A brief Introduction to few-shot learning

深度学习?
小样本学习!

Support Set

犰狳

[qiú yú]

穿山甲

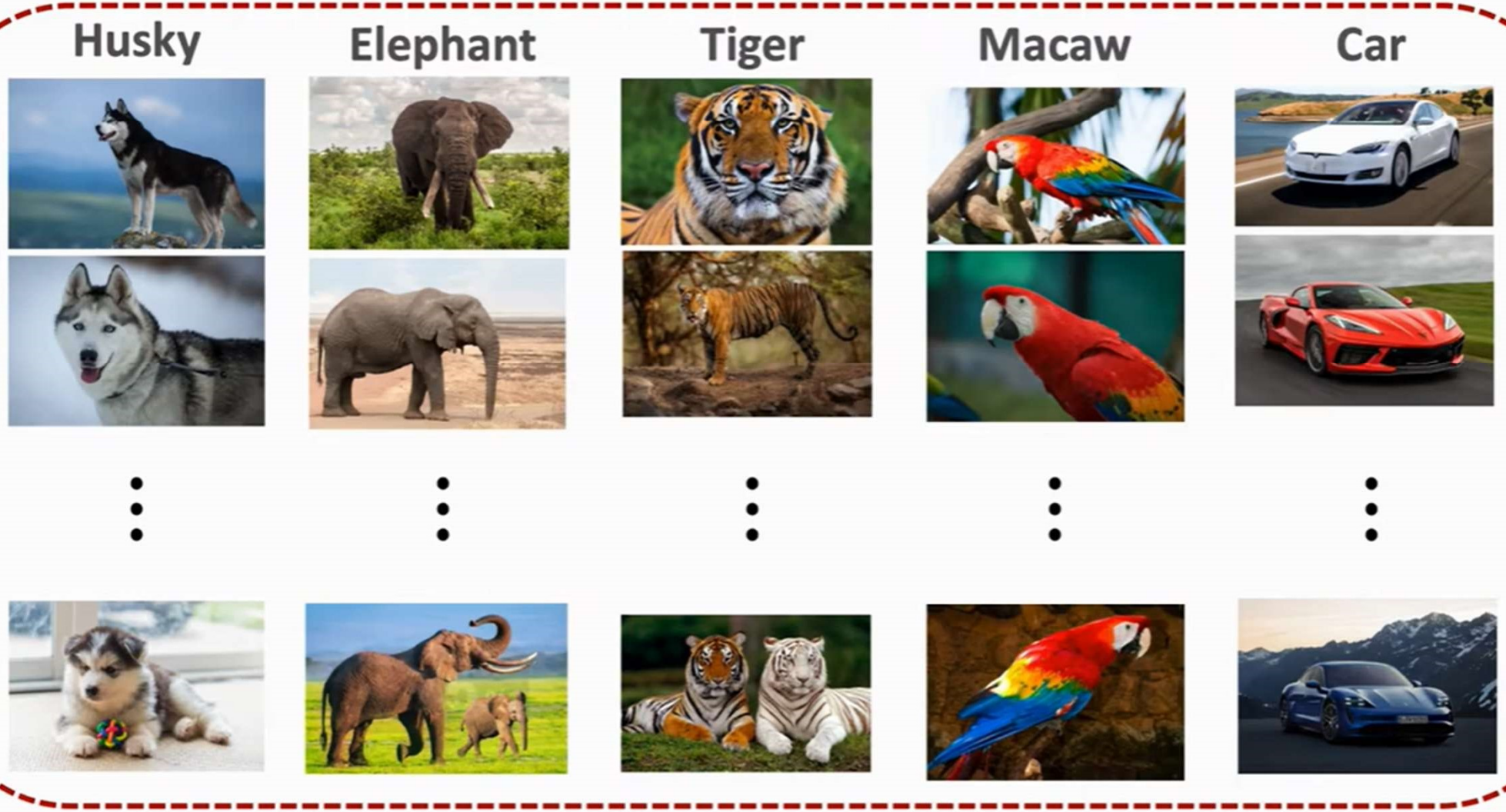
Armadillo

Pangolin

Query

Armadillo or Pangolin?

Training Set



深度学习：识别新的样本
小样本学习：识别异同similarity

Are they the same kind of animal?

Are they the same kind of animal?

Few-Shot Learning

Query:

Support Set:

Fox	Squirrel	Rabbit	Hamster	Otter	Beaver
					

Few-Shot Learning

Query:

预训练+finetune

Support Set:

Fox

Squirrel

Rabbit

Hamster

Otter

Beaver

Meta learning（元学习）

- 训练模型自主学习的能力
  （死记硬背式的学习和方法论式的学习）

- 小学1年级学calculus，大学一年级学calculus的区别

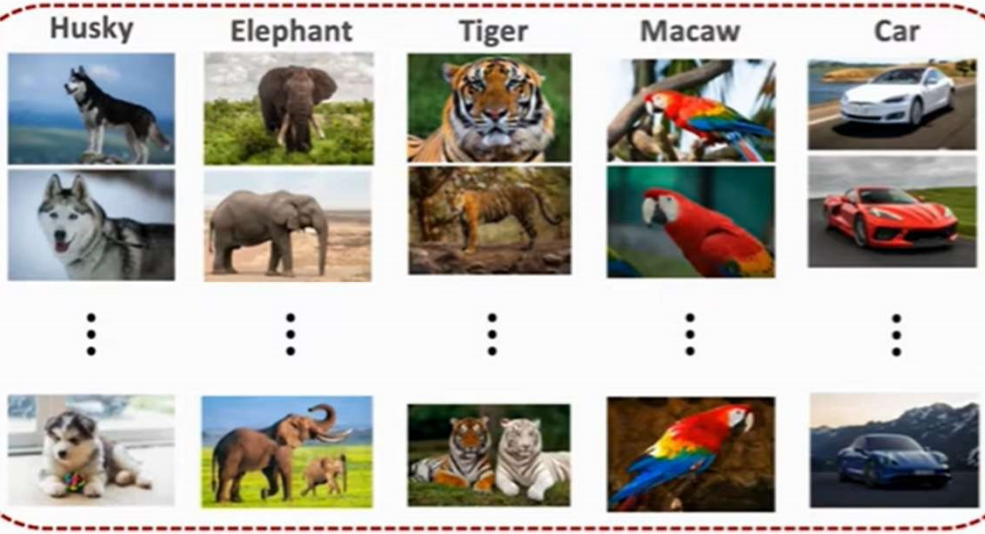
Meta Learning

Support set:

Training Set, Support Set, and Query

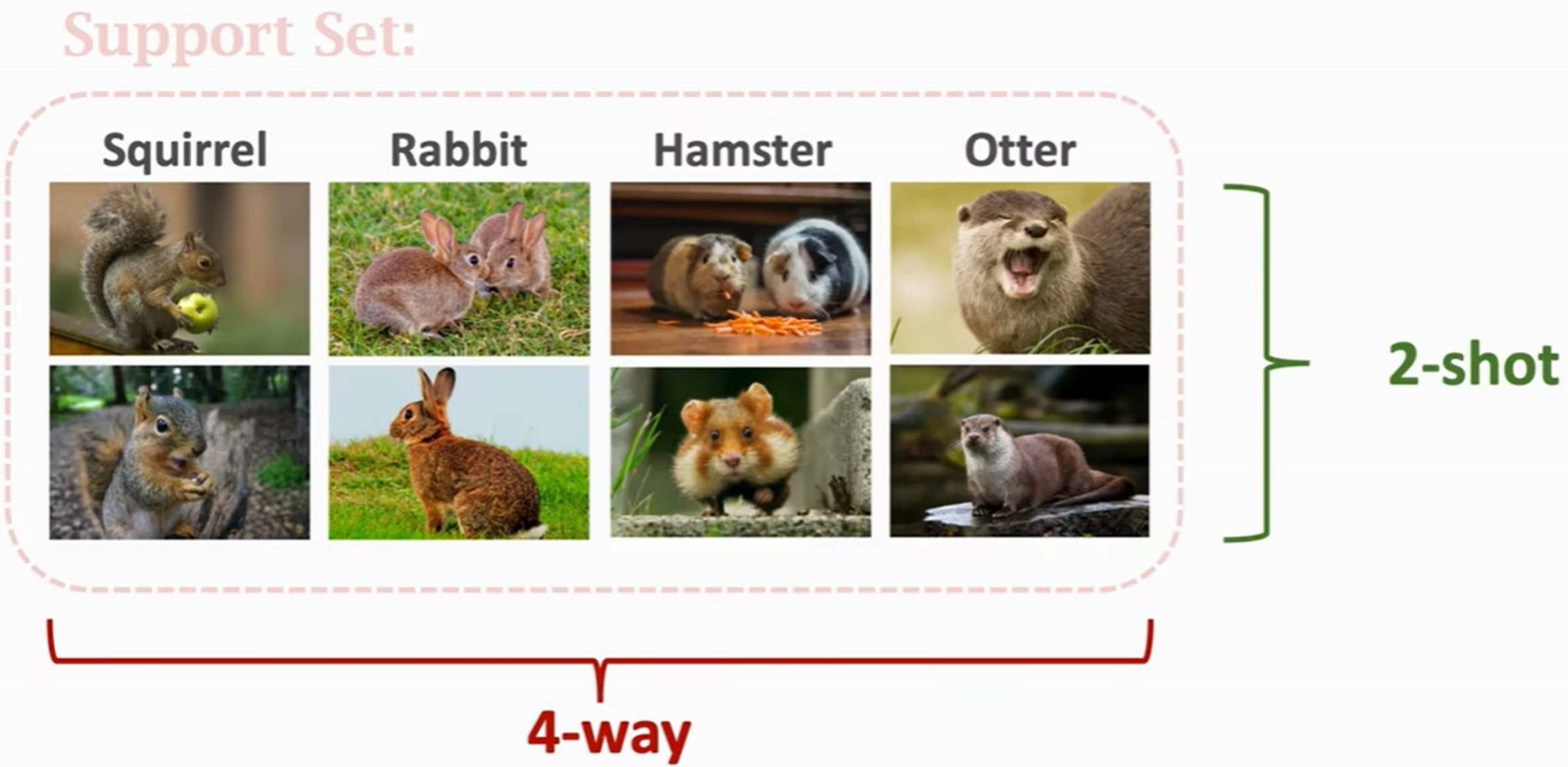
Support Set:

Training Set



Query Sample

k-way *n*-shot Support Set



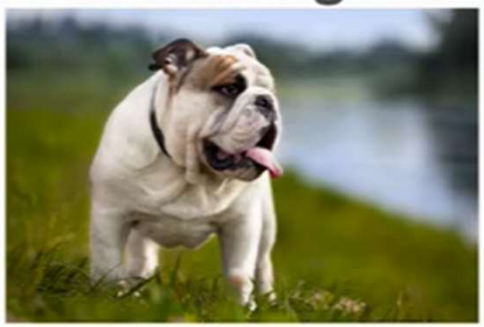
Basic Idea

- Learn a similarity function: $\mathrm{sim}(\mathbf{x}, \mathbf{x}')$.
- Ideally, $\mathrm{sim}(\mathbf{x}_1, \mathbf{x}_2) = 1$, $\mathrm{sim}(\mathbf{x}_1, \mathbf{x}_3) = 0$, and $\mathrm{sim}(\mathbf{x}_2, \mathbf{x}_3) = 0$.

Bulldog

$\mathbf{x}_1$

Bulldog



$\mathbf{x}_2$

Fox

$\mathbf{x}_3$

Basic Idea

- First, learn a similarity function from large-scale training dataset.

深度学习：样本是什么
预训练模型：相似度是什么

Basic Idea

What is in the image?

Query:

sim = 0.2	sim = 0.1	sim = 0.03	sim = 0.05	sim = 0.7	sim = 0.5
Greyhound	Bulldog	Armadillo	Pangolin	Otter	Beaver
					

Positive Samples

( ,  , 1)

( ,  , 1)

( ,  , 1)

Negative Samples

( ,  , 0)

( ,  , 0)

( ,  , 0)

Training Siamese Network

$\mathbf{x}_1$

$\mathbf{x}_2$

f

f

$\mathbf{h}_1$

$\mathbf{h}_2$

$\mathbf{z}$

Dense Layers

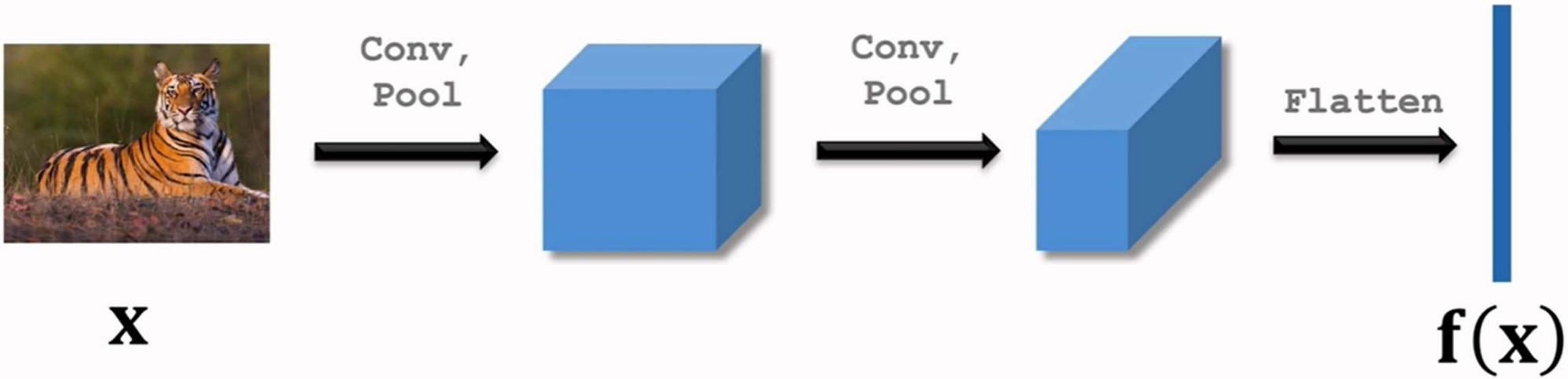
Sigmoid

$\text{sim}(\mathbf{x}_1, \mathbf{x}_2)$

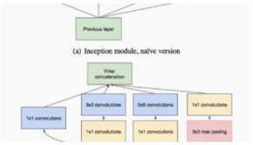
卷积神经网络
更好的特征提取器?
魔改网络结构!

Pretraining

- Pretrain a CNN for feature extraction (aka embedding).
- The CNN can be pretrained using standard supervised learning or Siamese network.



CNN常见CV架构对比



MAT研究所：以下是在ILSVRC竞赛中赢得的一些流行的CNN架构。LeNet-5AlexNetVGGNetGoogLeNetResNet1. DCNN架构通常，大多数深层卷积神经网络由一组关键的基本层组成，包括卷积... 阅读全文

▲ 赞同 1 ▼ 添加评论 2021-06-28

聊一聊CV中的backbone结构：

凤舞九天：这里简单介绍下resnet系列和inception 系列，因为目前resnet系列（包括SE、resnest等网络结构）仍然是业... 阅读全文

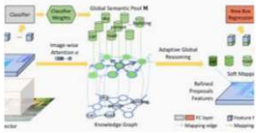
▲ 赞同 18 ▼ 添加评论 2021-06-26

发现CV方向的很多问题已经被研究的很透彻了，不知道怎样才能找到问题作出改进发论文?

匿名用户：问题多了去了，过参数化网络的局部极小空间的几何结构弄清楚了么？ 网络结构和功能的对应清楚么？ 当网络结构和参数需要同时优化的时候... 阅读全文

▲ 赞同 8 ▼ 添加评论 2019-05-24

图神经网络和cvnlp的结合。请问现有的论文是否大多数是节点分类?



Spaceman：并不局限于节点的分类，图模型有利于进行结构化建模，比如在CV中可以把结构信息嵌入到模型中。例如CVPR19的Reasoning RCNN。对于类别数极多的目标检测任务，物体的类别大致是一个长尾分布，而尾... 阅读全文

▲ 赞同 2 ▼ 1 条评论 2020-03-31

[CV - Object Detection - 2022]目标检测系列 - 网络结构设计和优化技巧

搜索发现

博人传漫画第 79 话
美联储紧急贷款升至新纪录
半天妖垃圾桶捞回餐食又端给顾客
美联储紧急贷款飙升至新纪录
日媒：日本东京湾出现石油泄漏
国内成品油迎来年内第二跌
央行降低存款准备金率 0.25 个百分点
姜广涛涉嫌刑事犯罪

Training Siamese Network

x_1

x_2

f

f

h_1

h_2

Dense Layers

z

Target = 1

Loss

Sigmoid

$\text{sim}(x_1, x_2)$

Training Siamese Network

x_1

x_2

f

f

h_1

h_2

z

Dense Layers

Target = 1

Loss

Sigmoid

$\text{sim}(x_1, x_2)$

Triplet Loss

$\mathbf{x}^{+}$
(positive)

$\mathbf{f}$

$\mathbf{f}(\mathbf{x}^{+})$

$\mathbf{x}^{a}$
(anchor)

$\mathbf{f}$

$\mathbf{f}(\mathbf{x}^{a})$

$\mathbf{x}^{-}$
(negative)

$\mathbf{f}$

$\mathbf{f}(\mathbf{x}^{-})$

$$d^{+} = \left|\left|\mathbf{f}(\mathbf{x}^{+}) - \mathbf{f}(\mathbf{x}^{a})\right|\right|_2^2$$

$$d^{-} = \left|\left|\mathbf{f}(\mathbf{x}^{a}) - \mathbf{f}(\mathbf{x}^{-})\right|\right|_2^2$$

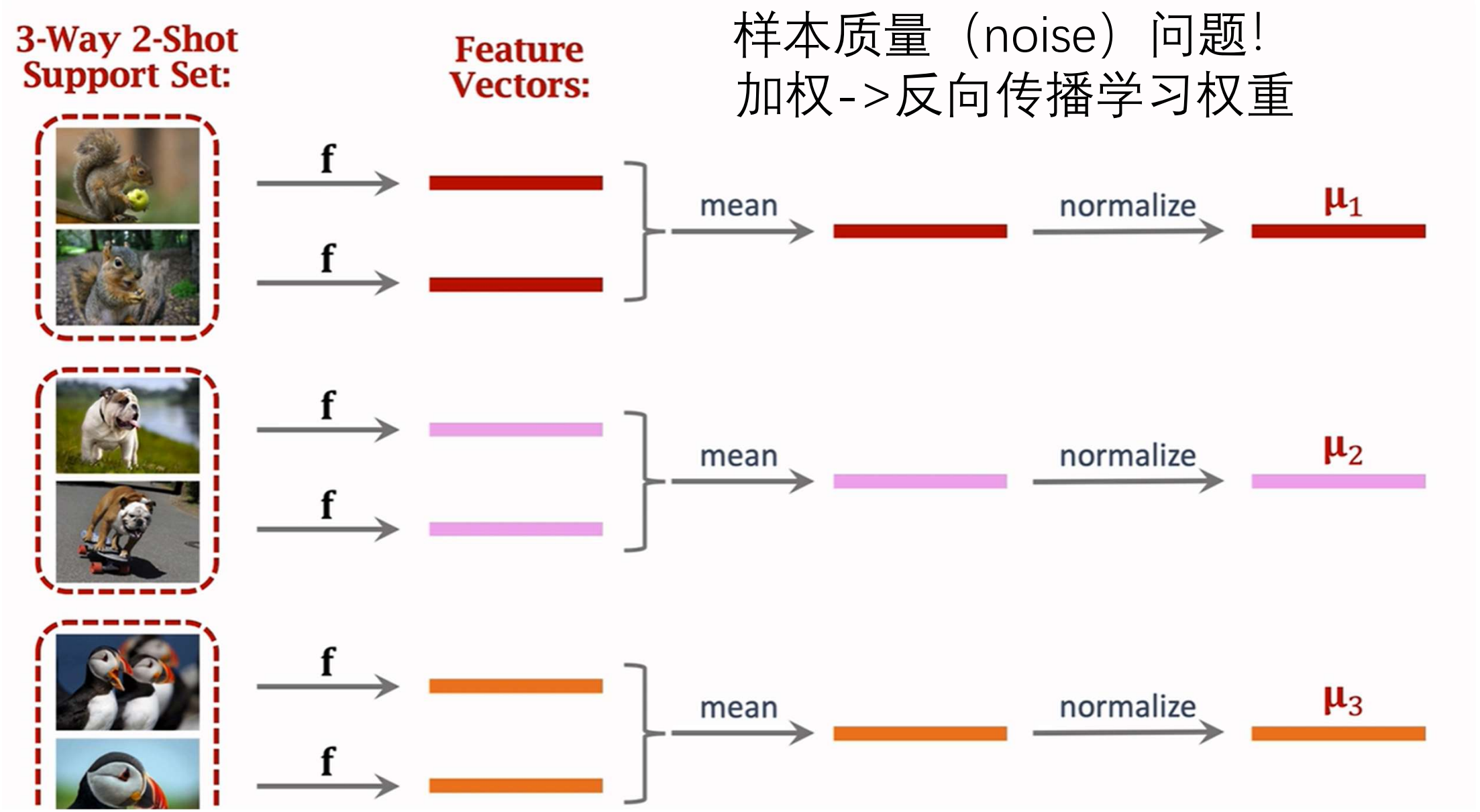
Triplet Loss

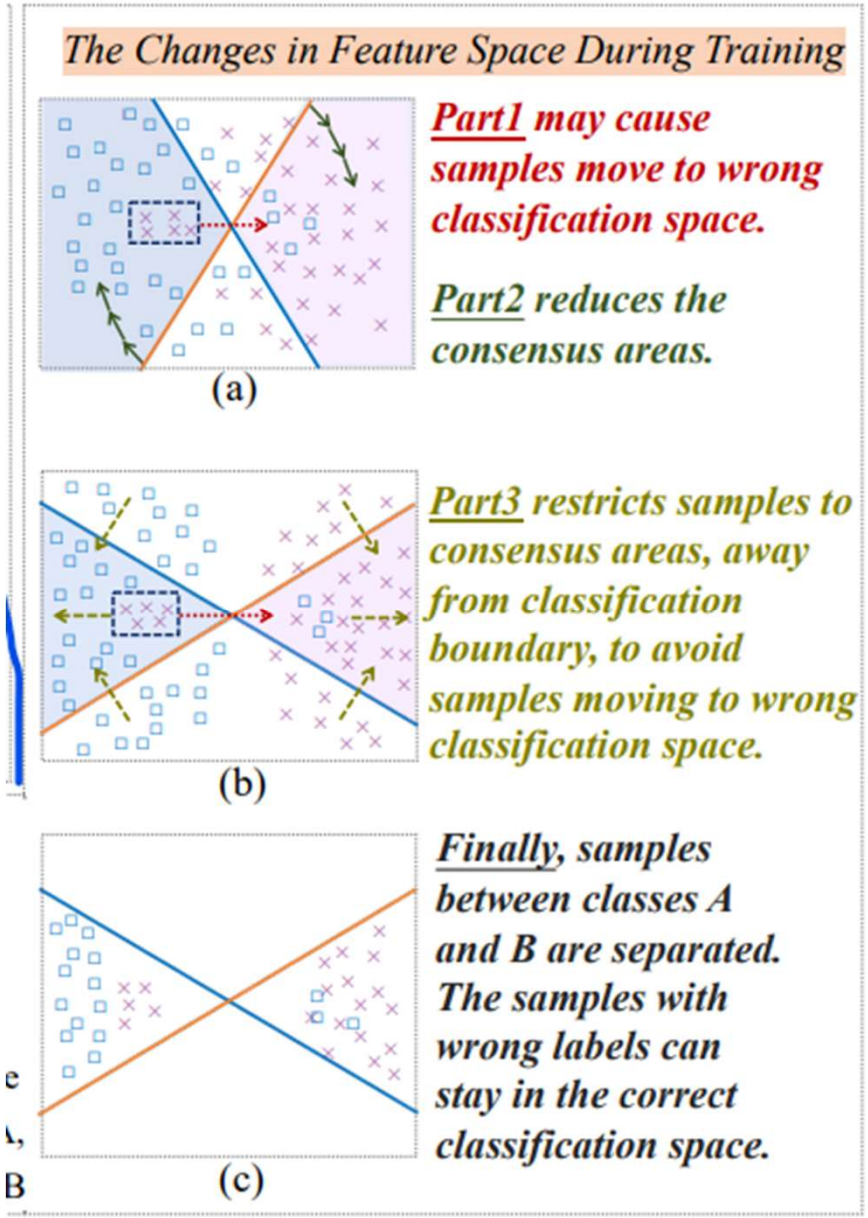
$\mathbf{x}^+$
(positive)

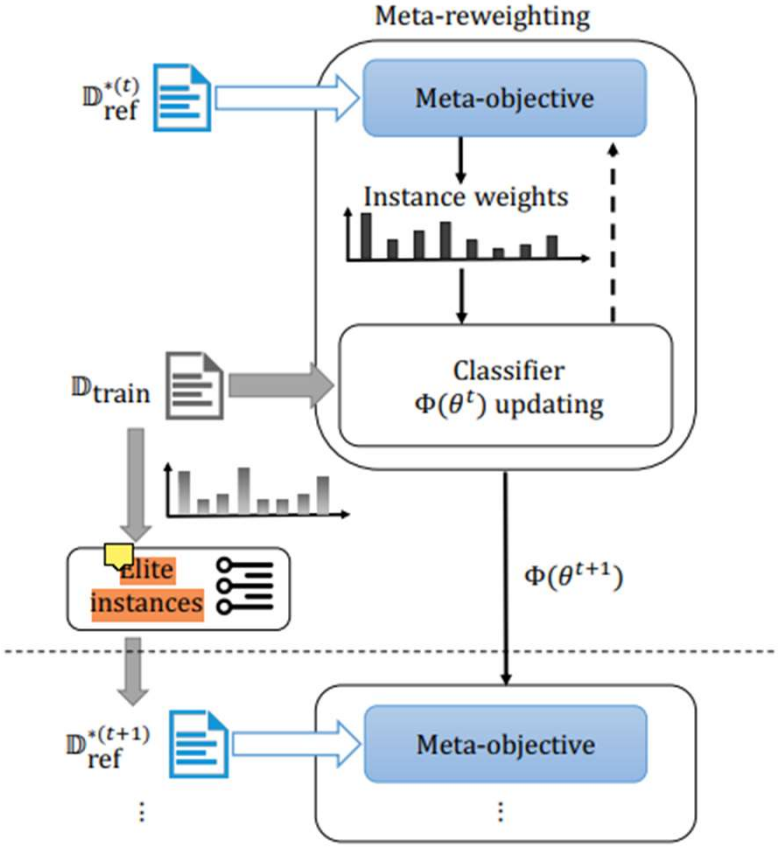
$\mathbf{x}^a$
(anchor)

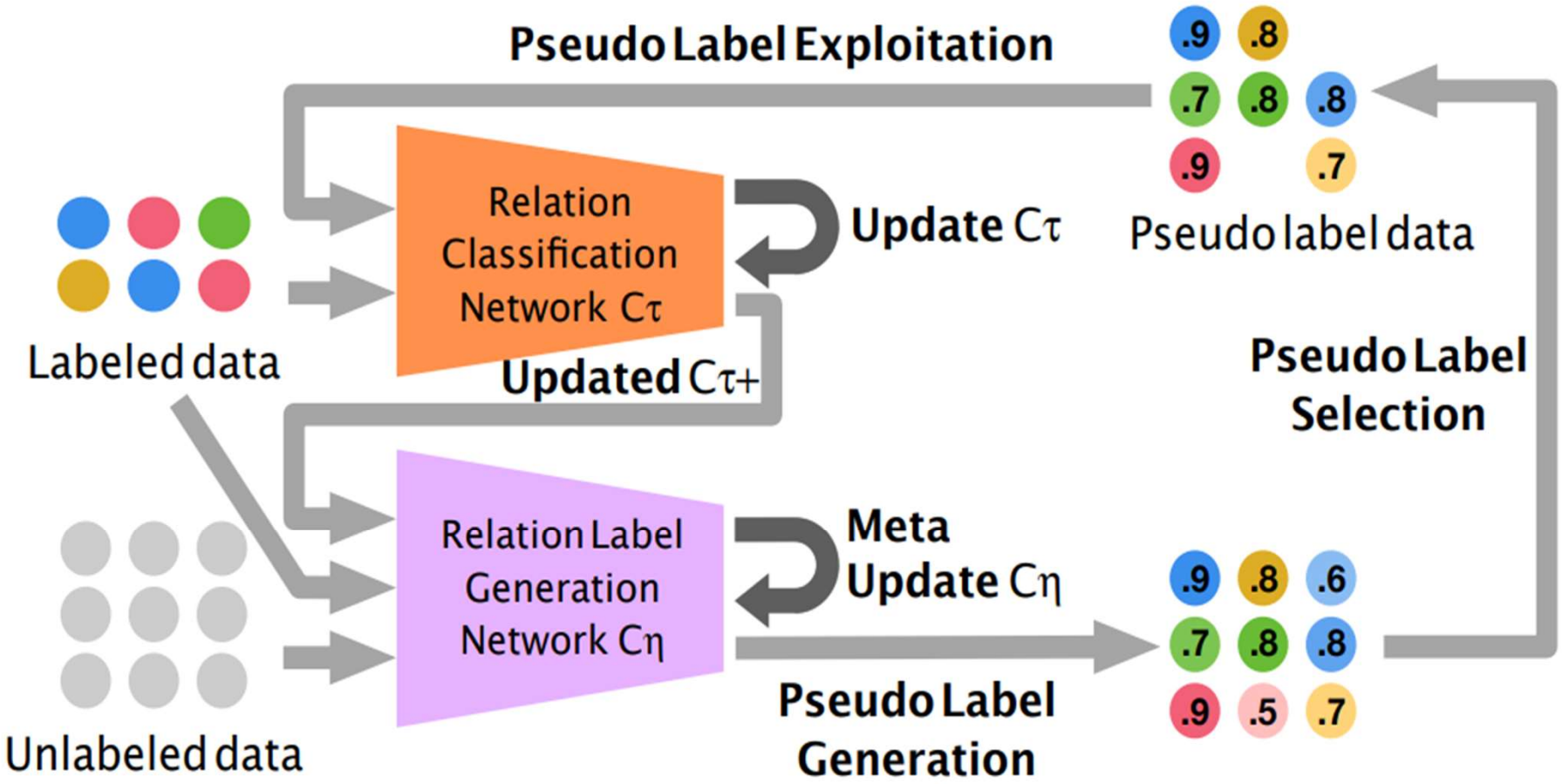
$\mathbf{x}^-$
(negative)

- Encourage $d^+ = \left|\left|\mathbf{f}(\mathbf{x}^+) - \mathbf{f}(\mathbf{x}^a)\right|\right|_2^2$ to be small.
- Encourage $d^- = \left|\left|\mathbf{f}(\mathbf{x}^a) - \mathbf{f}(\mathbf{x}^-)\right|\right|_2^2$ to be big.
- If $d^- \geq d^+ + \alpha$, then no loss. ($\alpha > 0$ is margin.)
- Otherwise, the loss is $d^+ + \alpha - d^-$.
- $\text{Loss}(\mathbf{x}^a, \mathbf{x}^+, \mathbf{x}^-) = \max\{0,\ d^+ + \alpha - d^-\}$.

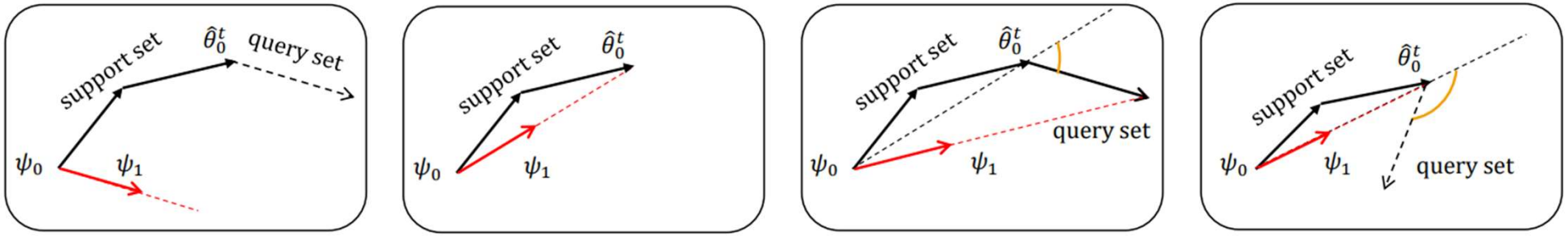

3-Way 2-Shot Support Set:
Feature Vectors:
样本质量（noise）问题!
加权->反向传播学习权重
f
f
mean
normalize
μ_1
f
f
mean
normalize
μ_2
f
f
mean
normalize
μ_3


The Changes in Feature Space During Training
Part1 may cause samples move to wrong classification space.
Part2 reduces the consensus areas.
(a)
Part3 restricts samples to consensus areas, away from classification boundary, to avoid samples moving to wrong classification space.
(b)
Finally, samples between classes A and B are separated. The samples with wrong labels can stay in the correct classification space.
(c)


Meta-reweighting
$\mathbb{D}_{\text{ref}}^{*(t)}$
Meta-objective
Instance weights
$\mathbb{D}_{\text{train}}$
Classifier
$\Phi(\theta^t)$ updating
elite instances
$\Phi(\theta^{t+1})$
$\mathbb{D}_{\text{ref}}^{*(t+1)}$
Meta-objective


Pseudo Label Exploitation
.9 .8
.7 .8 .8
.9 .7
Pseudo label data
Relation Classification Network Cτ
Update Cτ
Labeled data
Updated Cτ+
Pseudo Label Selection
Relation Label Generation Network Cη
Meta Update Cη
.9 .8 .6
.7 .8 .8
.9 .5 .7
Unlabeled data
Pseudo Label Generation

基于梯度的元学习方法



(a) MAML(First order) (b) REPTILE (c) AMGS (ours, scheme 1) (d) AMGS (ours, scheme 2)

当前的研究方向

- 正负样本的选择 →举例: 猫狗分类 猫狮分类？？？
- 更好的representation → 网络结构（各种net ……）encoder decoder 特征提取（contrastive learning）
- 过拟合问题：regularization技术、损失函数的设计
- 对训练数据噪声的鲁棒性
- 元学习的策略
- 基于强化学习的方法？？？ Future and discussion (GPT)