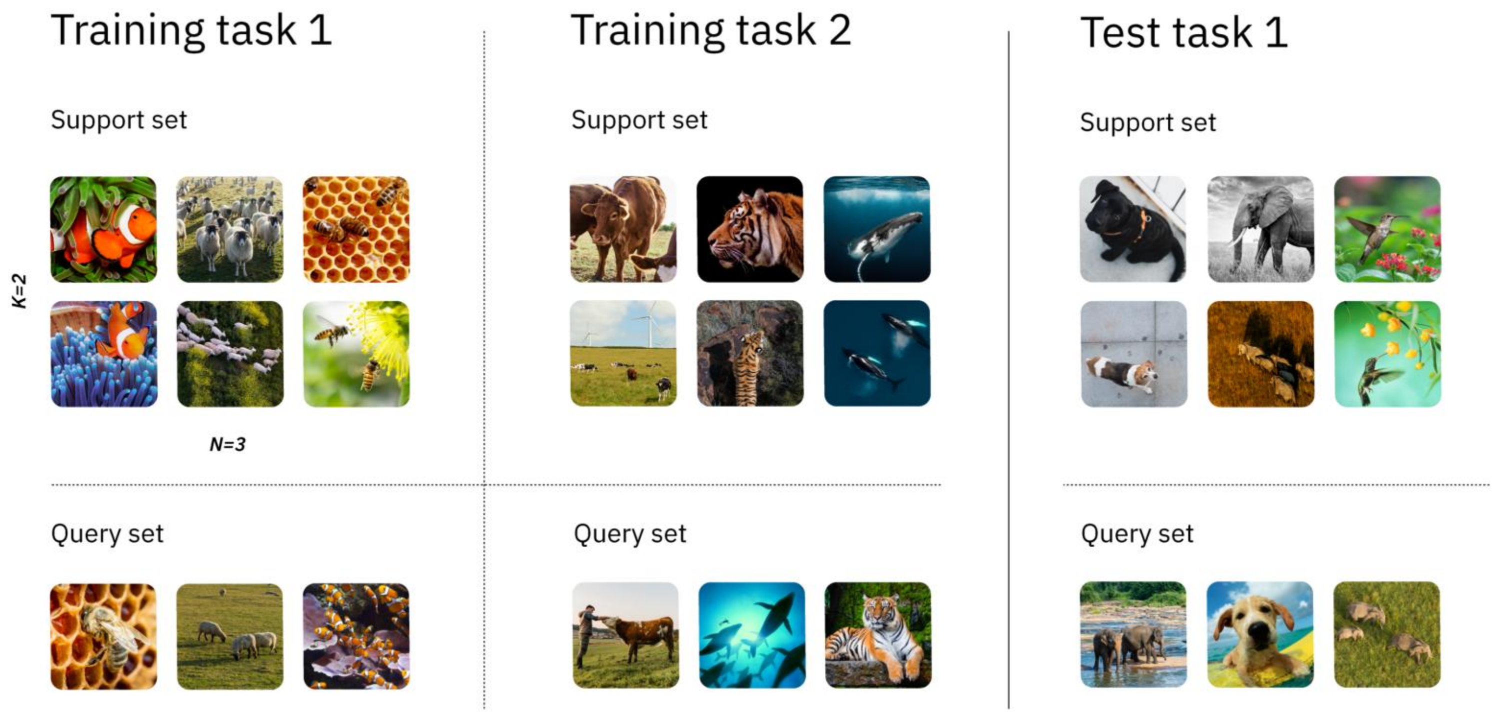
Few-Shot Learning

Unsupervised Domain Adaptation



Tenenbaum, Joshua B., et al. "How to grow a mind: Statistics, structure, and abstraction." *science* 331.6022 (2011): 1279-1285.

Few-Shot Learning



N-way-K-shot classification

3-way-2-shot

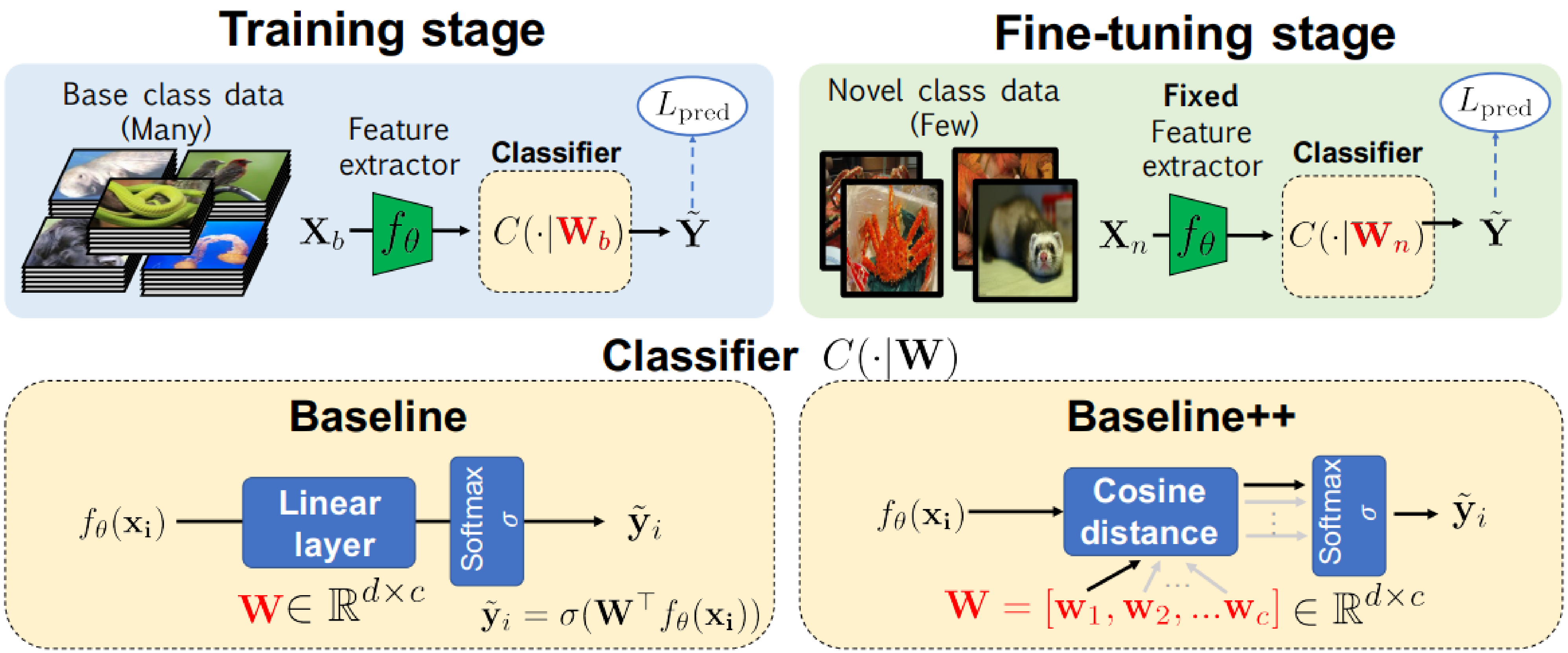


Figure 1: **Baseline and Baseline++ few-shot classification methods.** Both the baseline and baseline++ method train a feature extractor f_θ and classifier $C(.|\mathbf{W}_b)$ with base class data in the training stage In the fine-tuning stage, we fix the network parameters θ in the feature extractor f_θ and train a new classifier $C(.|\mathbf{W}_n)$ with the given labeled examples in novel classes. The baseline++ method differs from the baseline model in the use of cosine distances between the input feature and the weight vector for each class that aims to reduce intra-class variations.

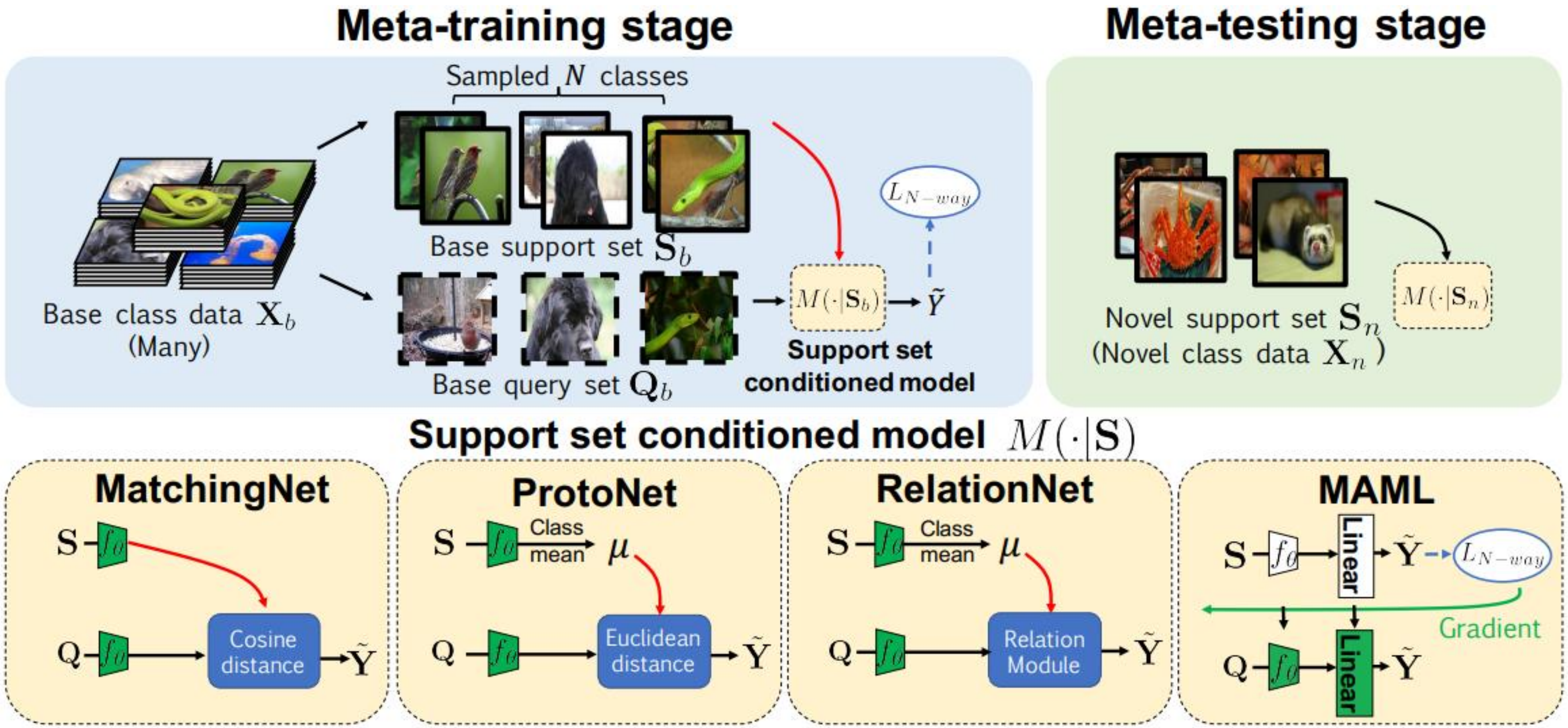
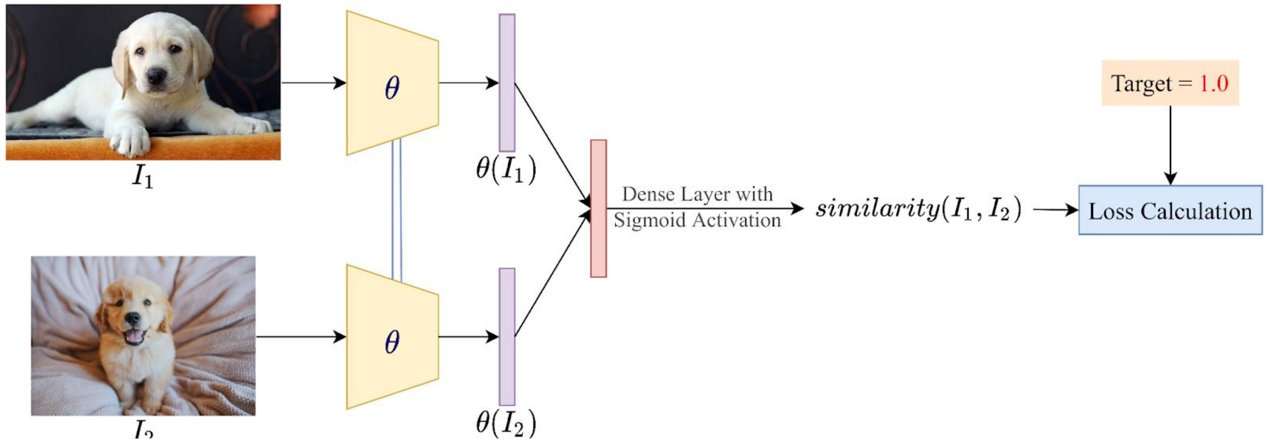
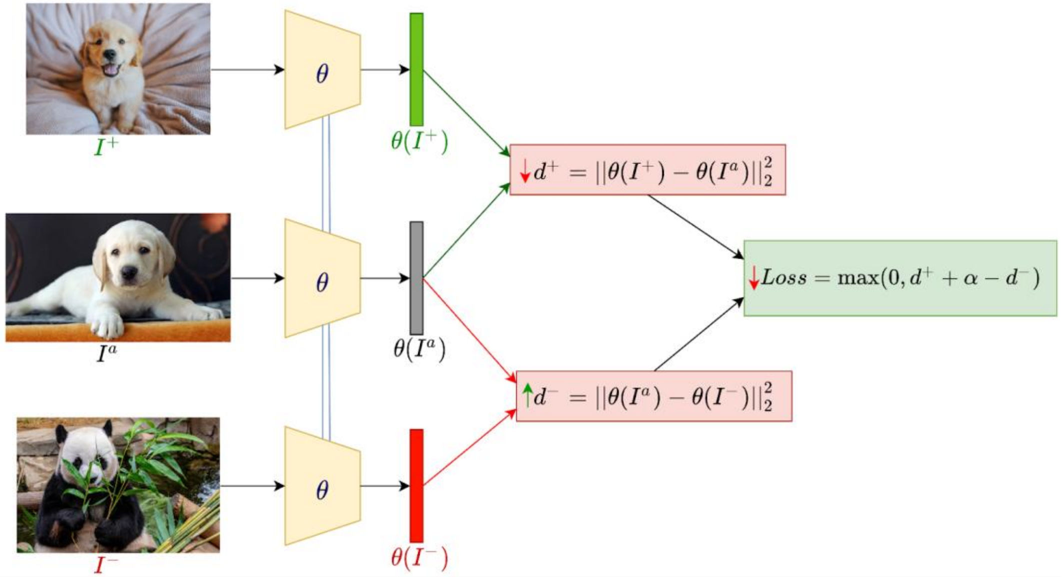


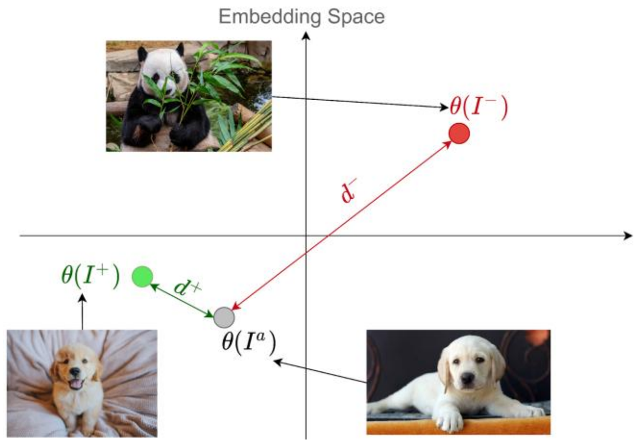
Figure 2: **Meta-learning few-shot classification algorithms.** The meta-learning classifier $M(\cdot|\mathbf{S})$ is conditioned on the support set $\mathbf{S}$. (*Top*) In the meta-train stage, the support set $\mathbf{S}_b$ and the query set $\mathbf{Q}_b$ are first sampled from random N classes, and then train the parameters in $M(.|\mathbf{S}_b)$ to minimize the N-way prediction loss $L_{\text{N-way}}$. In the meta-testing stage, the adapted classifier $M(.|\mathbf{S}_n)$ can predict novel classes with the support set in the novel classes $\mathbf{S}_n$. (*Bottom*) The design of $M(\cdot|\mathbf{S})$ in different meta-learning algorithms.

1 Metric Learning (learning to compare)

Siamese Neural Networks

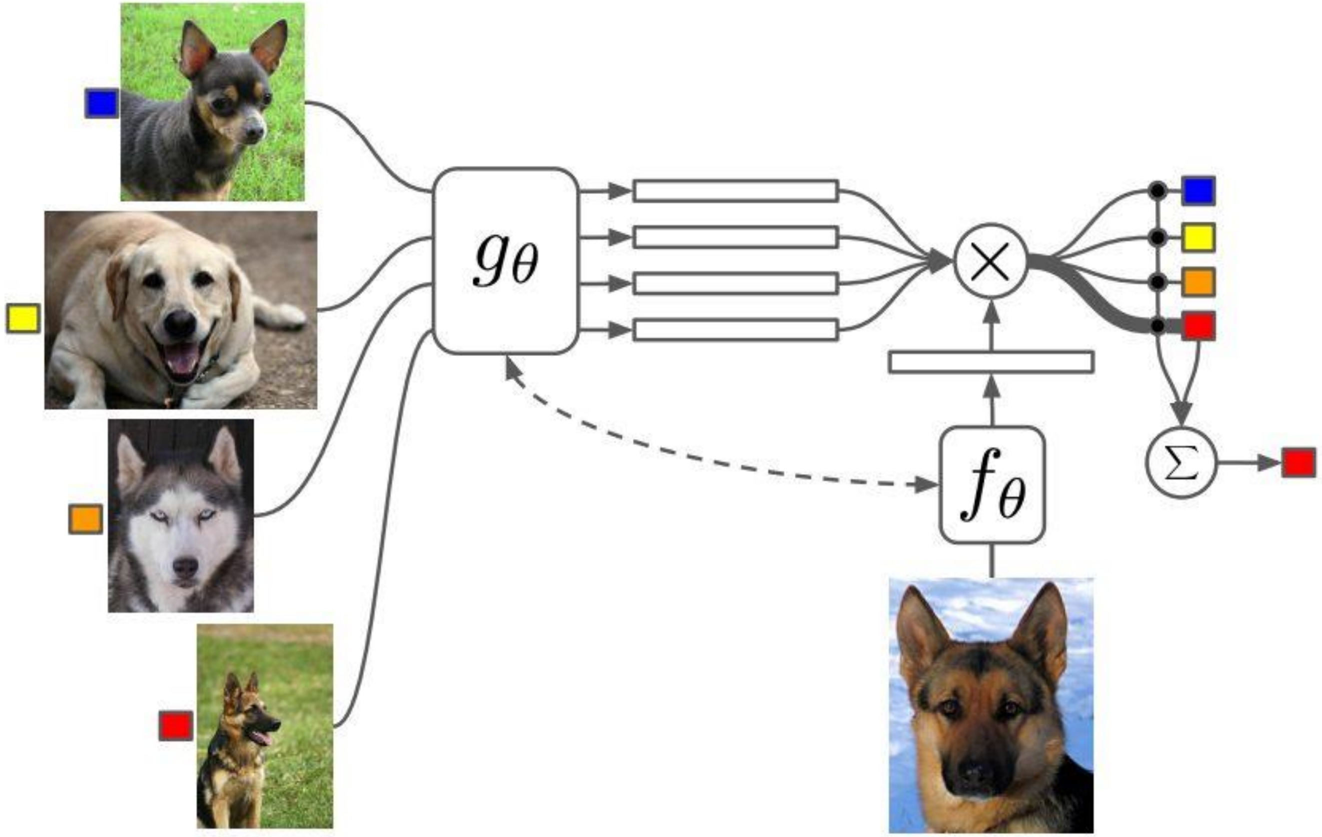






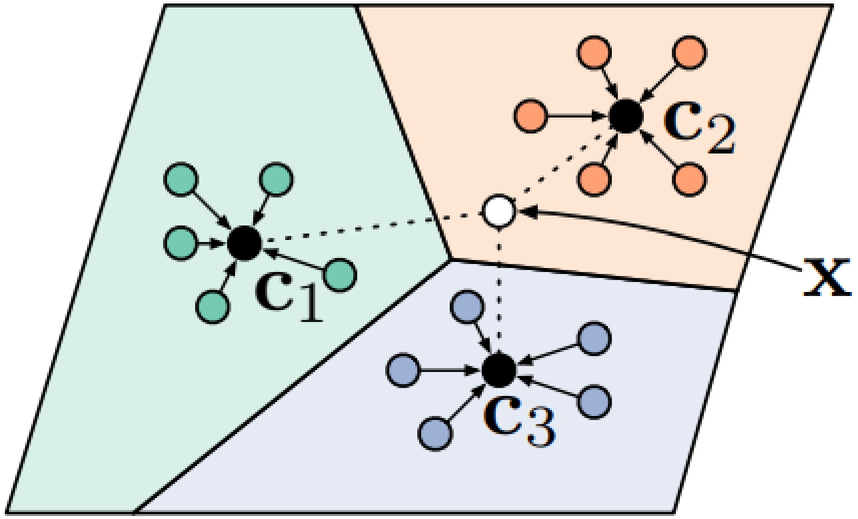
1 Metric Learning (learning to compare)

Matching Networks



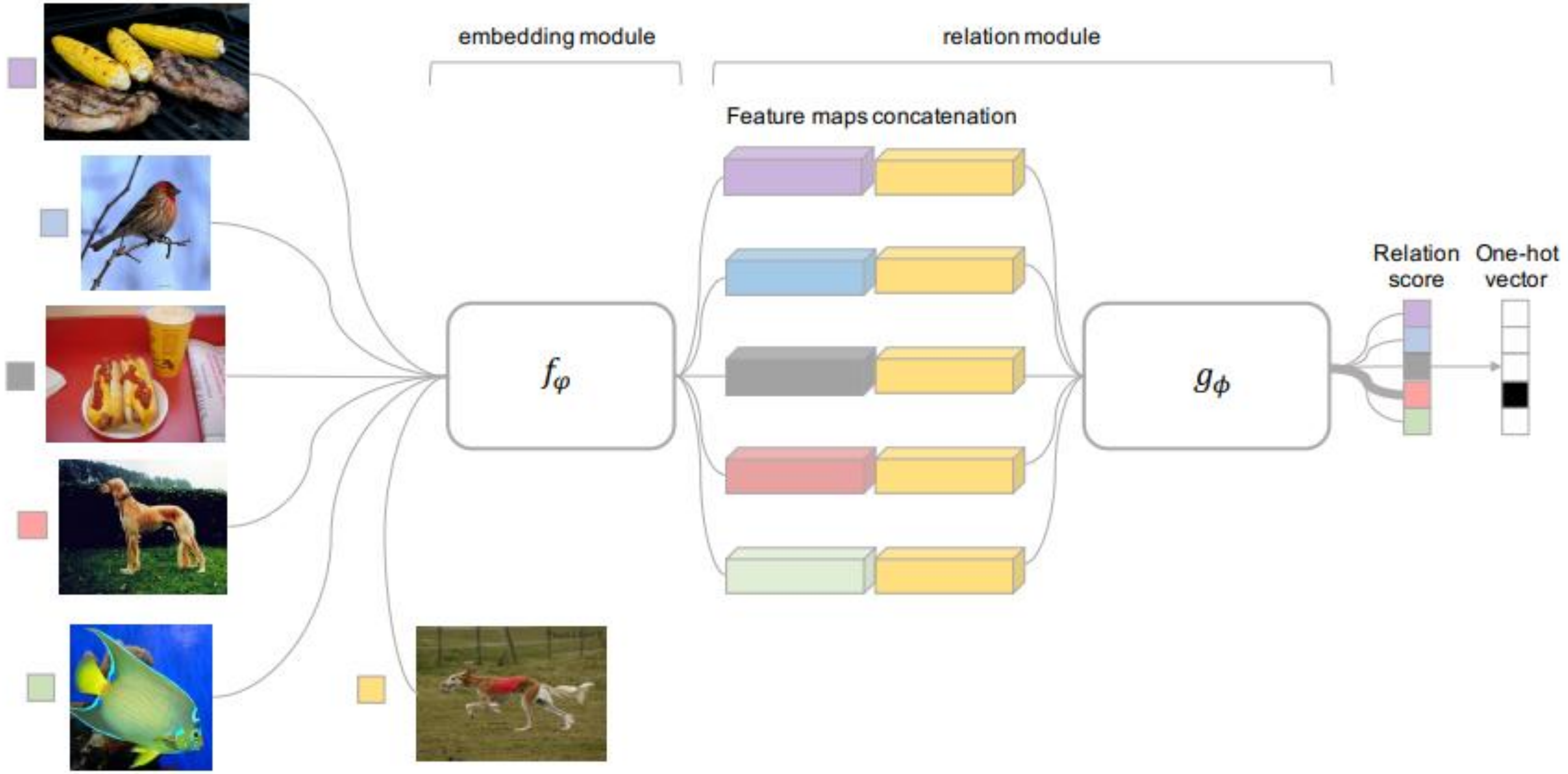
1 Metric Learning (learning to compare)

Prototypical Networks

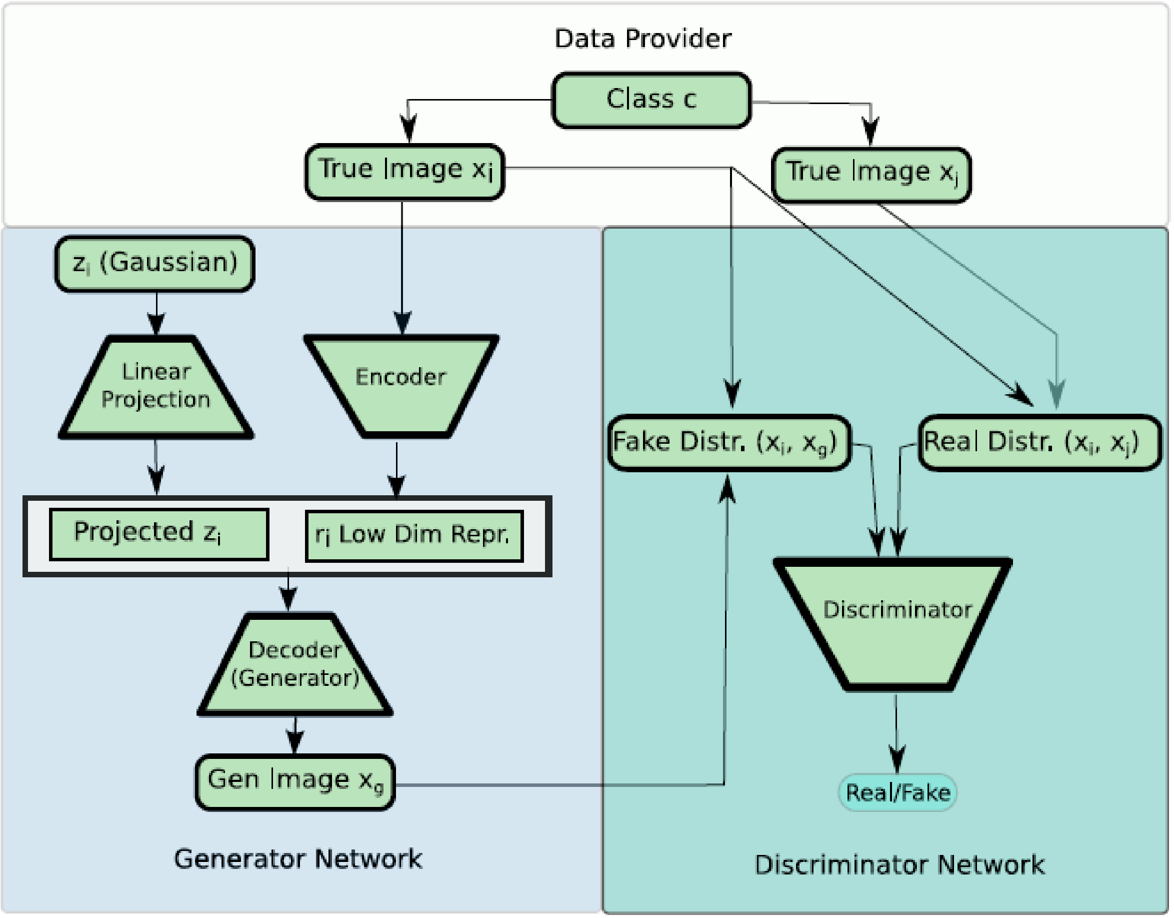


1 Metric Learning (learning to compare)

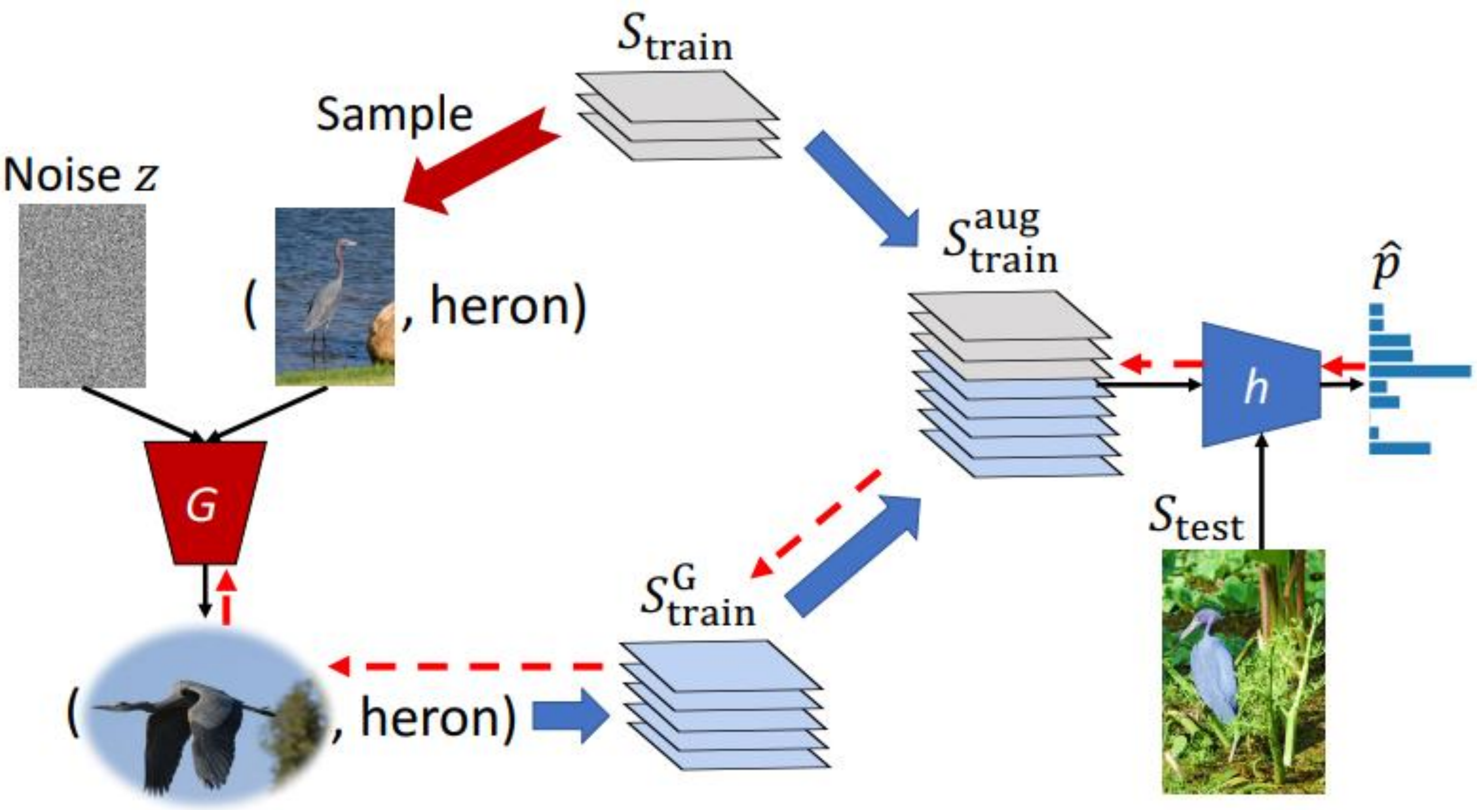
Relation Networks



2 Data augmentation methods (learning to augment)

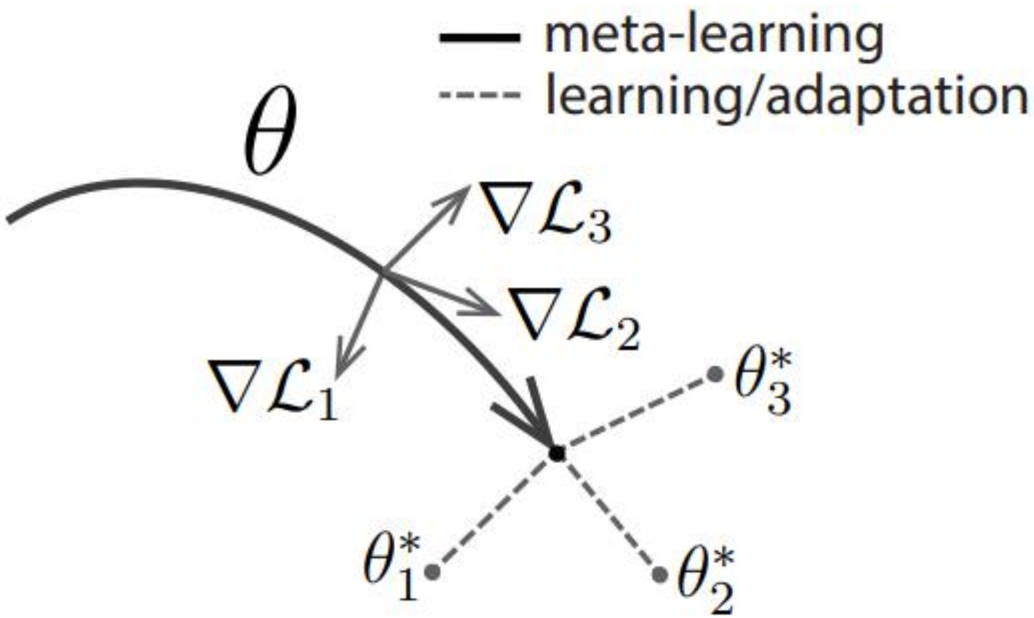


2 Data augmentation methods (learning to augment)



3 Meta-learning (learning to learn)

Model Agnostic Meta-Learning (MAML)

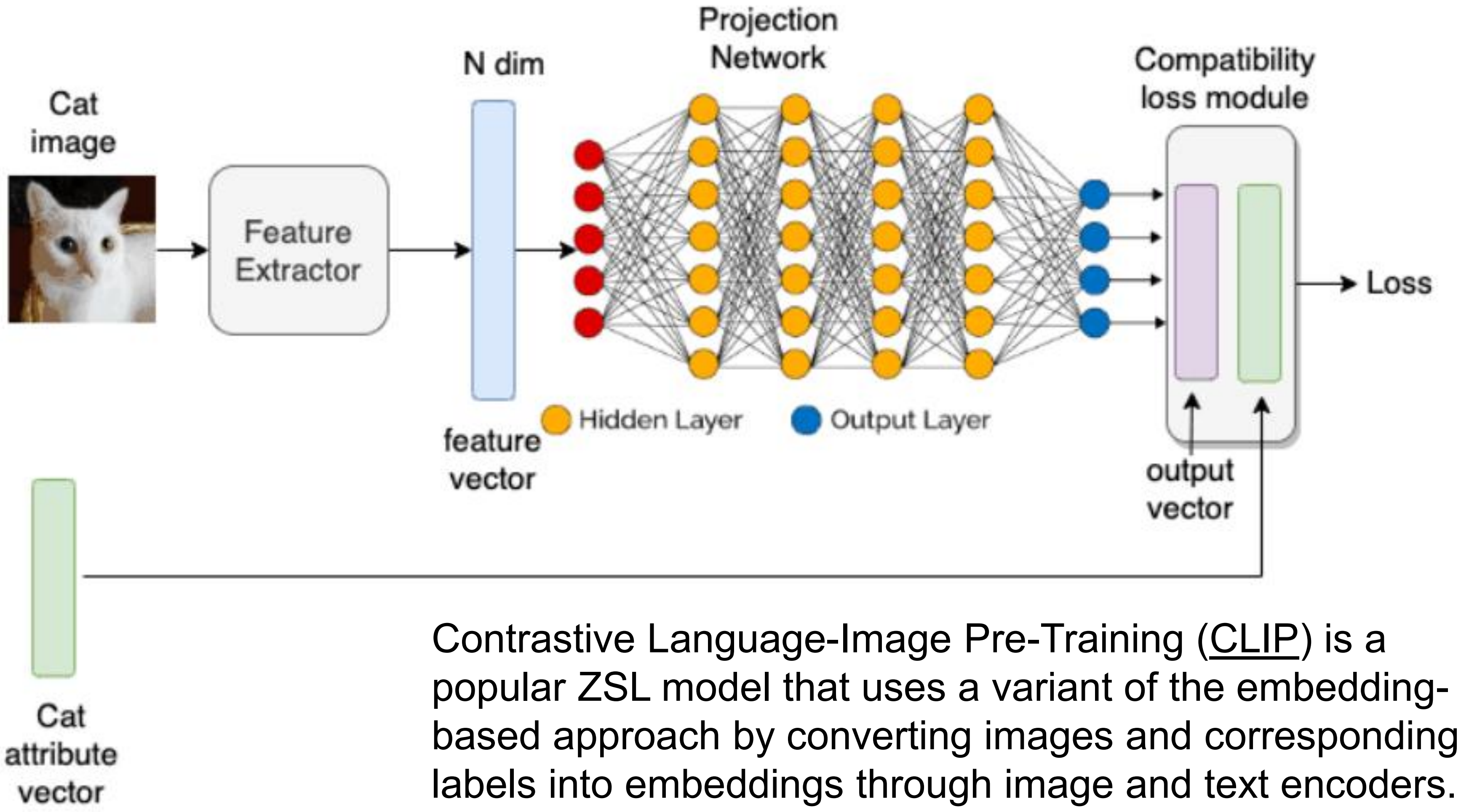


Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters

randomly initialize θ
while not done **do**
  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
  for all $\mathcal{T}_i$ **do**
    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
    Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
  end for
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
end while

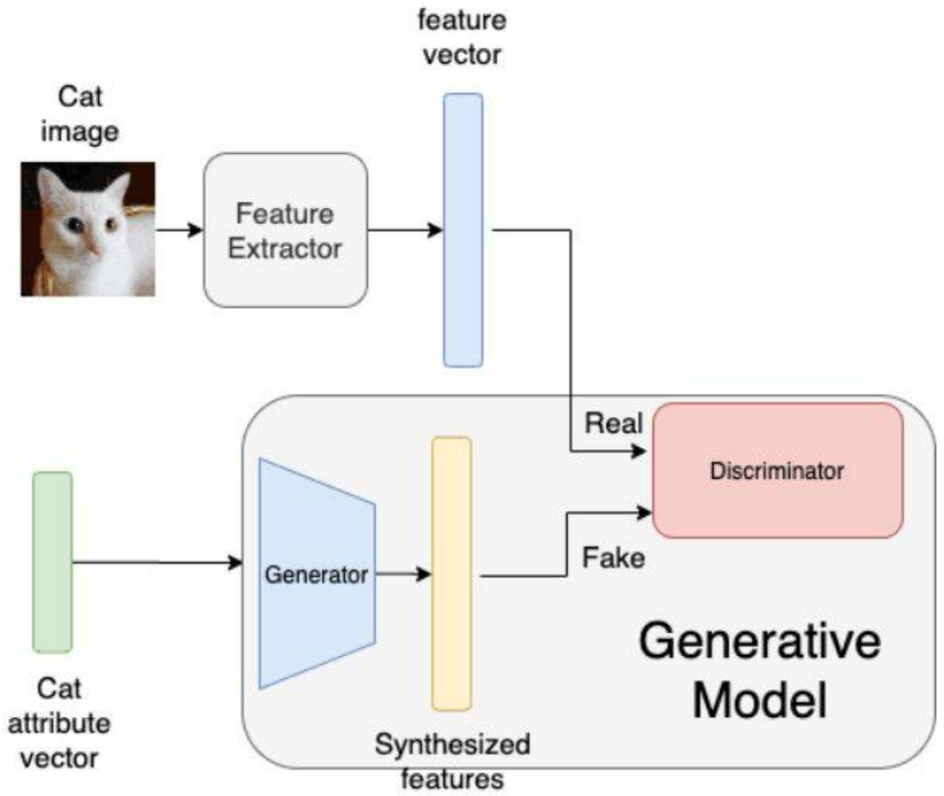
Zero-Shot Learning
Embedding-Based Approach



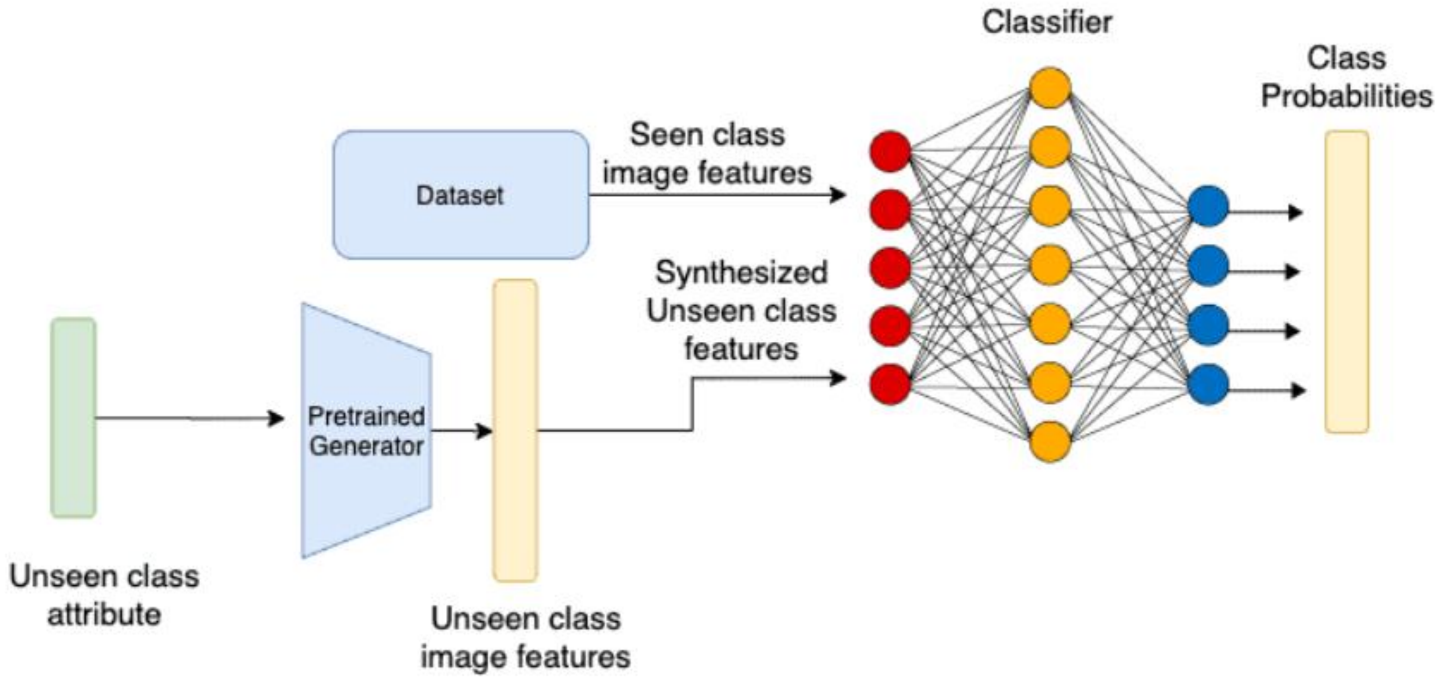
Contrastive Language-Image Pre-Training (CLIP) is a popular ZSL model that uses a variant of the embedding-based approach by converting images and corresponding labels into embeddings through image and text encoders.

Zero-Shot Learning

Generative-Based Approach



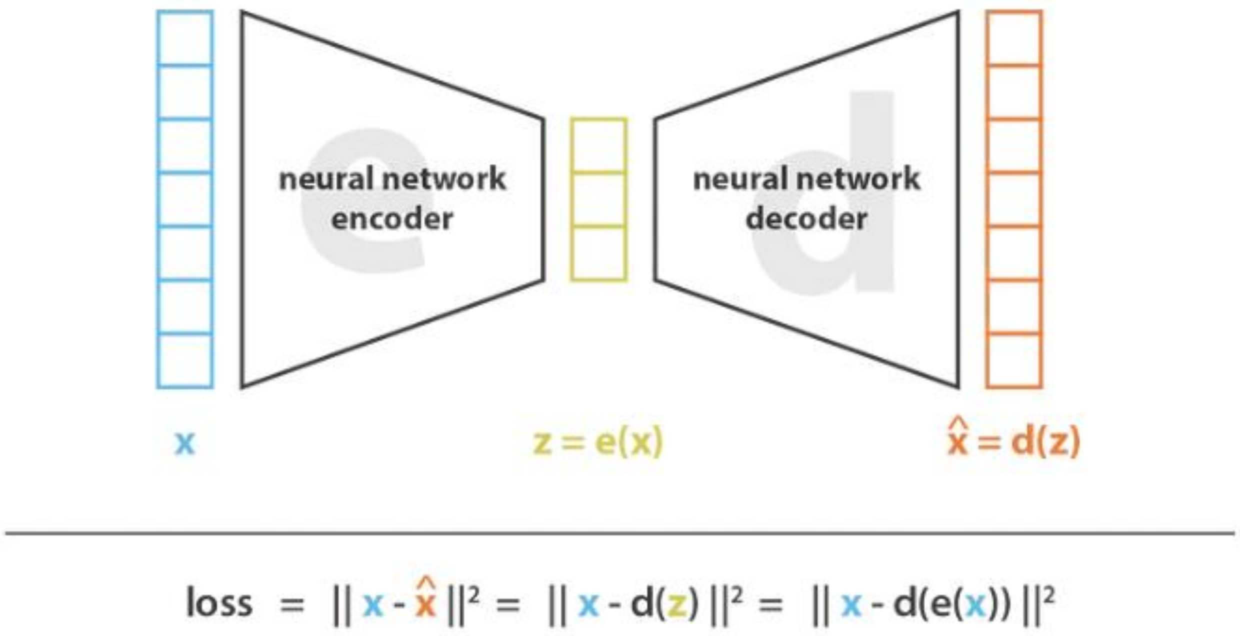
GANs: Training the Generator



GANs: Using the Generator to create synthetic feature vectors

Zero-Shot Learning

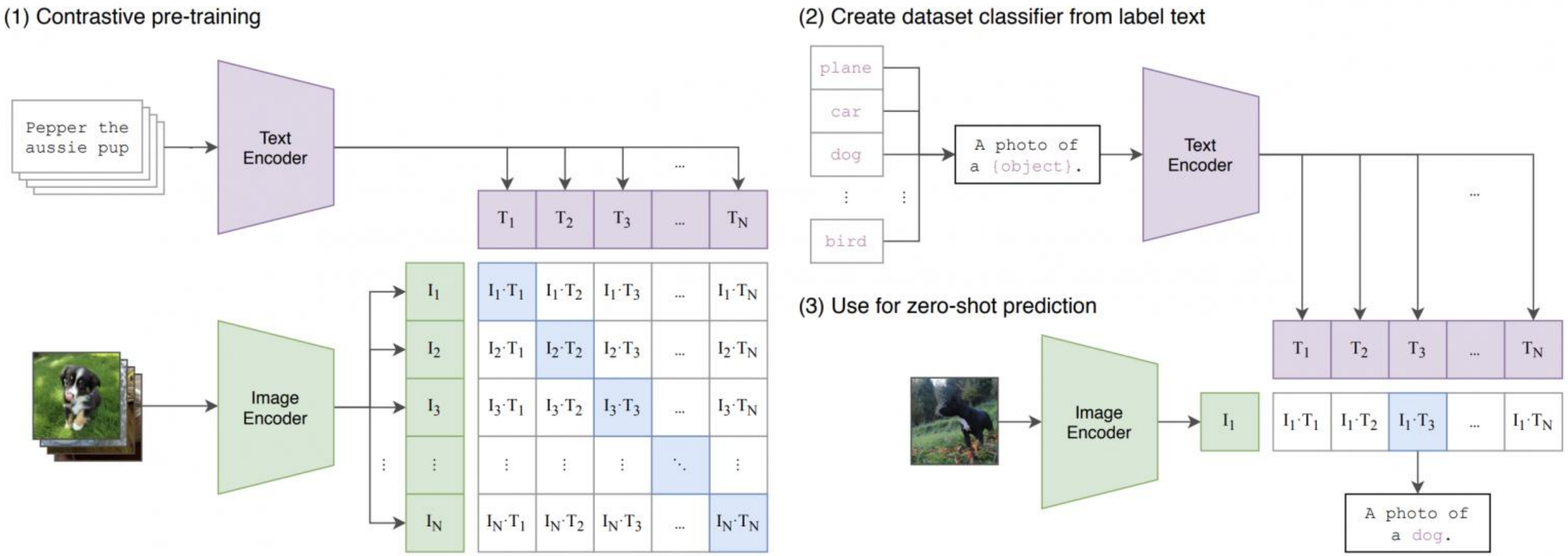
Generative-Based Approach



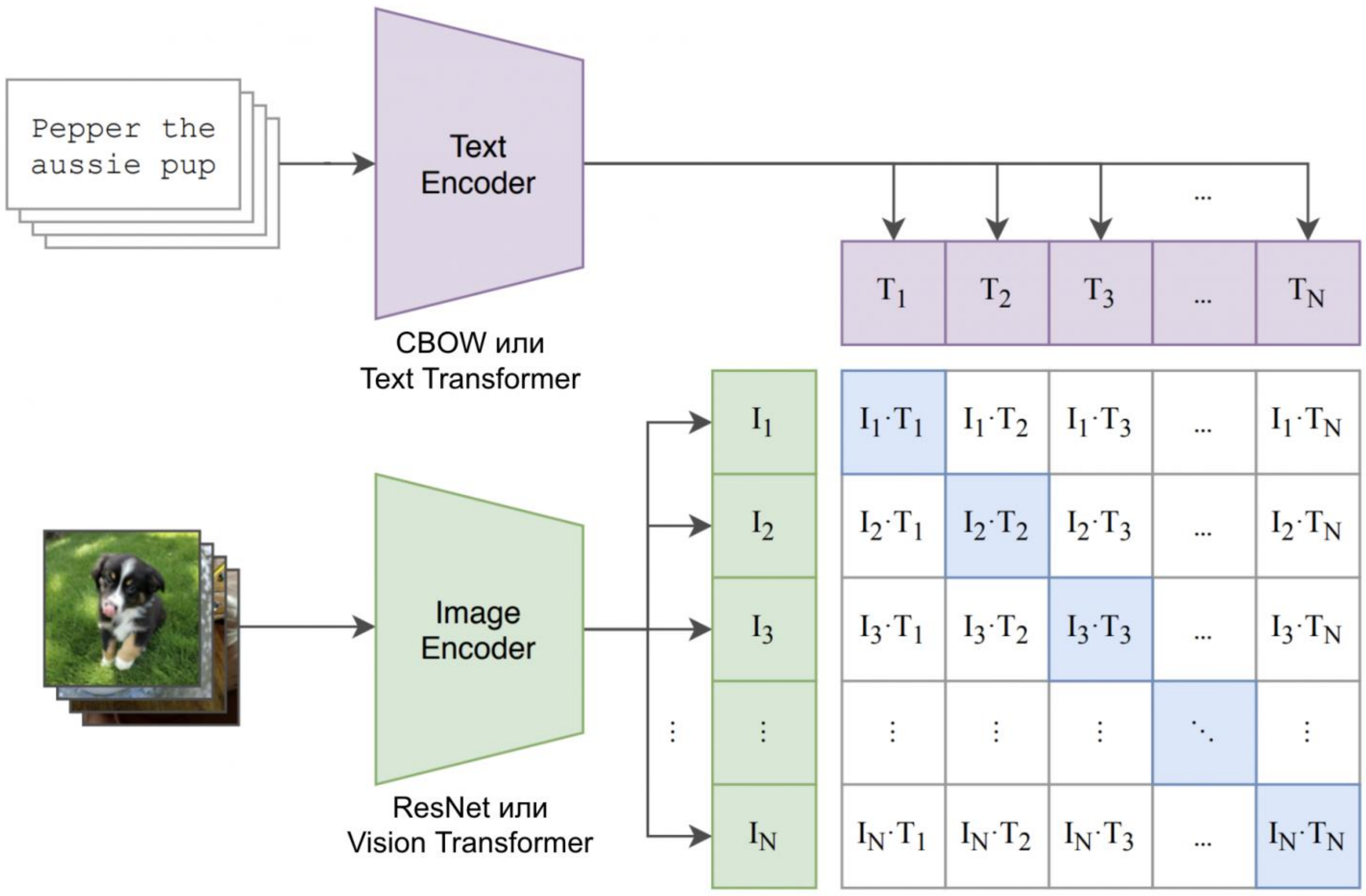
N-Shot Learning Applications

- Medical Image Analysis
- Visual-Question Answering (VQA)
- Autonomous Driving
- Image Retrieval and Action Recognition
- Text Classification
- Face Recognition

CLIP

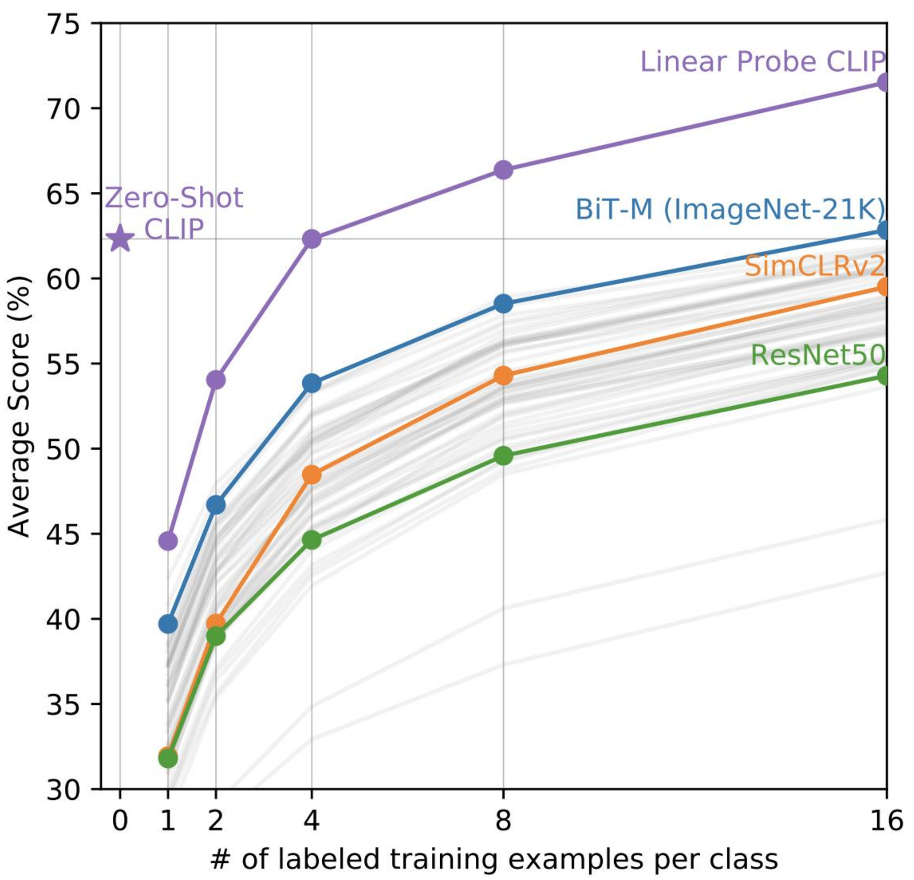

(1) Contrastive pre-training
Pepper the aussie pup
Text Encoder
Image Encoder
T1 T2 T3 ... TN
I1 I2 I3 ⋮ IN
I1·T1 I1·T2 I1·T3 ... I1·TN
I2·T1 I2·T2 I2·T3 ... I2·TN
I3·T1 I3·T2 I3·T3 ... I3·TN
IN·T1 IN·T2 IN·T3 ... IN·TN
(2) Create dataset classifier from label text
plane
car
dog
bird
A photo of a {object}.
Text Encoder
T1 T2 T3 ... TN
(3) Use for zero-shot prediction
Image Encoder
I1
I1·T1 I1·T2 I1·T3 ... I1·TN
A photo of a dog.

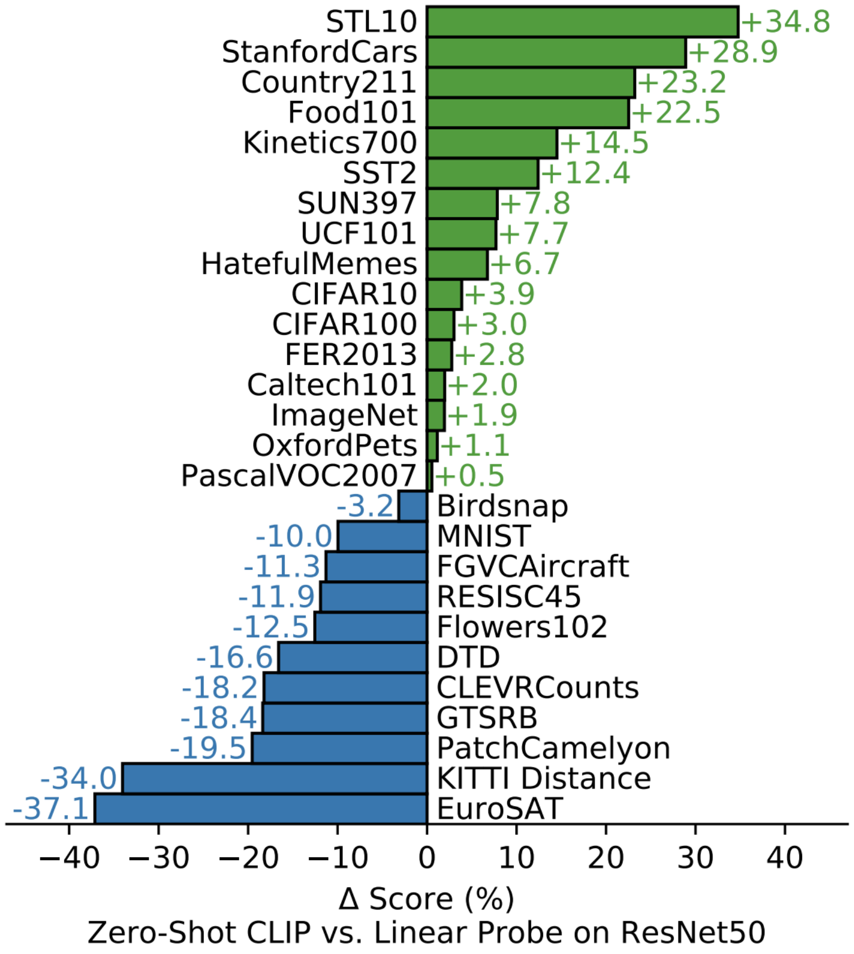
CLIP



Pepper the aussie pup
Text Encoder
CBOW или
Text Transformer
T1
T2
T3
...
TN
Image Encoder
ResNet или
Vision Transformer
I1
I2
I3
IN
I1·T1
I1·T2
I1·T3
I1·TN
I2·T1
I2·T2
I2·T3
I2·TN
I3·T1
I3·T2
I3·T3
I3·TN
IN·T1
IN·T2
IN·T3
IN·TN

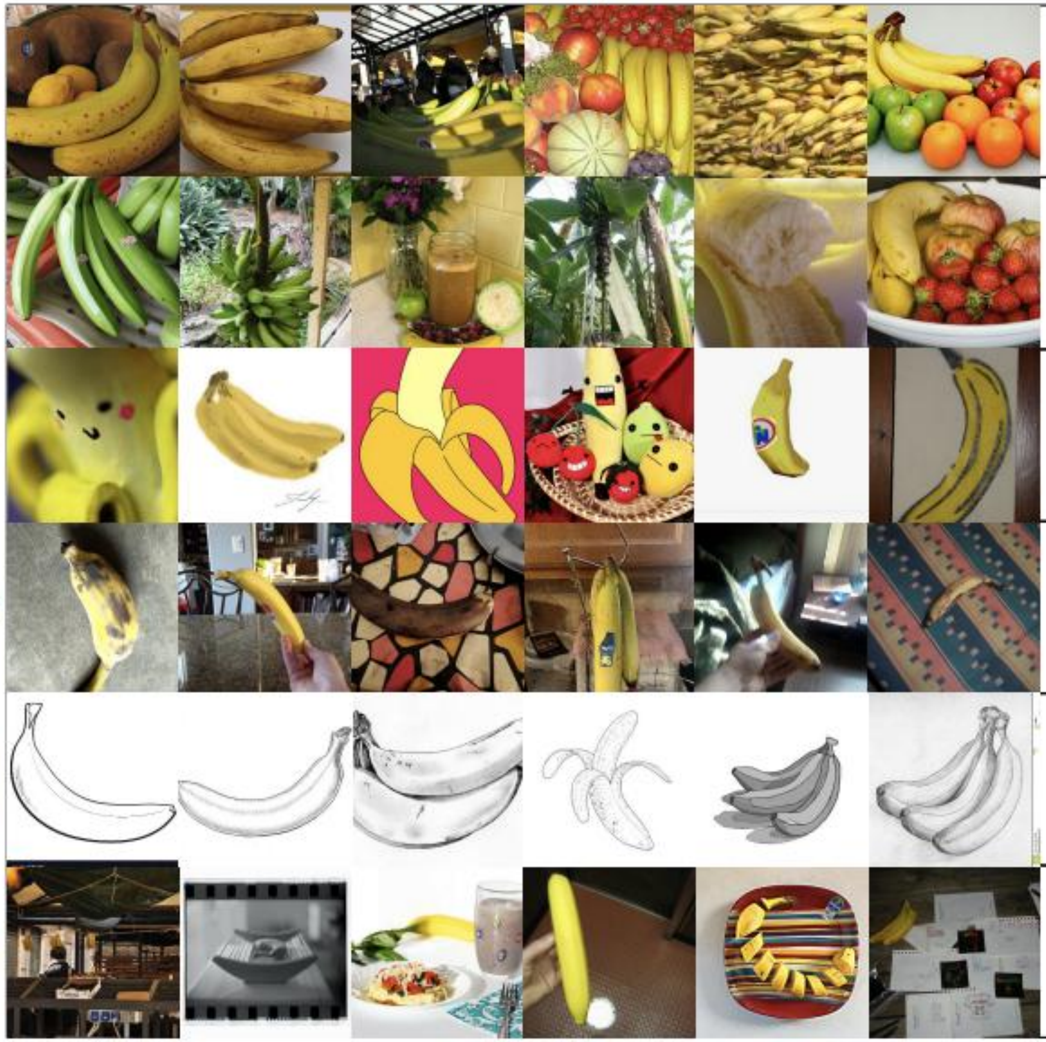
CLIP


Linear Probe CLIP
Zero-Shot
CLIP
BiT-M (ImageNet-21K)
SimCLRv2
ResNet50
Average Score (%)
# of labeled training examples per class
75
70
65
60
55
50
45
40
35
30
0
1
2
4
8
16


STL10 +34.8
StanfordCars +28.9
Country211 +23.2
Food101 +22.5
Kinetics700 +14.5
SST2 +12.4
SUN397 +7.8
UCF101 +7.7
HatefulMemes +6.7
CIFAR10 +3.9
CIFAR100 +3.0
FER2013 +2.8
Caltech101 +2.0
ImageNet +1.9
OxfordPets +1.1
PascalVOC2007 +0.5
-3.2 Birdsnap
-10.0 MNIST
-11.3 FGVCAircraft
-11.9 RESISC45
-12.5 Flowers102
-16.6 DTD
-18.2 CLEVRCounts
-18.4 GTSRB
-19.5 PatchCamelyon
-34.0 KITTI Distance
-37.1 EuroSAT
−40
−30
−20
−10
0
10
20
30
40
Δ Score (%)
Zero-Shot CLIP vs. Linear Probe on ResNet50

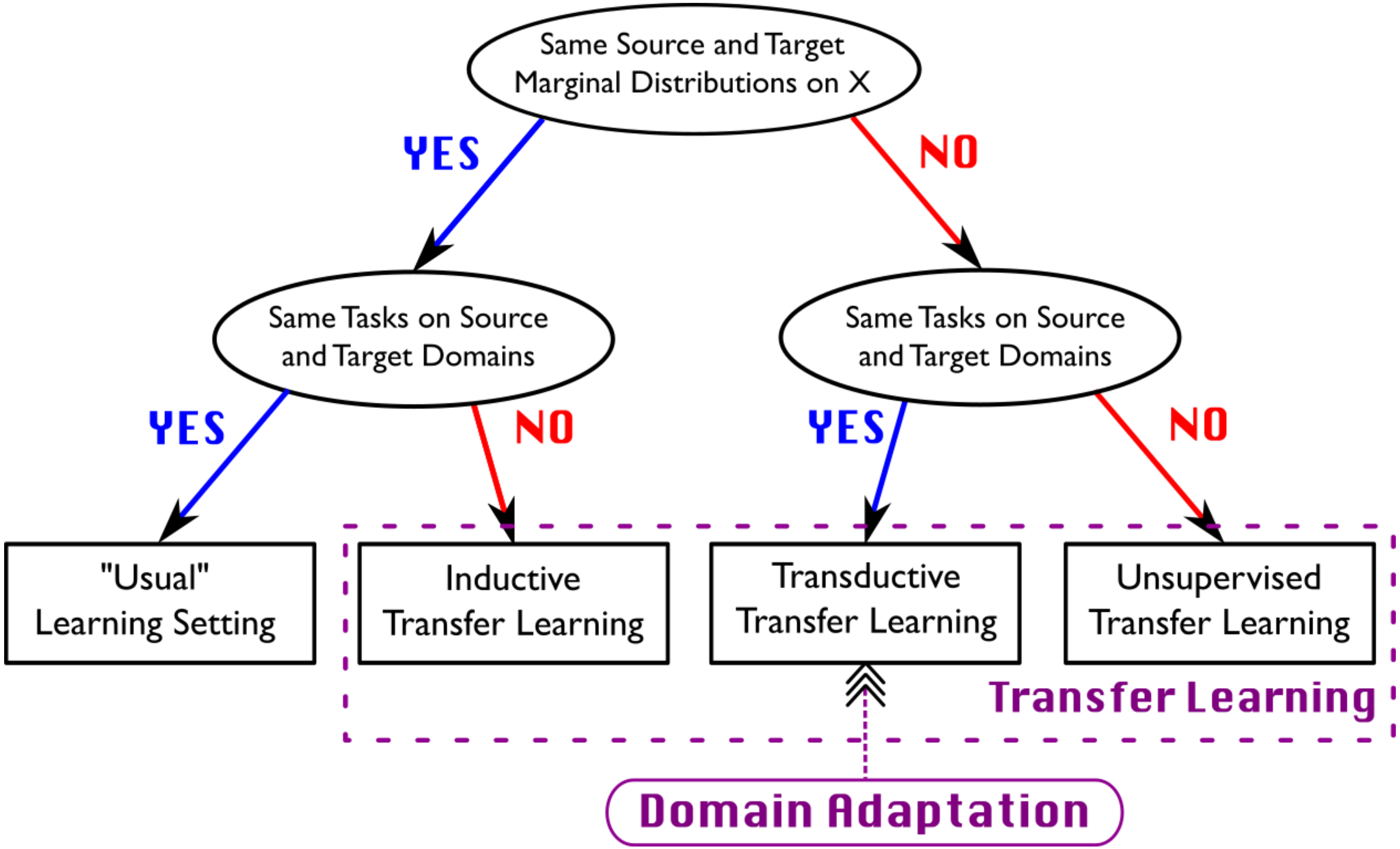
CLIP

Dataset Examples

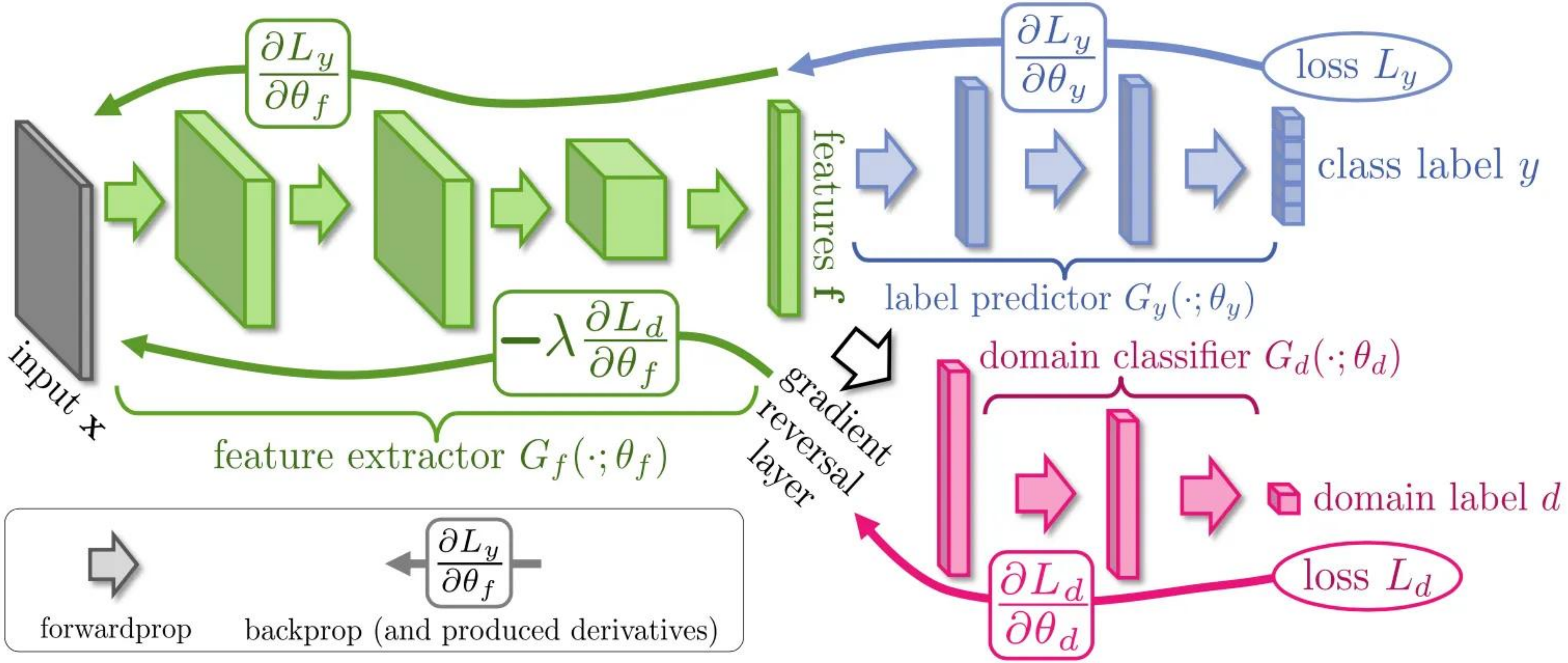


	ImageNet ResNet101	Zero-Shot CLIP	Δ Score
ImageNet	**76.2**	**76.2**	0%
ImageNetV2	64.3	**70.1**	+5.8%
ImageNet-R	37.7	**88.9**	+51.2%
ObjectNet	32.6	**72.3**	+39.7%
ImageNet Sketch	25.2	**60.2**	+35.0%
ImageNet-A	2.7	**77.1**	+74.4%

Domain Adaptation



Unsupervised Domain Adaptation



Ganin, Y., & Lempitsky, V. (2014)

Unsupervised Domain Adaptation by Backpropagation

https://arxiv.org/pdf/1409.7495
https://jmlr.org/papers/volume17/15-239/15-239.pdf

Полезные материалы

Few-Shot Learning

https://www.youtube.com/watch?v=Xuat7kHYwno&list=PL1pUDpkFOnIzeLCZ5aZgSXVZ8BcpCYN8Y

https://www.youtube.com/watch?v=ppC9ruaVuQQ

https://www.ibm.com/topics/few-shot-learning

https://github.com/sicara/easy-few-shot-learning

CLIP

https://github.com/openai/CLIP

https://habr.com/ru/articles/539312/

Unsupervised Domain Adaptation

https://github.com/adapt-python/adapt

https://arxiv.org/abs/2409.15264v1

https://github.com/ViLab-UCSD/UDABench_ECCV2024

https://www.youtube.com/watch?v=5SsEZvIYqqM&list=PLOQ9wdSxLW097UBdObI2vdereGJgA74Nb&index=25