台灣人工智慧學校

友達光電 AU Optronics

# Learning with Small Data: Few-Shot Learning and Its Applications

Yu-Chiang Frank Wang 王鈺強, Professor

Dept. Electrical Engineering, National Taiwan University

2021/07/22

國立臺灣大學 National Taiwan University

# About Myself

- **Education**

- **Ph.D./M.S. in Electrical & Computer Engineering** 2002 – 2009
  Carnegie Mellon University, Pittsburgh, USA

- **B.S. in Electrical Engineering** 1997 – 2001
  National Taiwan University, Taipei, Taiwan

- **Experiences**

- **Professor** 2019 – present
  GICE & Dept. EE, National Taiwan University


- **AI Advisory Consultant** 2019 – present
  ASUS Intelligent Cloud Services (AICS)
- **Associate Professor**, National Taiwan Univ. 2017 – 2019

- **Deputy Director**, CITI, Academia Sinica 2015 – 2017
- **Associate/Assistant Research Fellow** 2009 – 2017
  Research Center for IT Innovation (CITI), Academia Sinica
- **Instructor** 2018 – present
  AI Academy, CHT Academy, III, ITRI

# About Myself (cont'd)

- **Industrial Collaboration**
  - **Collaborators** (current ones in blue):
    Google, Qualcomm, ASUS, Inventec, Chuanhwa Telecomm,
    E.SUN Commercial Bank, Tron Future Tech, Novatek, III,
    TSMC, ITRI, Viscovery, Digital Drift, BitMark, AsiaInfo
    AI Labs Taiwan, Umbo CV, KaikuTek, Theia, Deep01, Moxa

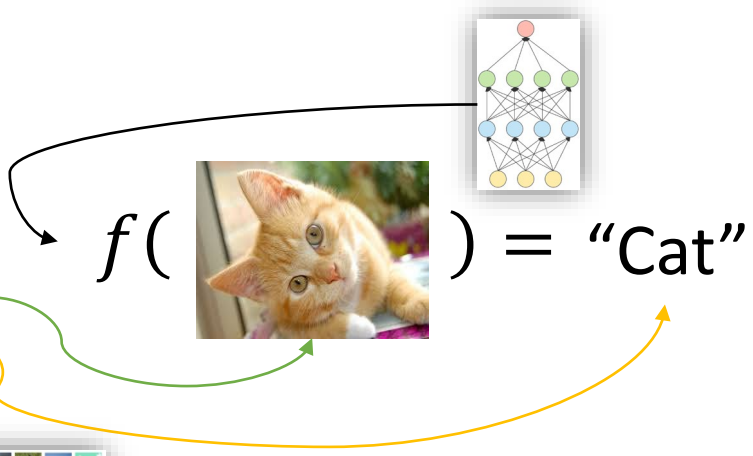
# What to Cover Today…
# Few-Shot Learning & Its Applications

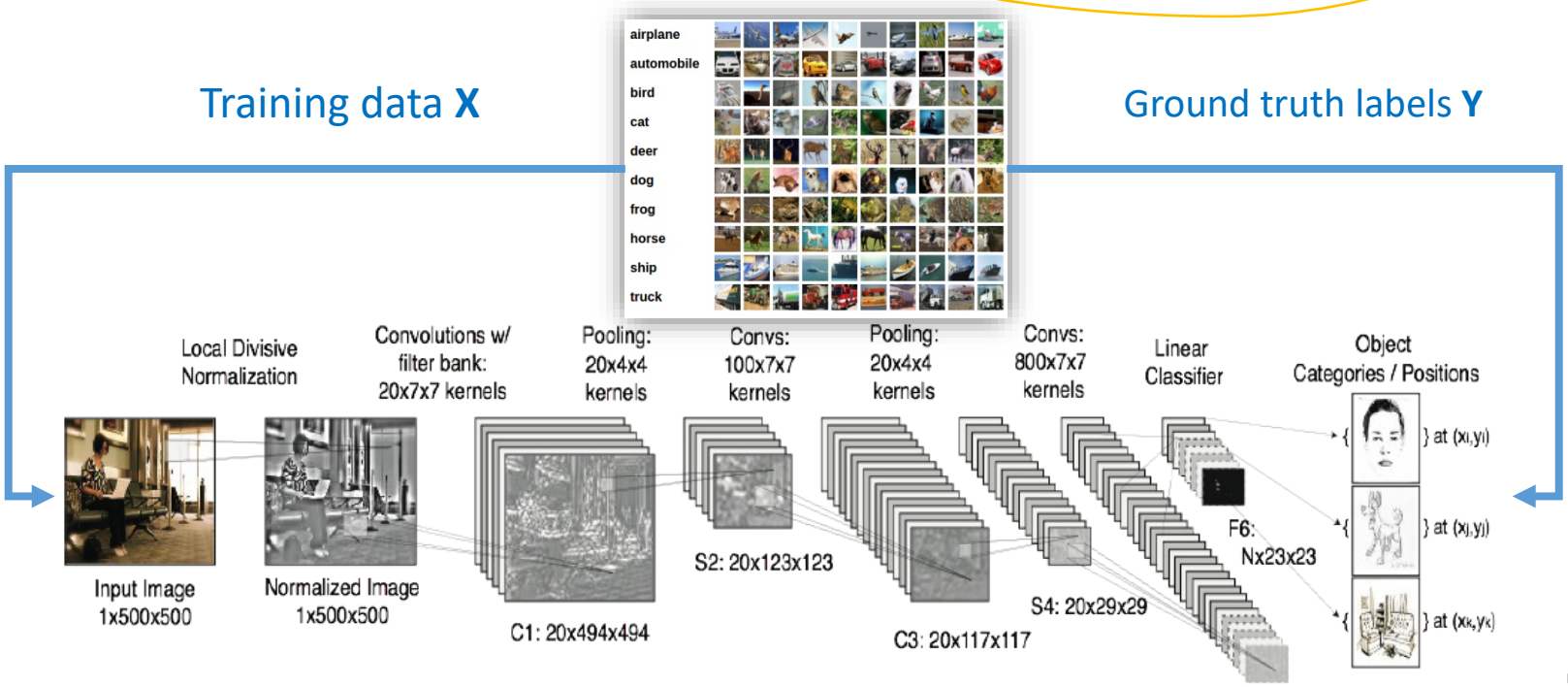
- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

Selected slide credits: C. Finn, S. Levine, & H.-Y. Lee

# Meta Learning 元學習

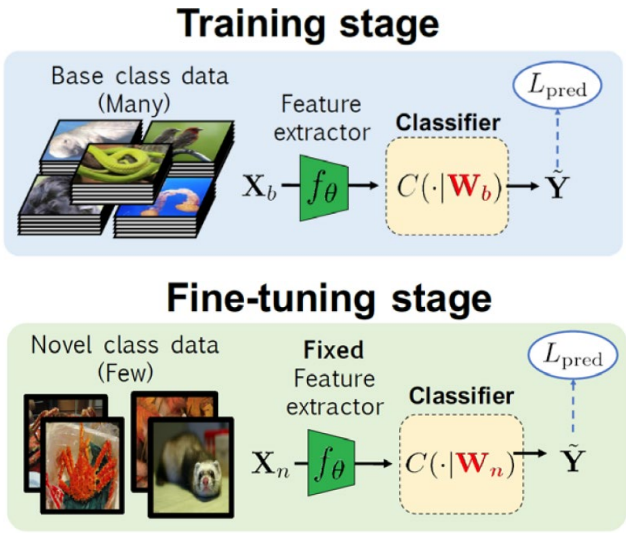
- Meta Learning ⊆ Machine Learning
- Machine Learning
  - Supervised Learning
    - Given training data **D = {X, Y}**, learn function/model ***f*** so that $f(x_i) = y_i$.

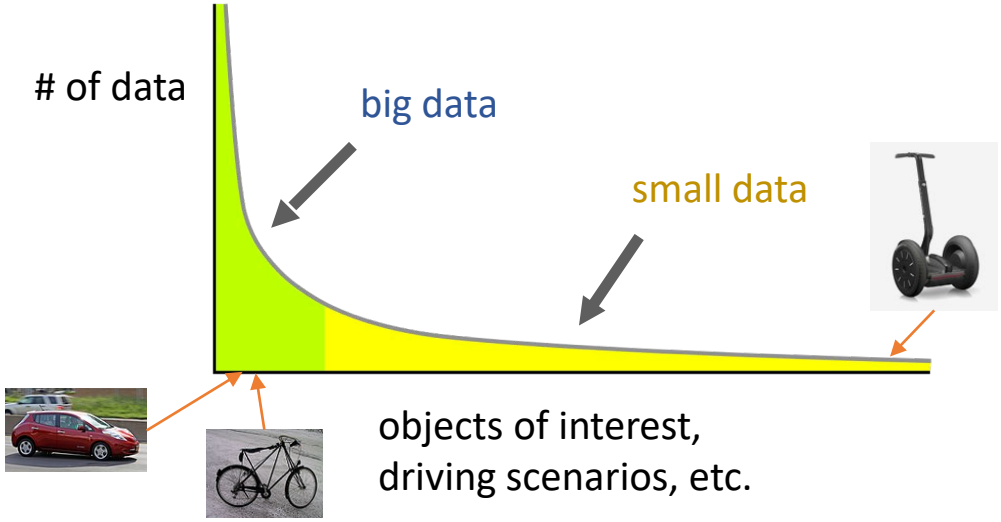


# What If Only Limited Amount of Data Available?

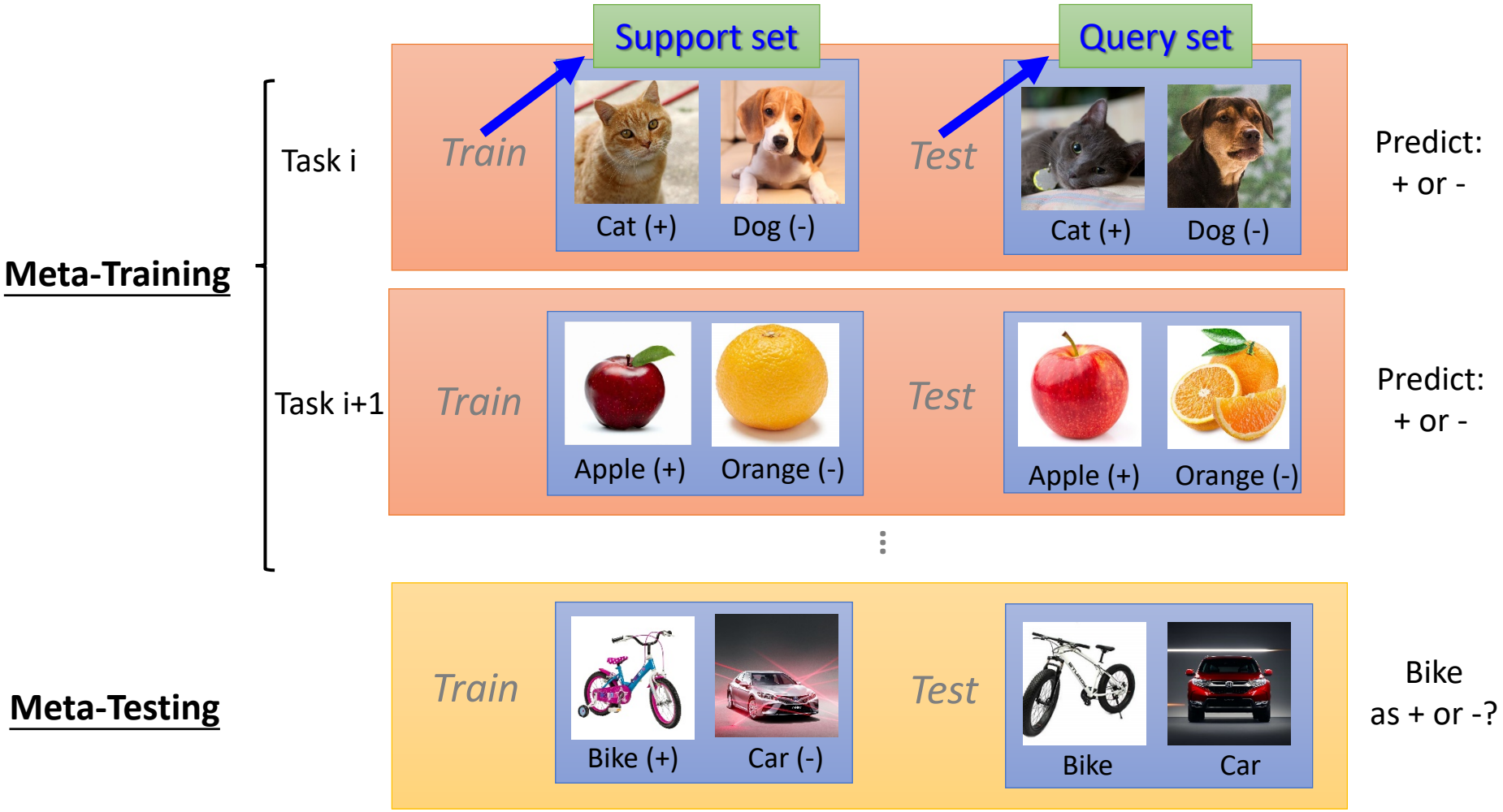
- **Naive transfer?**
  - **Model finetuning**:
    - Train a learning model (e.g., CNN) on large-size data (base classes), followed by finetuning on small-size data (novel classes).
    - That is, freeze feature backbone (learned from base classes) and learn/update classifier weights for novel classes.
  - Question: What would be the concern/limitation?



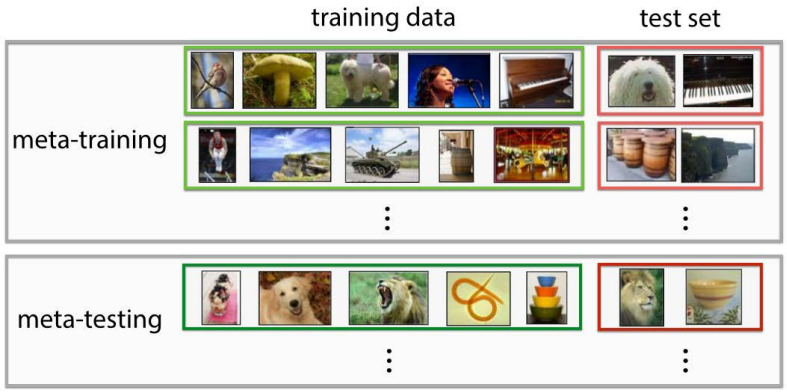
# Meta Learning = Learning to Learn

- Let's consider the following "2-way 1-shot" learning scheme:



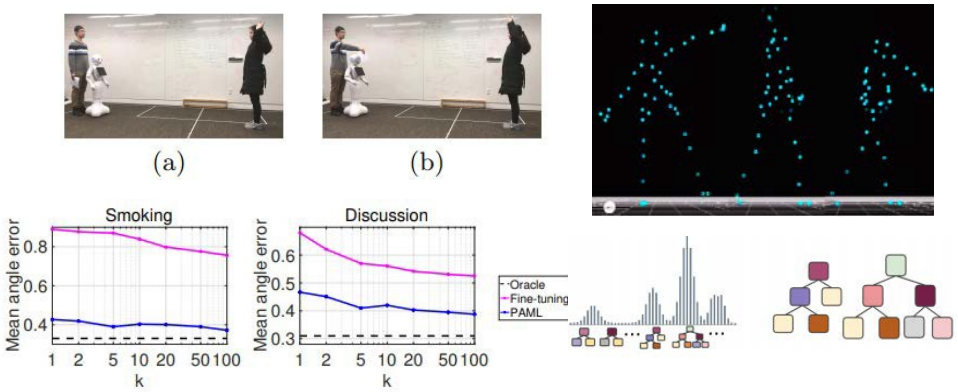
# Selected Applications of Meta Learning for Computer Vision

- Few-Shot Image Classification



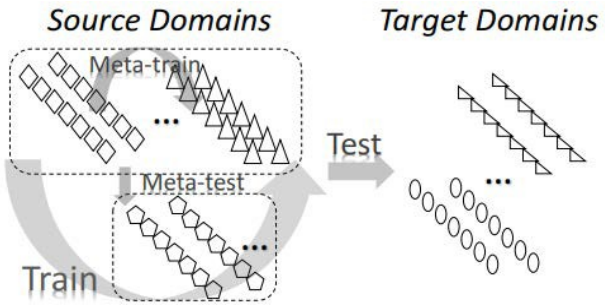

Vinyals et al., NIPS 2016

- Human Pose/Motion Prediction



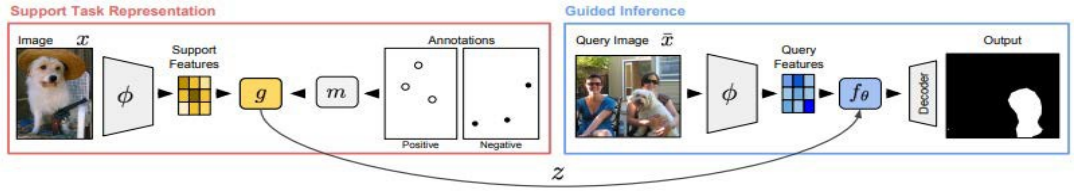

Gui et al., ECCV 2018

- Domain Transfer/Generalization




Li et al., AAAI 2018

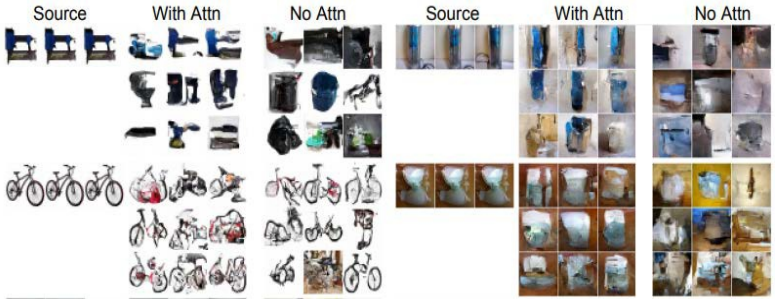
- Few-Shot Image Segmentation




Wang et al., ICCV 2019

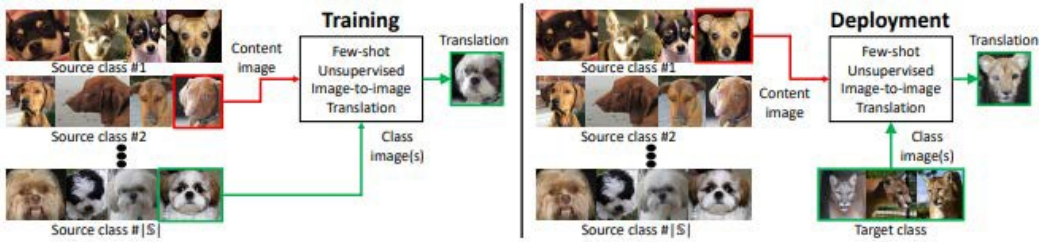
# Selected Applications of Meta Learning for Computer Vision (cont'd)

- Few-Shot Image Generation



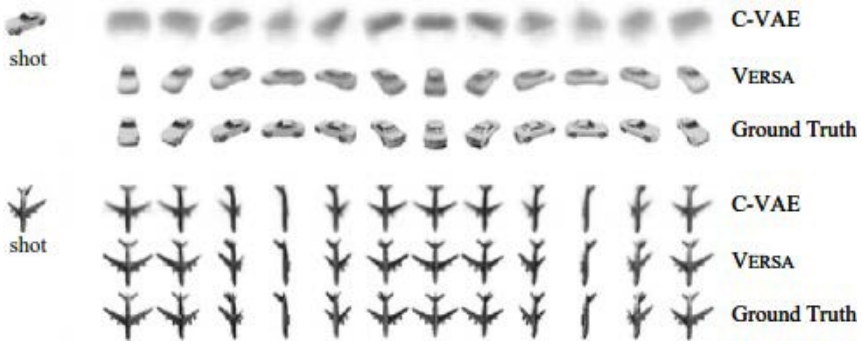

Reed et al., ICLR 2018

- Few-Shot Image-to-Image Translation



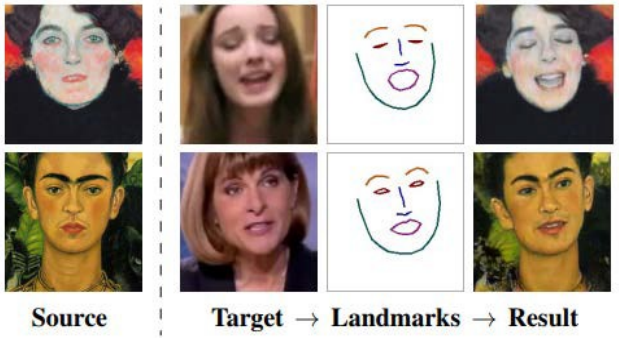

Liu et al., ICCV 2019

- Generation of Novel Viewpoints



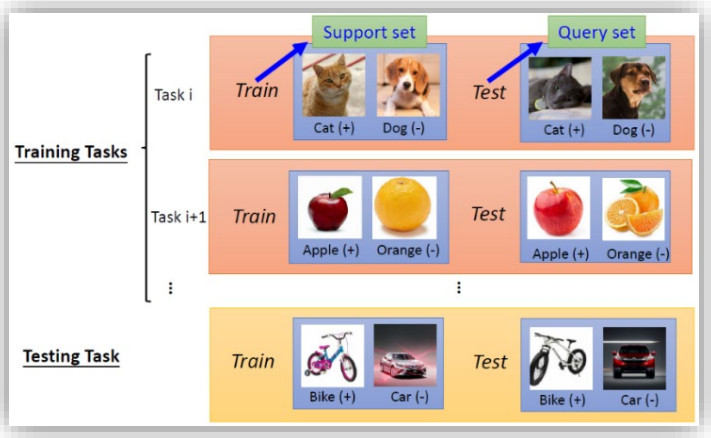

Gordon et al., NIPS Workshop 2018

- Generating Talking Heads from Images




Zakharov et al., ICCV 2019

# Meta Learning (cont'd)




- Two Perspectives…
  - *Probabilistic* View
    - Extract prior info from a set of (meta training) tasks, allowing efficient learning of a new task
    - Learning a new task uses this prior and (small) training set to infer most likely posterior model parameters
    - Easy to understand meta learning algorithms
    - E.g., MAML
  - ***Mechanistic* View**
    - A learning model (e.g., DNN) reads in a meta-dataset, which consists of many datasets, each for a different task
    - Then, the model observes new data points (for a novel task) and make predictions accordingly
    - Easy to implement meta learning algorithms
    - E.g., Relation Net, Prototypical Net, etc.

# A Quick Example

Person ID:
"Brad Pitt"

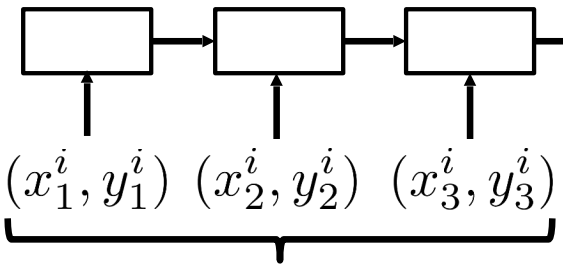
- Meta training: $\theta^\star = \arg\max_\theta \log p(\theta|\mathcal{D}_{\text{meta-train}})$
- Meta testing: $\phi^\star = \arg\max_\phi \log p(\phi|\mathcal{D}, \theta^\star)$

$\mathcal{D}_{\text{meta-train}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$

$\mathcal{D} = \{(x_1, y_1), \ldots, (x_k, y_k)\}$

$\mathcal{D}_i = \{(x_1^i, y_1^i), \ldots, (x_k^i, y_k^i)\}$

meta-training




[1, 0, 0]?
[0, 1, 0]?
[0, 0, 1]?

$\mathcal{D}_i$

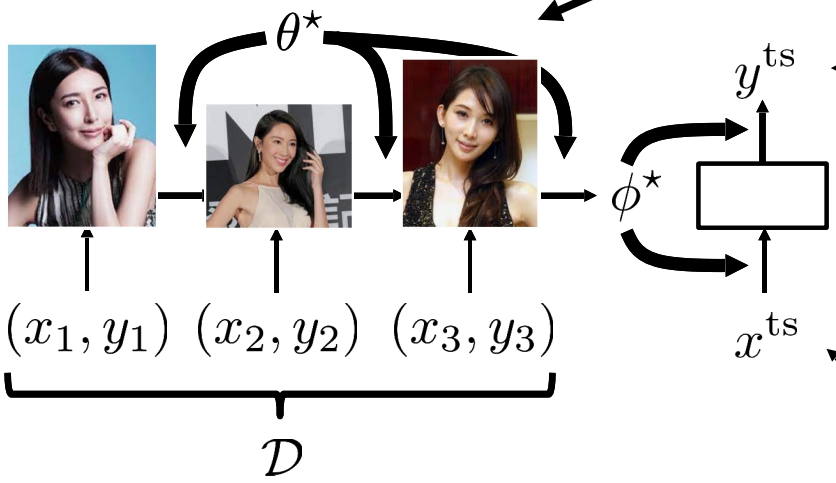
# A Quick Example (cont'd)

- Meta training: $\theta^\star = \arg\max_\theta \log p(\theta|\mathcal{D}_{\text{meta-train}})$
- Meta testing: $\phi^\star = \arg\max_\phi \log p(\phi|\mathcal{D}, \theta^\star)$

$\mathcal{D}_{\text{meta-train}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$

$\mathcal{D} = \{(x_1, y_1), \ldots, (x_k, y_k)\}$

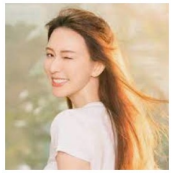
meta-testing




test label

[1, 0, 0]

test input

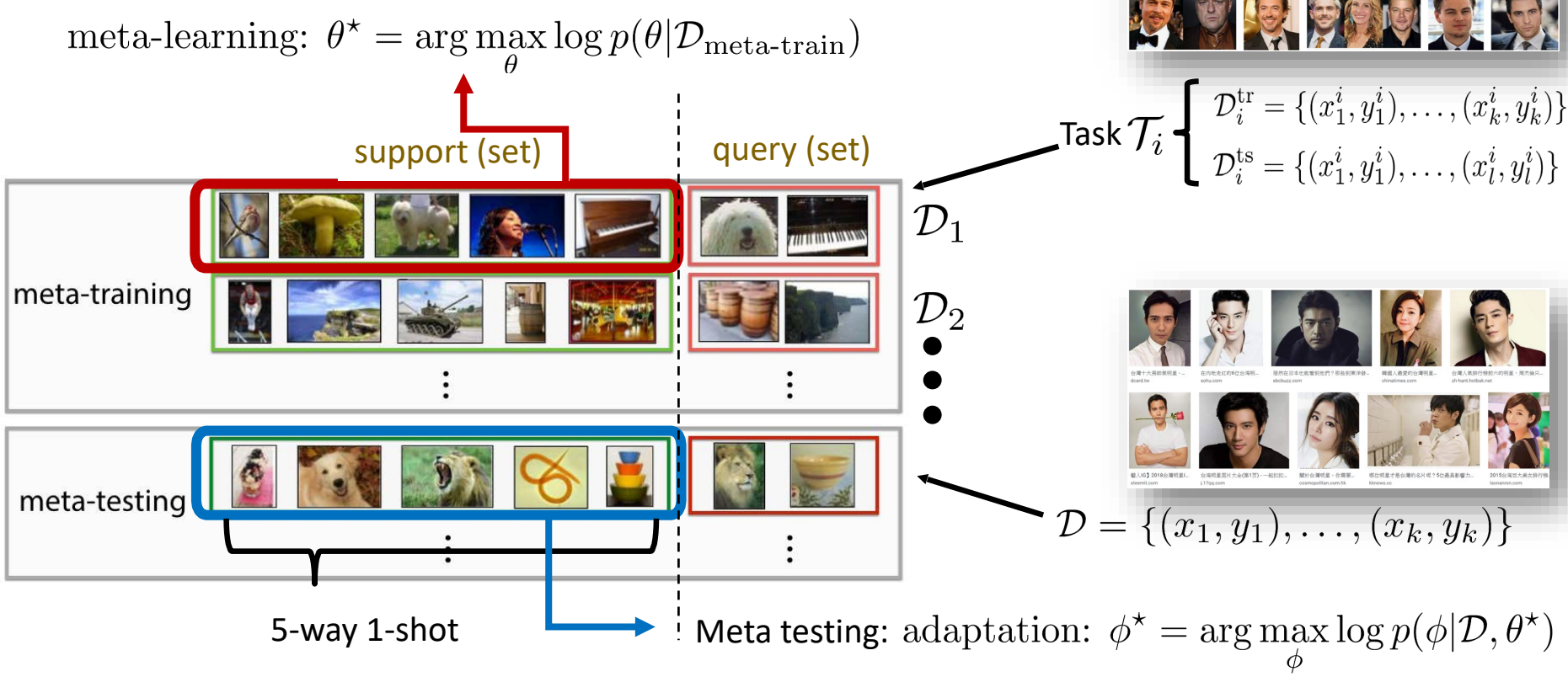


✓ **Key Idea:**

The condition/mechanism of meta-training and meta-testing must match.

In other words, meta learning is to learn the mechanism, ***not*** to fit the data/labels.

# Meta-Learning Terminology




✓ **Key Idea:**

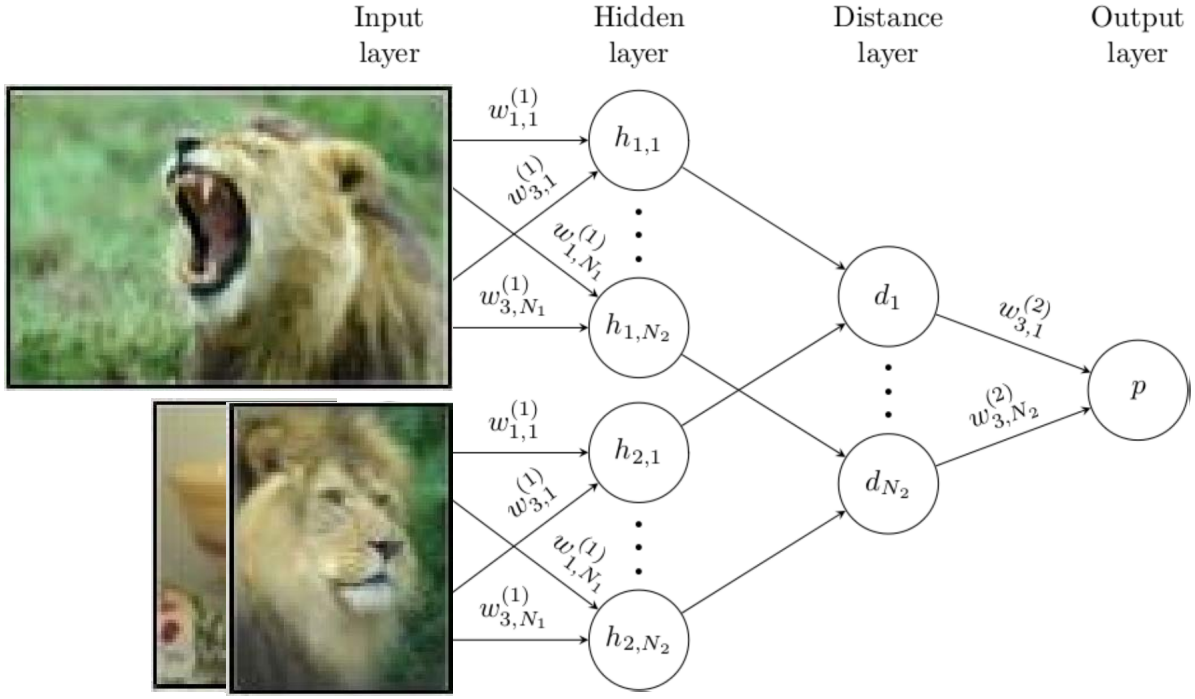
- Meta learning: learn a N-way K-shot learning mechanism, ***not*** fitting data/labels
- The conditions (i., N-way K-shot) of meta-training and meta-testing must match.

# What to Cover Today…
# Few-Shot Learning & Its Applications

- Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# Approach #1:
# Non-Parametric Approach (Mechanistic View)

- Can models learn to compare?
- E.g., Siamese Network
  - Learn a network to determine whether a pair of images are of the same category.



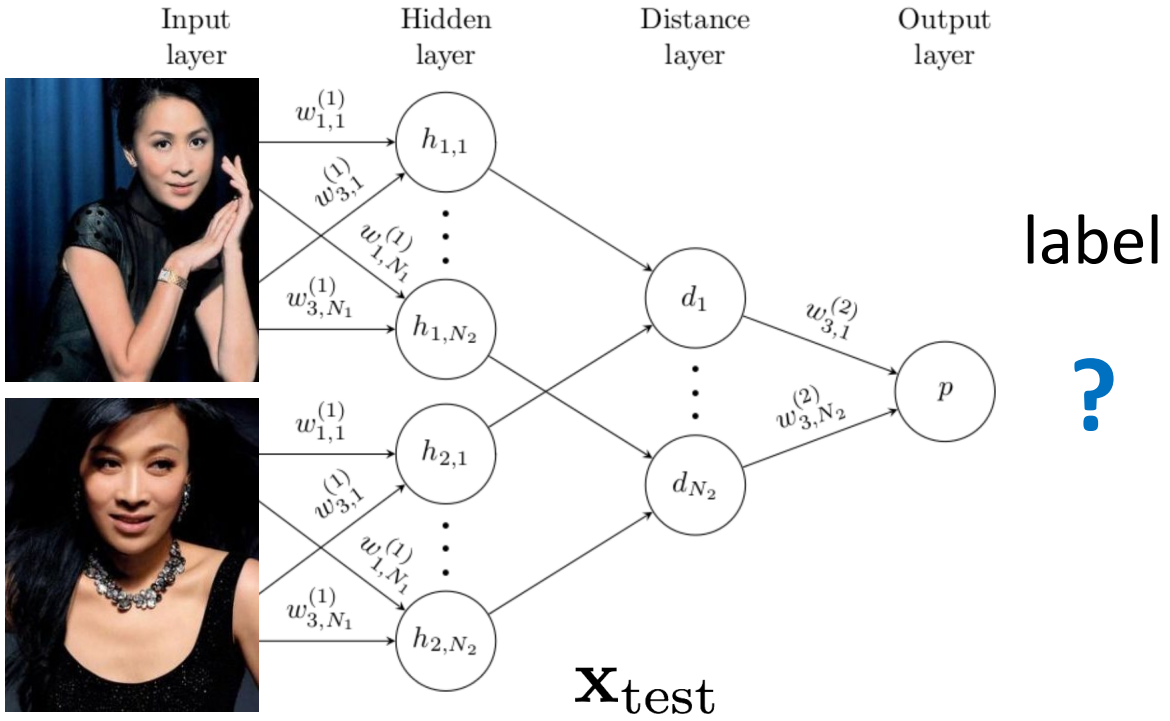

label

**1**

Koch et al., Siamese Neural Networks for One-Shot Image Recognition, ICML WS 2015

# Learn to Compare (cont'd)

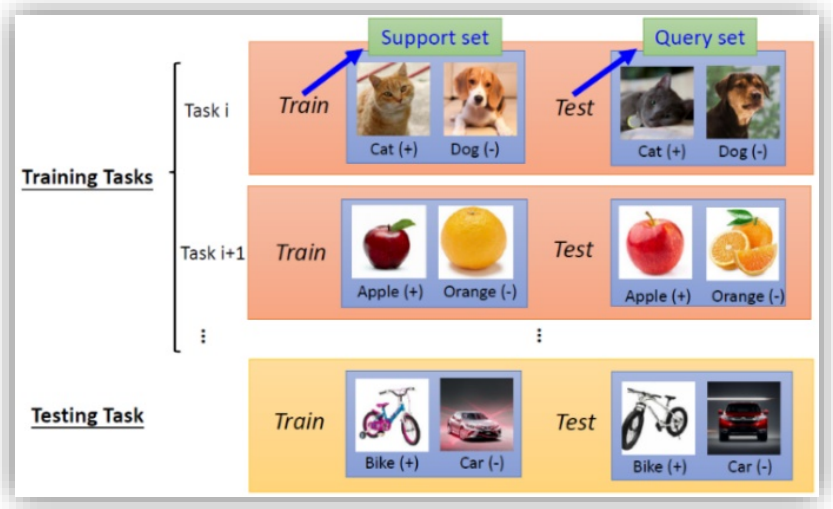
- Siamese Network (cont'd)
  - Meta-training/testing: learn to match (i.e., 2-way image matching)
    - Question: output label of the following example is 1 or 0? (i.e., same ID or not)
  - And, can we perform multi-way classification (beyond matching)?



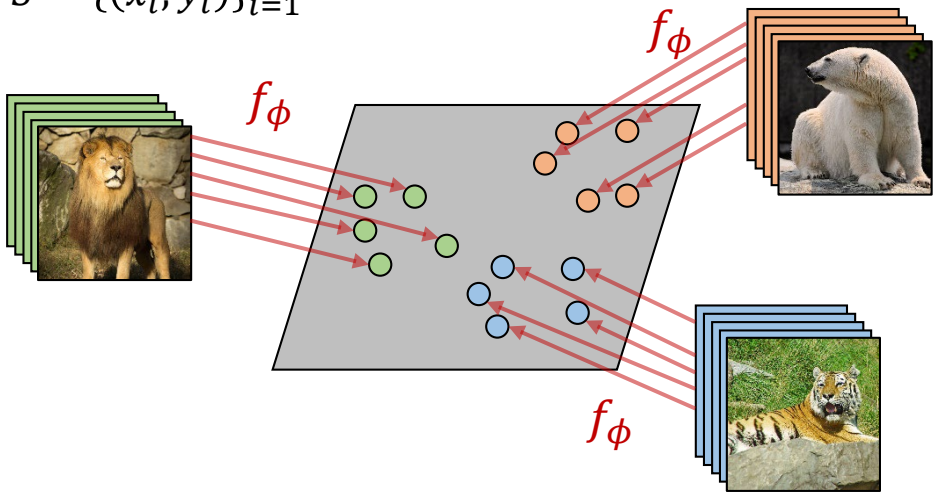

Koch et al., Siamese Neural Networks for One-Shot Image Recognition, ICML WS 2015

# Learn to Compare…with the Representative Ones!

- **Prototypical Networks**
  - Learn a model which properly describes data in terms of intra/inter-class info.
  - It learns a prototype for each class, with data similarity/separation guarantees.

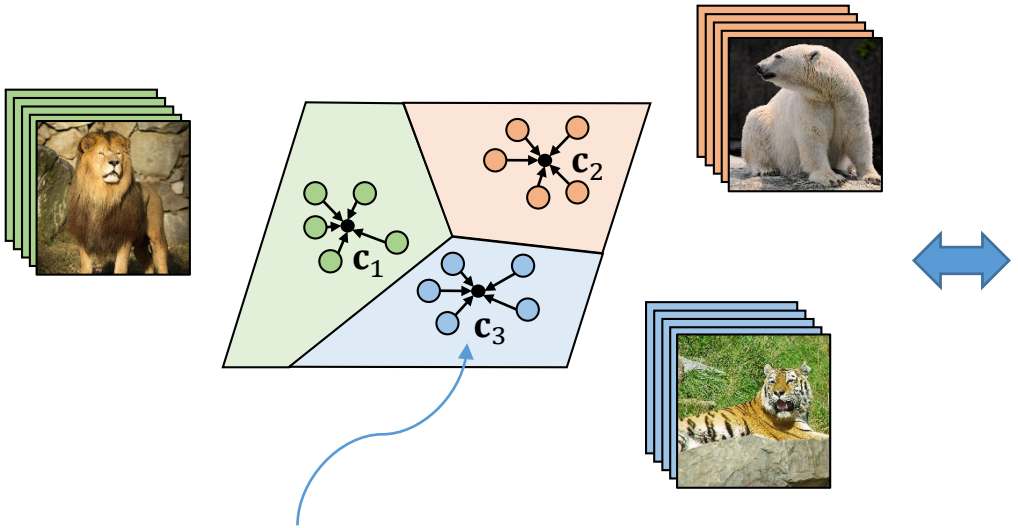


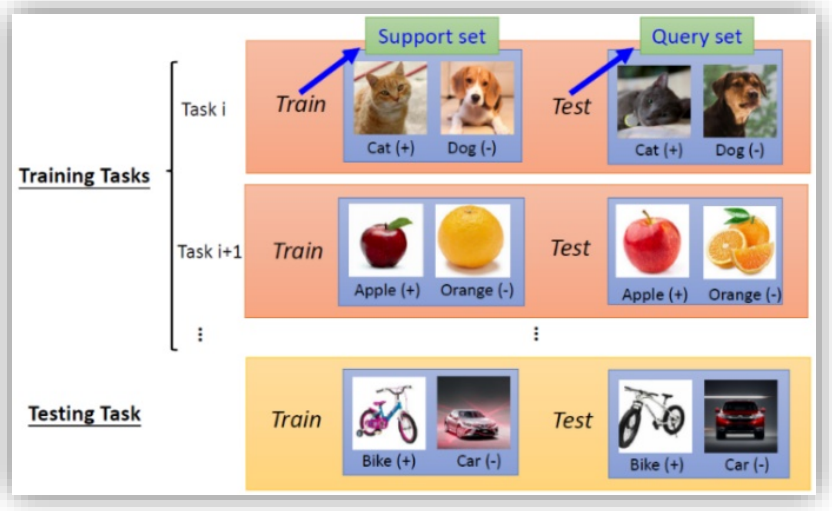

support set
$S = \{(x_i, y_i)\}_{i=1}^{k}$




Snell et al., Prototypical Networks for Few-Shot Learning, NIPS 2017

- **Prototypical Networks** (cont'd)
  - Learn a model which properly describes data in terms of intra/inter-class info.
  - It learns a prototype for each class, with data similarity/separation guarantees.
  - For DL version, the above embedding space is derived by a non-linear mapping $f_\phi$ and the representatives (or anchors) of each class is the **mean feature vector** $\mathbf{c}_k$.




$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i)$$, where $S_k \subset S$ is the subset of support set $S$ with class $k$

Snell et al., Prototypical Networks for Few-Shot Learning, NIPS 2017

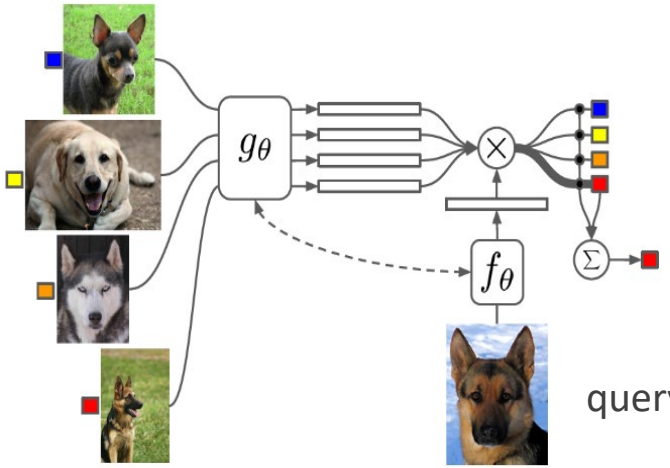
# Learn to Compare

- **Matching Networks**
  - Inspired by the **attention** mechanism,
    access an augmented memory containing useful info to solve the task of interest
  - The authors proposed a weighted nearest-neighbor classifier,
    with attention over a learned embedding from the support set $S = \{(x_i, y_i)\}_{i=1}^{k}$,
    so that the label of the query $\hat{x}$ can be predicted.

$c(.,.)$: cosine similarity

$$\hat{y} = \sum_{i=1}^{k} a(\hat{x}, x_i) y_i \quad \text{with} \quad a(\hat{x}, x_i) = e^{c(f(\hat{x}), g(x_i))} / \sum_{j=1}^{k} e^{c(f(\hat{x}), g(x_j))}$$

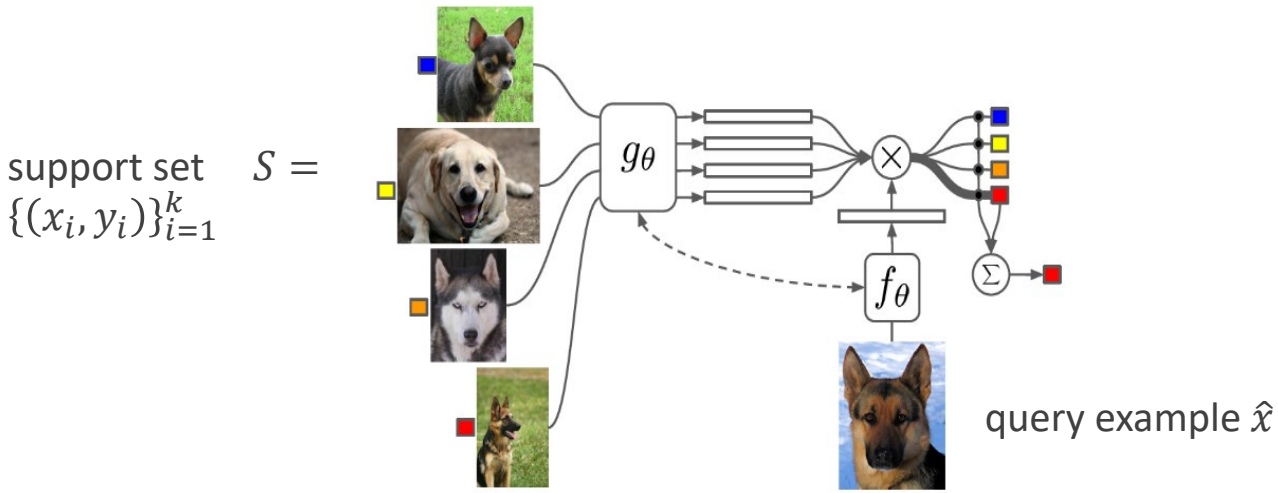
support set $S = \{(x_i, y_i)\}_{i=1}^{k}$




Vinyals et al., "Matching Networks for One Shot Learning," NIPS, 2016

# Learn to Compare

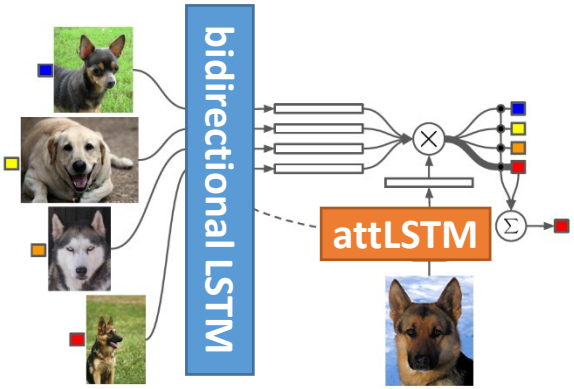
- **Matching Networks** (cont'd)
  - Simple form: $g = f$
  - Similar to Siamese network
  - Also similar to prototypical network for **one**-shot learning




Vinyals et al., "Matching Networks for One Shot Learning," NIPS, 2016

- **Matching Networks** (cont'd)
  - Full context embedding (FCE):
  - Each element in $S$ should not be embedded independently of other elements
    - $g(x_i) \rightarrow g(S)$ as a **bidirectional LSTM** by considering the whole $S$ as a **sequence**
  - Also, $S$ should be able to modify the way we embed $\hat{x}$
    - $f(\hat{x}) \rightarrow f(\hat{x}, S)$ as an **LSTM** with **read-attention** over $g(S)$: $\text{attLSTM}(f'(\hat{x}), g(S), K)$, where $f'(\hat{x})$ is the (fixed) CNN feature, and $K$ is the number of unrolling steps
  - Experiment results on *mini*ImageNet

support set
$S = \{(x_i, y_i)\}_{i=1}^{k}$



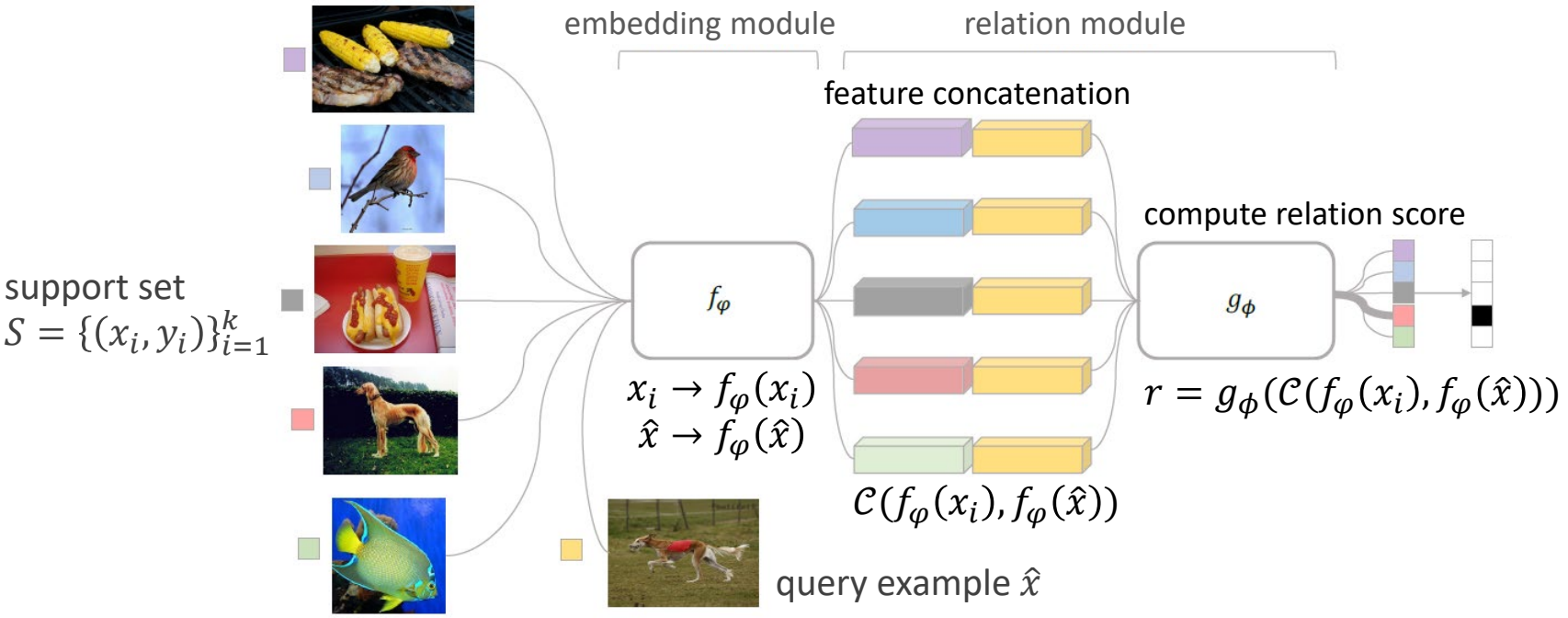

query example $\hat{x}$

| Model | Matching Fn | Fine Tune | 5-way Acc 1-shot | 5-way Acc 5-shot |
|---|---|---|---|---|
| PIXELS | Cosine | N | 23.0% | 26.6% |
| BASELINE CLASSIFIER | Cosine | N | 36.6% | 46.0% |
| BASELINE CLASSIFIER | Cosine | Y | 36.2% | 52.2% |
| BASELINE CLASSIFIER | Softmax | Y | 38.4% | 51.2% |
| MATCHING NETS (OURS) | Cosine | N | 41.2% | 56.2% |
| MATCHING NETS (OURS) | Cosine | Y | 42.4% | 58.0% |
| MATCHING NETS (OURS) | Cosine (FCE) | N | 44.2% | 57.0% |
| MATCHING NETS (OURS) | Cosine (FCE) | Y | **46.6%** | **60.0%** |

Vinyals et al., "Matching Networks for One Shot Learning," NIPS, 2016

# Learn to Compare with Self-Learned Metrics

- **Relation Network**
  - Metric-learning approaches typically focus on learning an embedding function with a **fixed metric** (e.g., Euclidean distance, cosine similarity, …)
  - The authors proposed to train a **Relation Network** (RN) to explicitly learn a transferrable **deep distance metric** comparing the relation between images

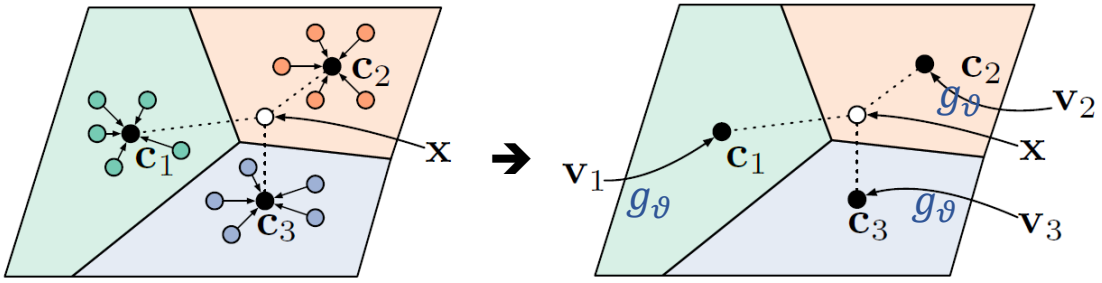



Sung et al., "Learning to Compare: Relation Network for Few-Shot Learning," CVPR, 2018

# Relation Networks (cont'd)

- Some works can be extended to **zero-shot learning** (if time permits):
  - Instead of few-shot images, the support set contains a **semantic embedding vector** ($\mathbf{v}_k$) for each of the training classes.
  - Thus, we can use a second **heterogeneous** embedding function to embed the semantic embedding vectors.
  - Extension of **Prototypical Network**:

$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i) \quad \rightarrow \quad \mathbf{c}_k = g_\vartheta(\mathbf{v}_k)$$

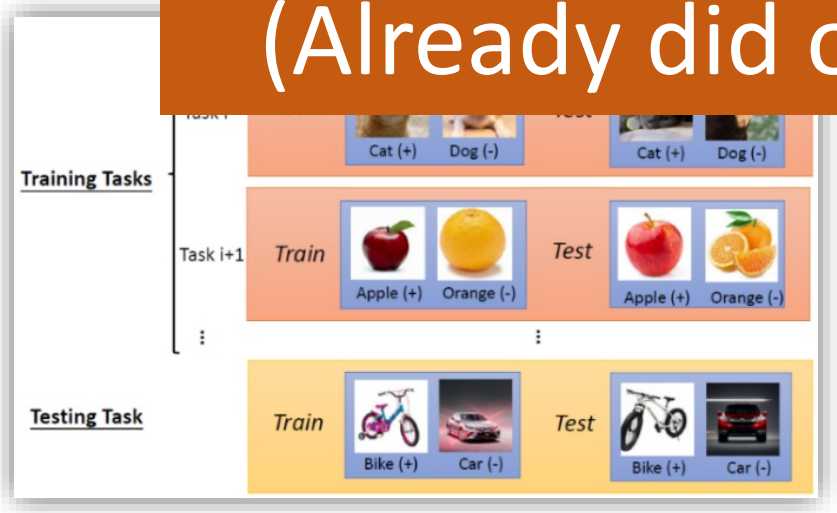


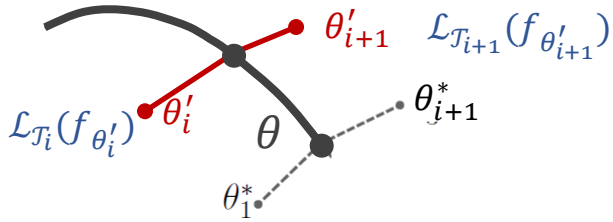

  - Relation Networks: $r = g_\phi(\mathcal{C}(f_\varphi(x_i), f_\varphi(\hat{x}))) \quad \rightarrow \quad r = g_\phi(\mathcal{C}(f_{\varphi_2}(\mathbf{v}_k), f_{\varphi_1}(\hat{x})))$

# Optimization-Based Approach (Probabilistic View)

- **M**odel-**A**gnostic **M**eta-**L**earning (MAML)*
  - Key idea:
    - Train over many tasks (with a small amount of data & few gradient steps), so that the learned parameter Θ would generalize to novel tasks
    - Learning to initialize/fine-tune
  - Meta-Learner $\Phi \rightarrow \Theta_0$:
    - Le… th…
    - Th… …/effectively.

**Will Not Cover This Today!**
**(Already did on July 8th)**




optimize model parameter $\theta$ so that it can quickly adapt to new tasks

*Finn, Abbeel, Levine, ICML 2017

# Some Takeaways for Existing Meta-Learning Approaches

**Parametric-based**

+ handles **varying** & **large K** well
+ **structure lends well** to **out-of-distribution tasks**

- **second-order optimization**

**Non-parametric based**

+ **simple**
+ entirely **feedforward**
+ **computationally fast** & **easy to optimize**

- **harder to generalize** to **varying K**
- hard to scale to **very large K**
- so far, **limited to classification**

Generally, well-tuned versions of each perform **comparably** on existing FSL benchmarks.

# What to Cover Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

Selected slide credits: C. Finn, S. Levine, & H.-Y. Lee

# Learning with Small Data
# → Learning with *Hallucinated* Data

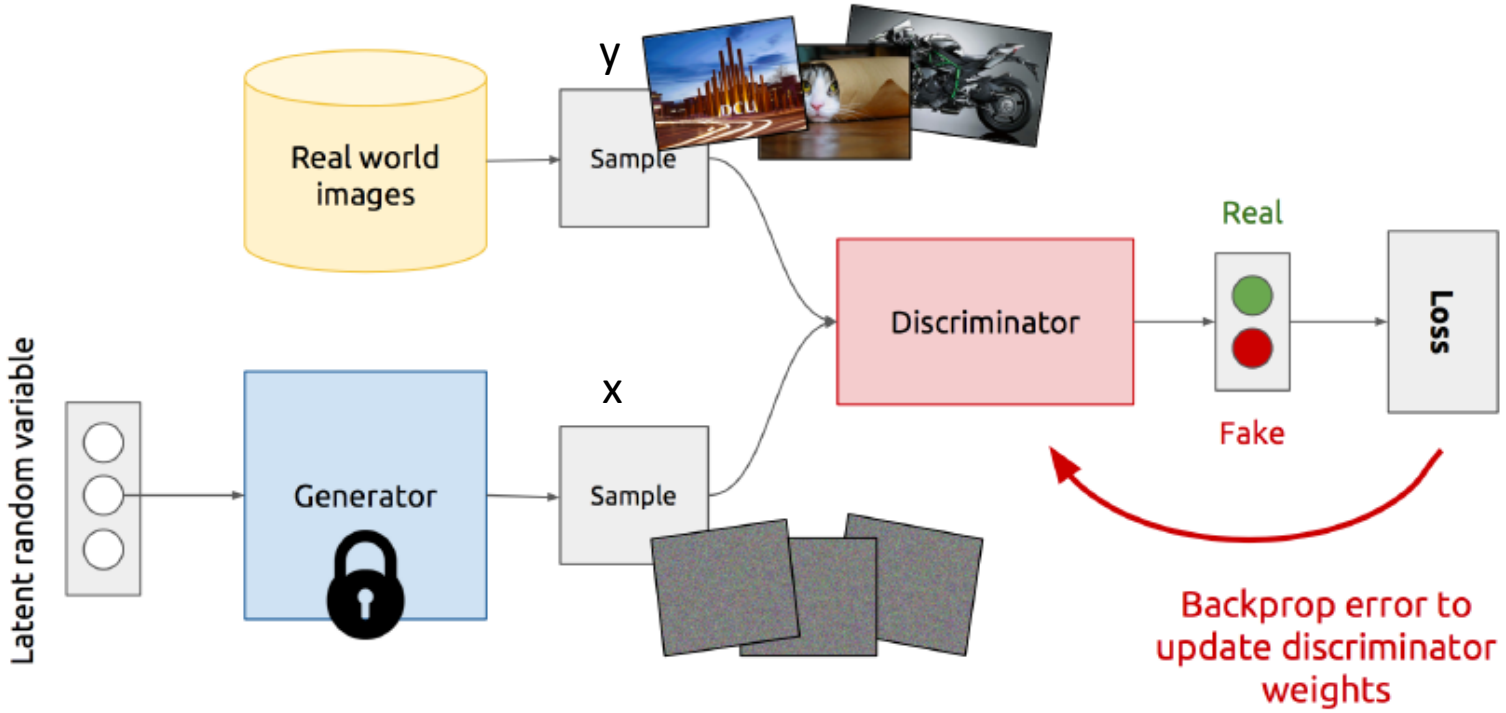
- Data Hallucination
  - Intra-class variations may be shared across categories.
    (e.g., camera pose, translation, lighting changes, and even articulation)
  - One can possibly *hallucinate* additional samples for novel classes


  - However, standard *data augmentation* techniques only utilize a limited amount of **a priori known invariances** (e.g., translations, rotations, flips, color jittering, etc.), failing to sufficiently describe desirable intra/inter-class information…

Wang et al., "Low-Shot Learning from Imaginary Data," CVPR, 2018

# A Super Brief Intro/Review for *Generative Adversarial Networks (GAN)*

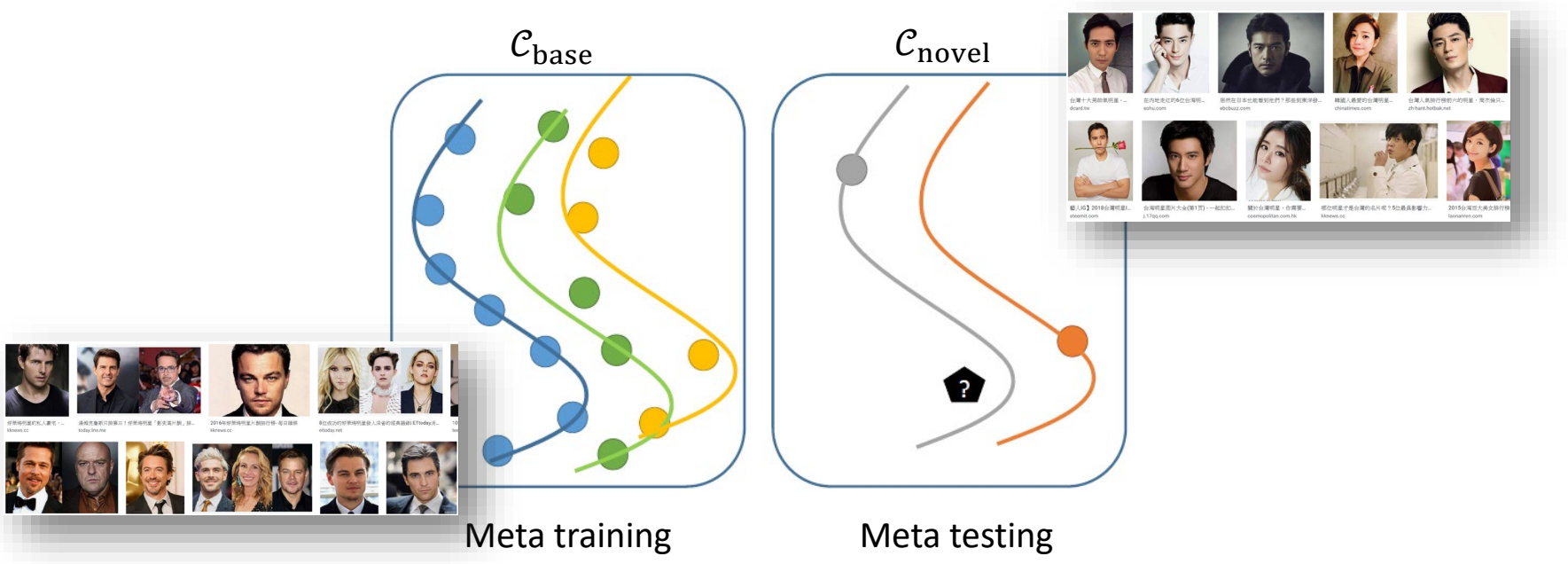
- Design of GAN
  - Loss: $\mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D\ (G(x)))] + \mathbb{E}[\log D(y)]$




Goodfellow et al., Generative Adversarial Nets, NIPS, 2014

# Learn to Augment…Data Hallucination for FSL (1/3)

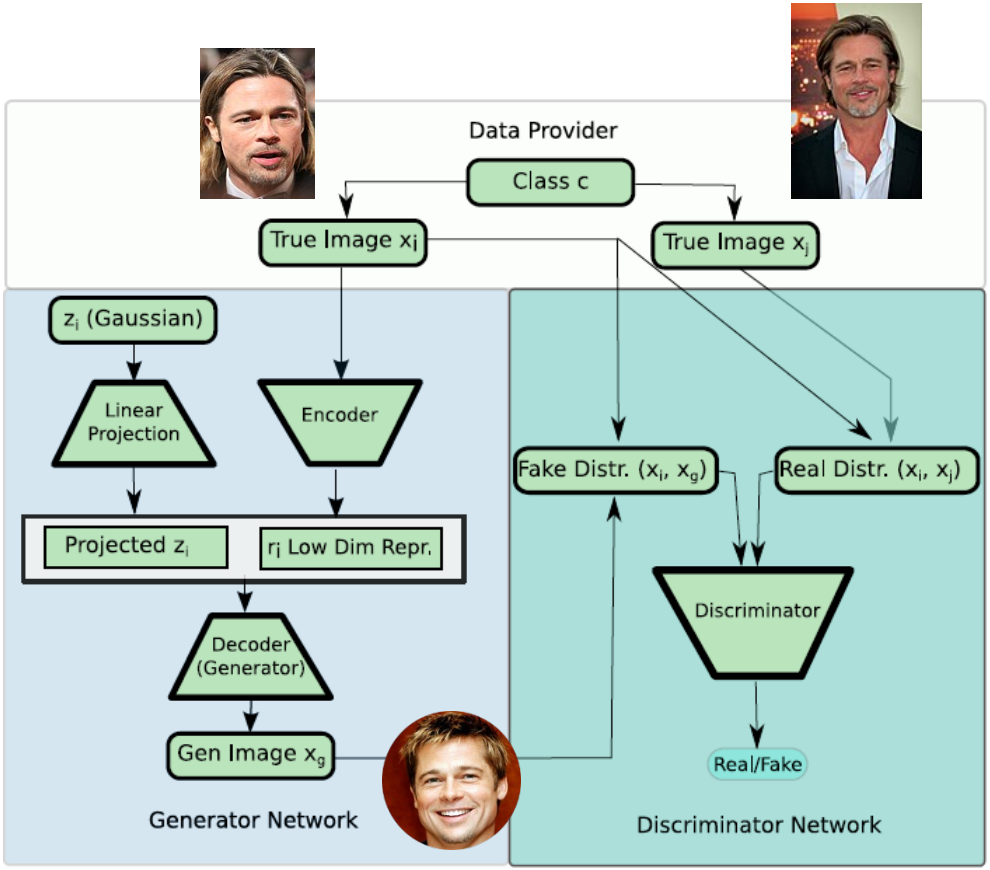
- Data Hallucination by Conditional GAN
  - Can we learn a model resulting in a desirable **invariance space**, which can be derived by a conditional GAN in the **source domain** ($\mathcal{C}_{\text{base}}$), and apply it to the **target domain** ($\mathcal{C}_{\text{novel}}$)?




Antoniou et al., "Data Augmentation Generative Adversarial Networks," ICLR Workshop, 2018

- Data Augmentation GAN

(Left) Generator
$\mathbf{r}_i = Enc(\mathbf{x}_i)$
$\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{I})$
$\mathbf{x}_g = Dec(\mathbf{z}_i, \mathbf{r}_i)$




Discriminator:
$D(\mathbf{x}_i, \mathbf{x}_j)$ ➔ Real **pair**
$D(\mathbf{x}_i, \mathbf{x}_g)$ ➔ Fake **pair**

**Question:**
Why not verify $\mathbf{x}_j$ and $\mathbf{x}_g$?
i.e., why conditioned on $\mathbf{x}_i$?

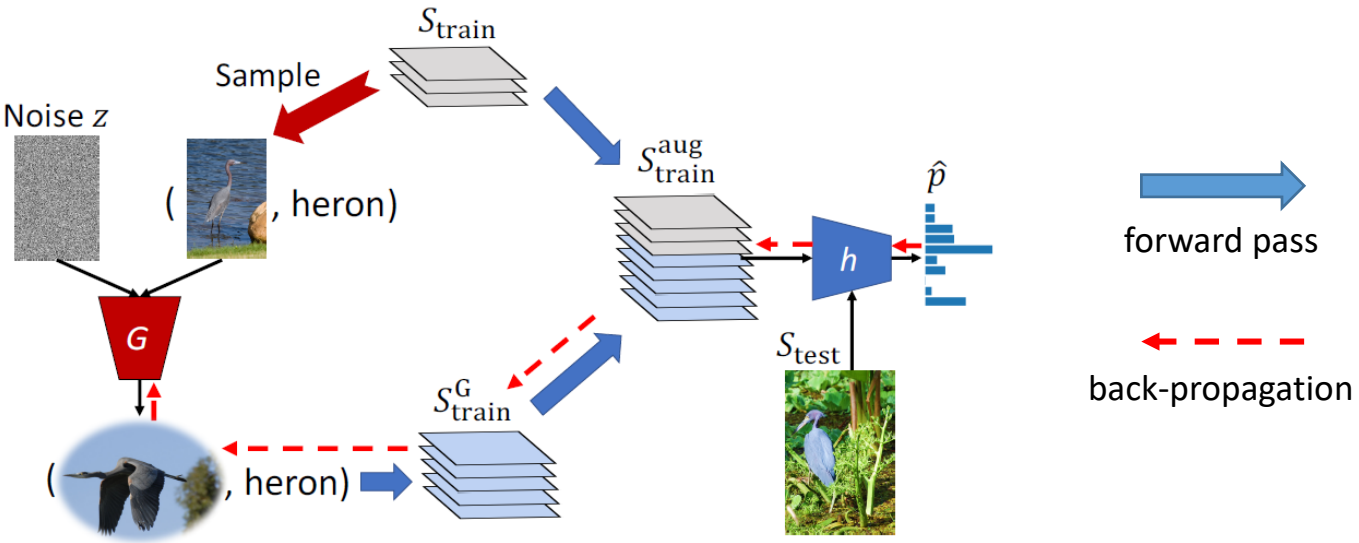
(1) prevent the generator from simply output the original image $\mathbf{x}_i$
(2) to improve diversity (aka. mode collapse)

➔ (1) or (2) or…?

Antoniou et al., "Data Augmentation Generative Adversarial Networks," ICLR Workshop, 2018

# Learn to Augment…Data Hallucination for FSL (2/3)

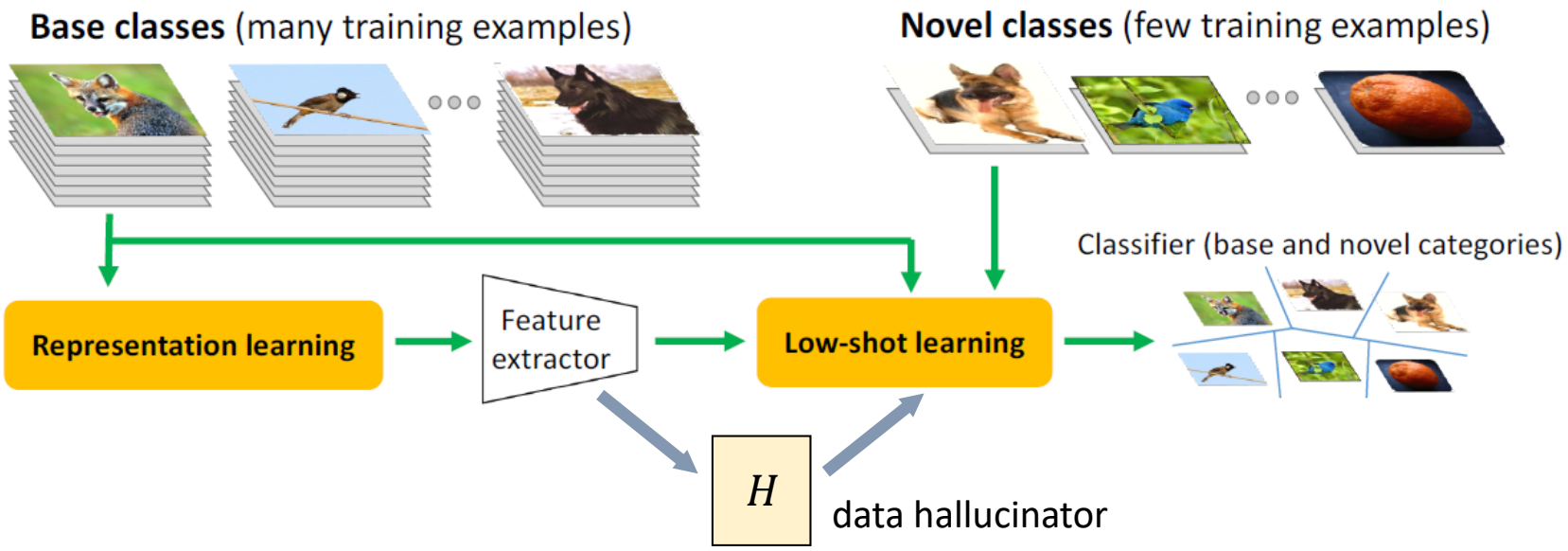
- Hallucinator Jointly Trained with FSL Classifiers
  - The hallucinated examples should be **useful** for classification tasks, rather than just being **diverse** or **realistic** (that may fail to improve FSL performances).
  - The authors proposed to train a **conditional-GAN-based** data hallucinator ($G(x, z)$) **jointly** with the meta-learning module ($h$) in an **end-to-end** manner.




Wang et al., Low-shot learning from imaginary data, CVPR 2018

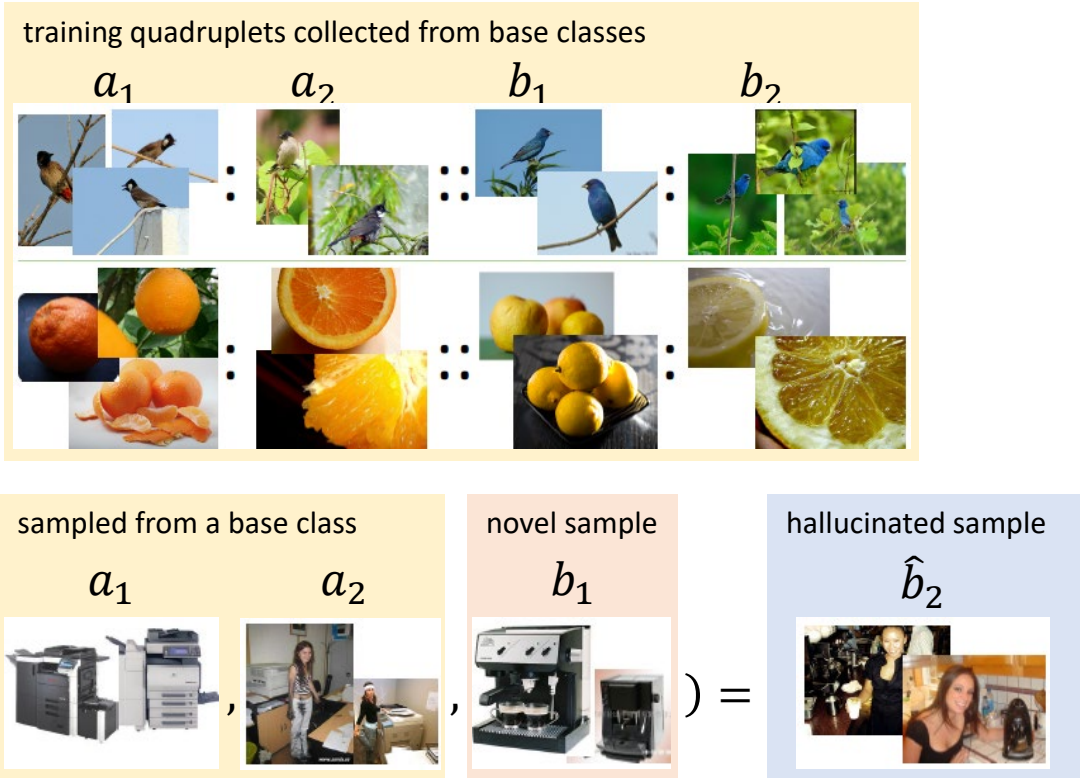
# Learn to Augment…Data Hallucination for FSL (3/3)

- Data Hallucination by Analogy
  - Modern recognition models are trained on large labeled datasets like ImageNet
  - To deal with the above challenges faced by **recognition systems in the wild**, one can alternatively consider the FSL benchmark in the following two phases:




Hariharan et al., Low-shot Visual Recognition by Shrinking and Hallucinating Features, ICCV, 2017

- Hallucination by Analogy (cont'd)
  - Analogy-based Data Hallucinator
    - Train $H$ using **analogy quadruplets** $(a_1, a_2, b_1, b_2)$,
      where $(a_1, a_2)$ belong to some class, $(b_1, b_2)$ belong to another class, and $a_1 : a_2 :: b_1 : b_2$ holds.



# What to Cover Today…
# Few-Shot Learning & Its Applications

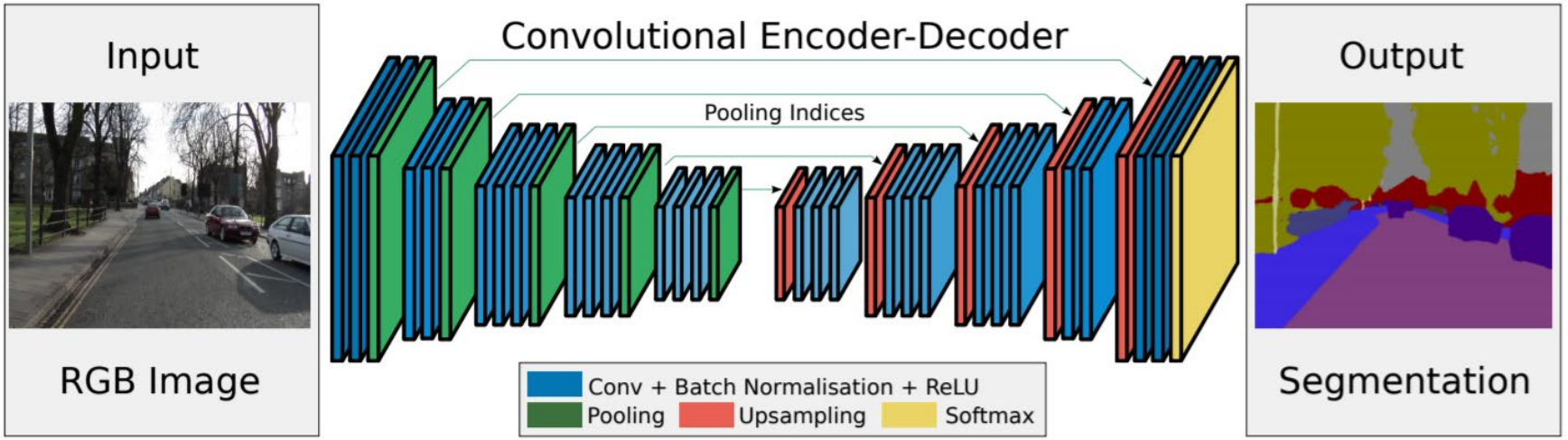
- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# Semantic Segmentation

- Goal
  - Assign a class label to each pixel in the input image
  - Don’t differentiate instances, only care about pixels

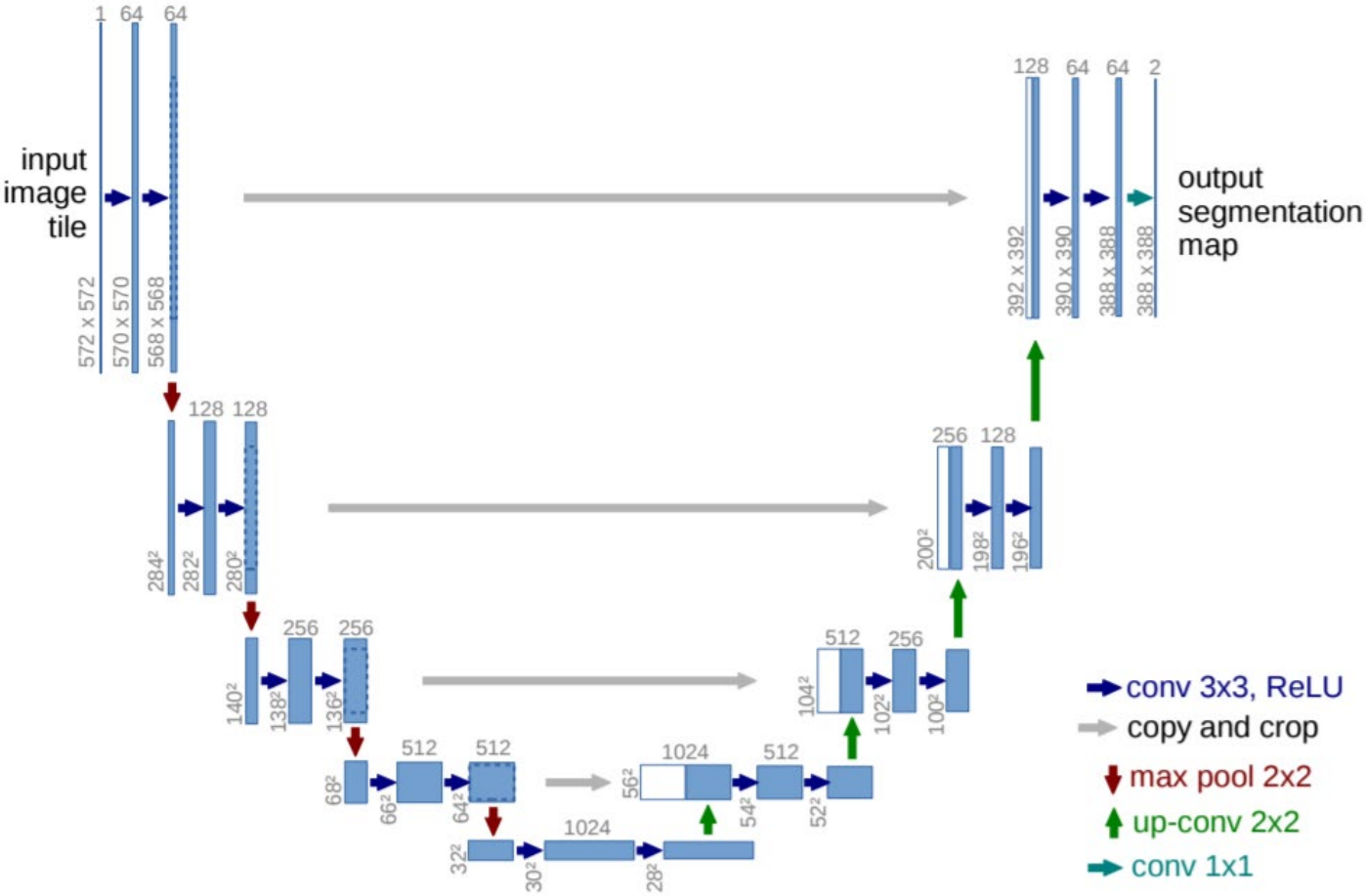
# SegNet

- Efficient architecture (memory + computation time)
- Upsampling reusing max-unpooling indices
- Reasonable results without performance boosting addition
- Comparable to FCN




"SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation" [link]

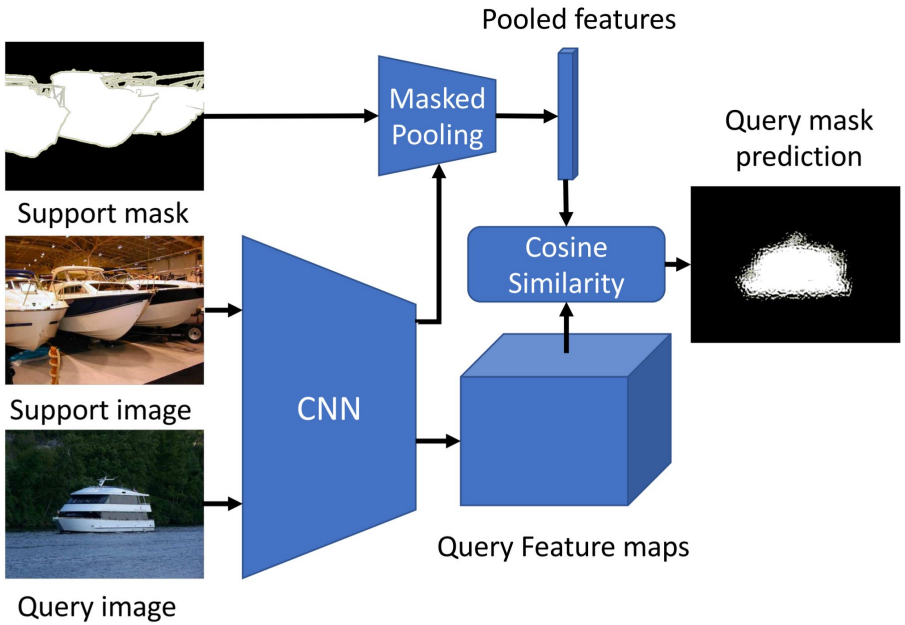
# U-Net




U-Net: Convolutional Networks for Biomedical Image Segmentation [link]

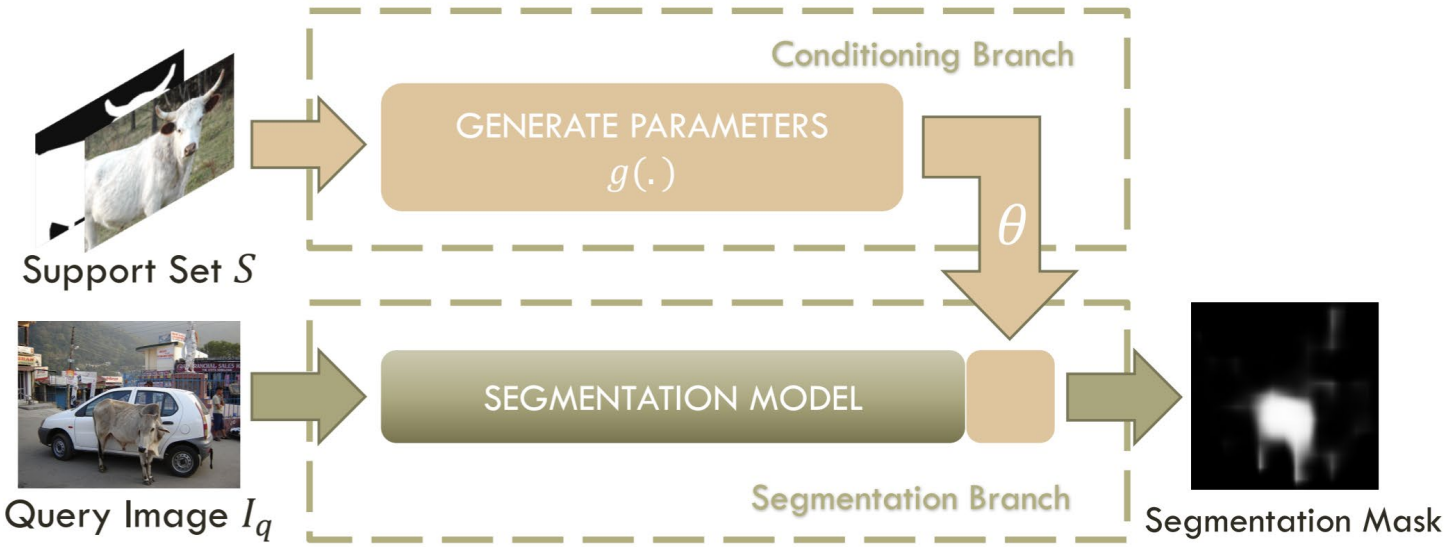
# Few-Shot Segmentation

- A large number of image categories are with pixel-wise ground truth labels, while a number of classes them are with limited amounts of GT pixel-wise labels.
- A **shared CNN backbone** produces feature maps for both support and query images.
- **Prototypes** for each class is obtained by **masked pooling** from support feature maps.
- Query feature maps are then compared with the prototypes in a **pixel-by-pixel** fashion.
- Typically, **cosine similarity** is adopted for pixel-wise feature comparison.




Meta-learning for few-shot segmentation

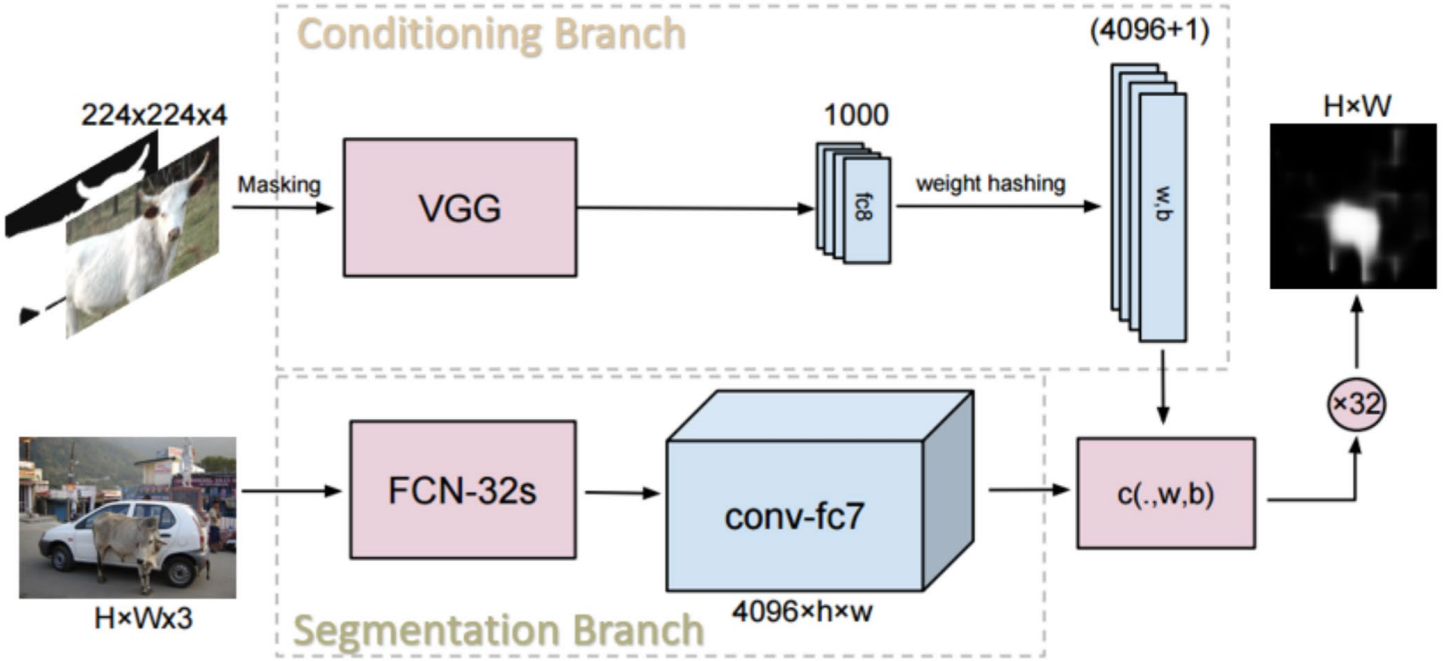
# OSLSM [BMVC 2017]




- $S$ is an annotated image from a new semantic class
- Input $S$ to a function $g$ that outputs a set of parameters $\theta$
- $\theta$ is used to parameterize part of the segmentation model which produces a segmentation mask given $I_q$

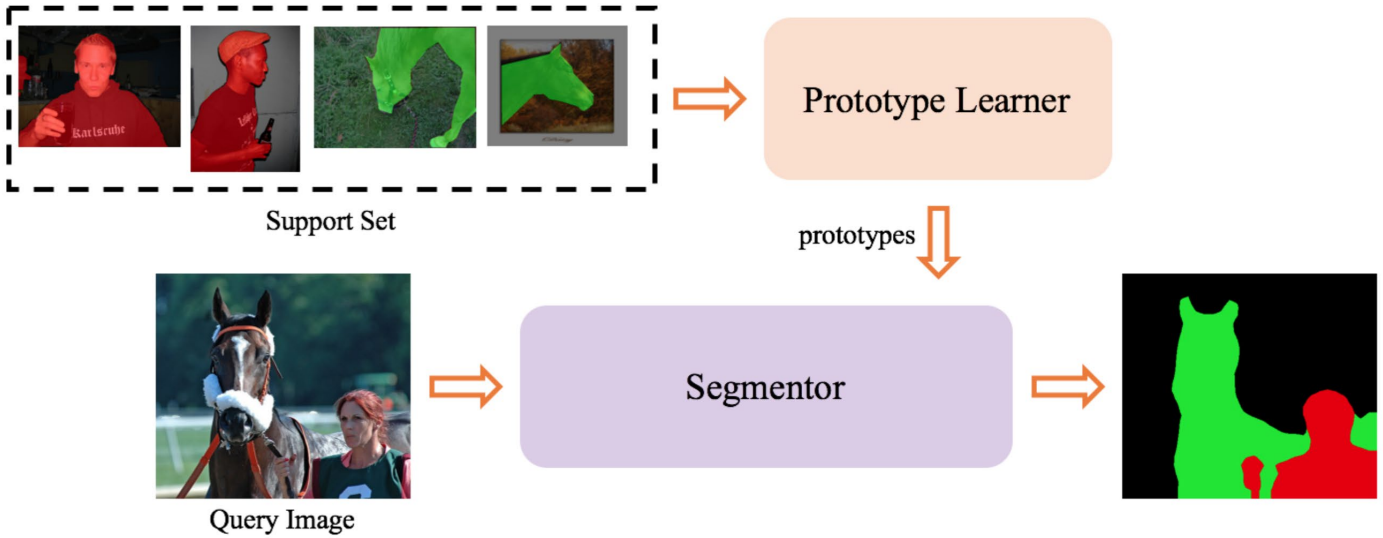
Shaban, Amirreza, et al. "One-shot learning for semantic segmentation." BMVC 2017

# OSLSM [BMVC 2017]




Shaban, Amirreza, et al. "One-shot learning for semantic segmentation." BMVC 2017

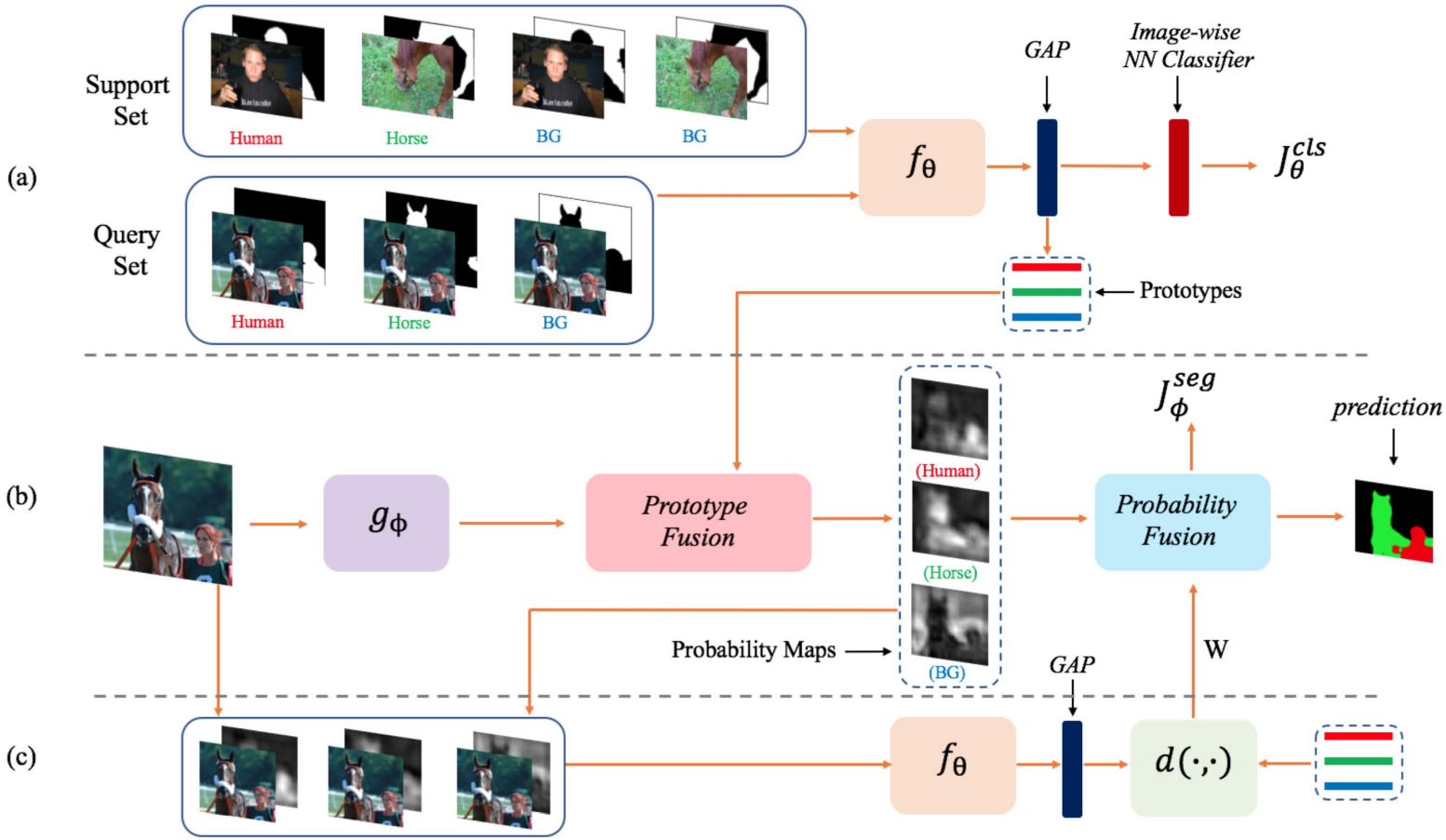
# Prototype Learning [BMVC 2018]




- A prototype is learned for each foreground class and the background class.
- Prototypes are used to predict rough segmentation maps for each class.
- The final prediction is optimized using probability fusion.

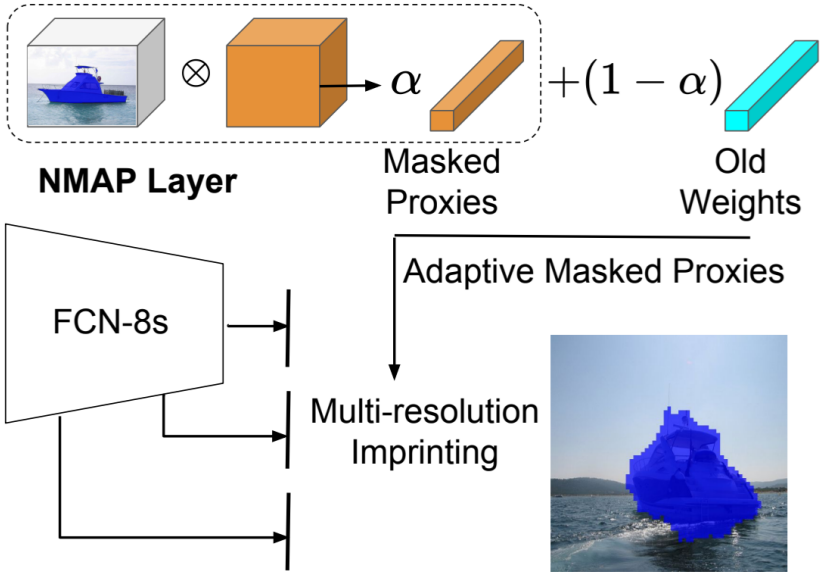
Dong, Nanqing, and Eric Xing. "Few-Shot Semantic Segmentation with Prototype Learning." BMVC. 2018.

# PL [BMVC 2018]




Dong, Nanqing, and Eric Xing. "Few-Shot Semantic Segmentation with Prototype Learning." BMVC. 2018.

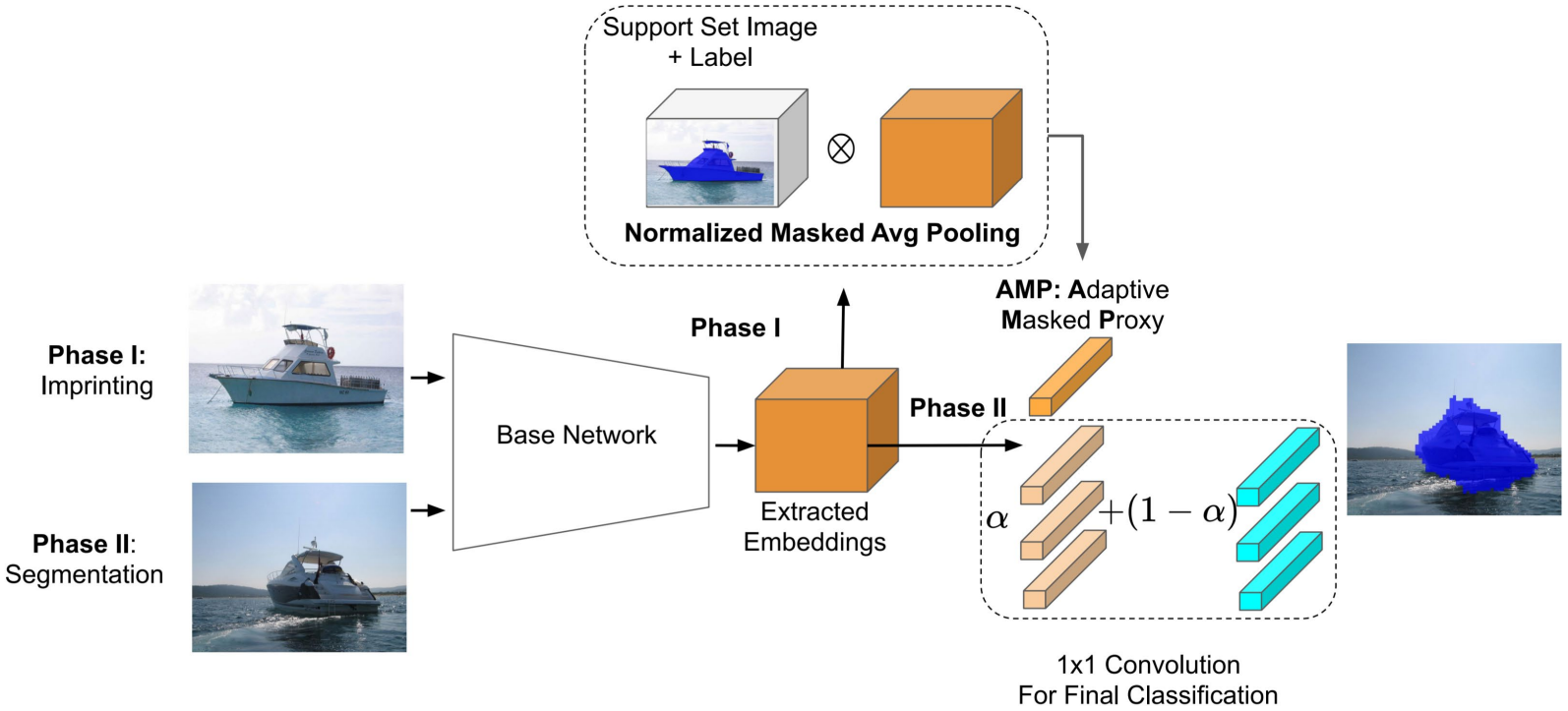
# AMP [ICCV 2019]




- Adaptive masked proxies (i.e., prototypes') are extracted for ach semantic class.
- Proxies update themselves in a continuous stream of data (e.g., video).
- Proxies from different resolution levels are used in multi-resolution imprinting

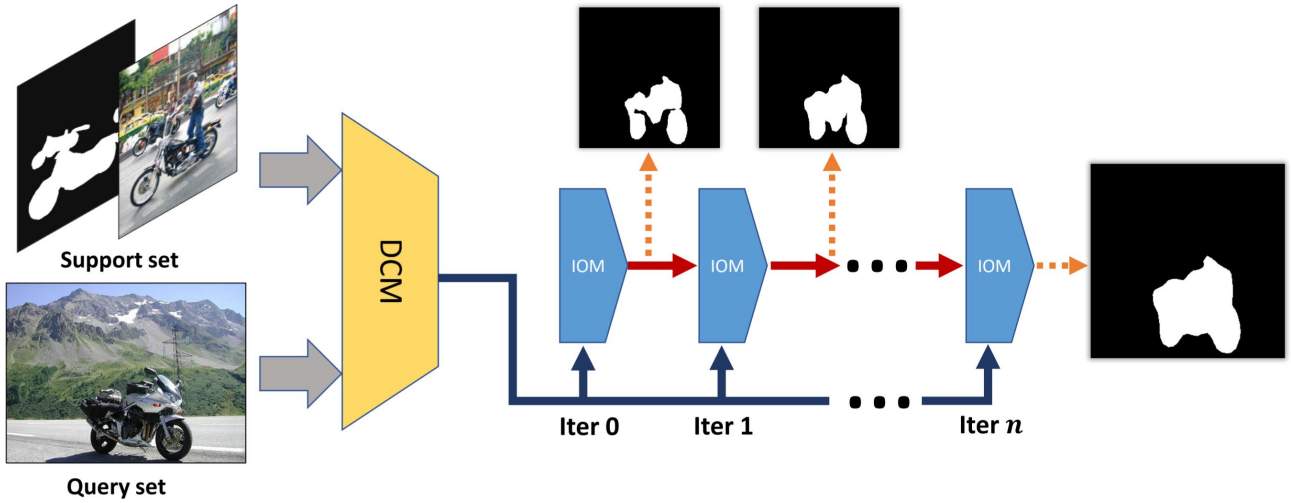
Jagersand et al., AMP: Adaptive masked proxies for few-shot segmentation, ICCV 2019.

# AMP [ICCV 2019]




Jagersand et al., AMP: Adaptive masked proxies for few-shot segmentation, ICCV 2019.

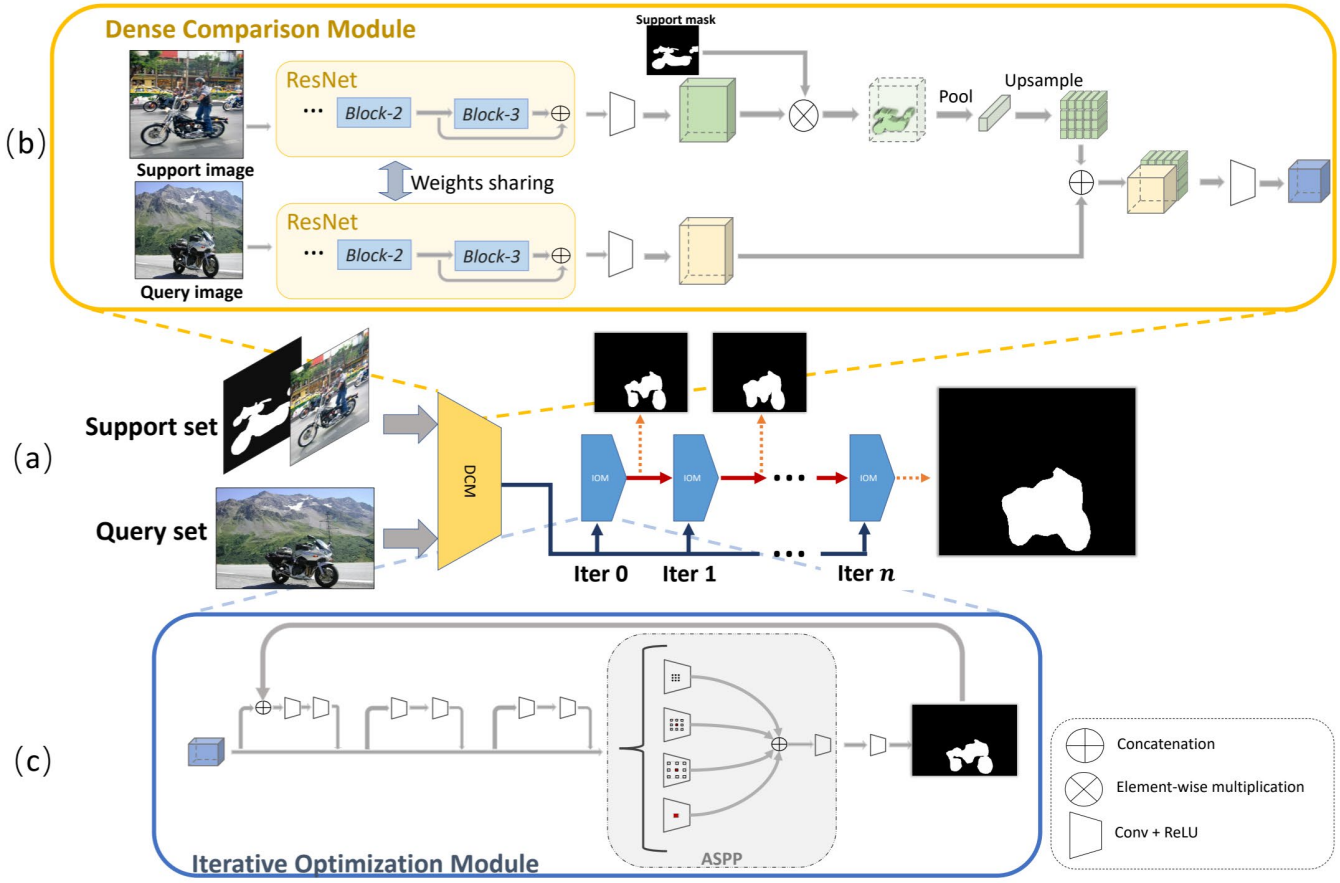
# CANet [CVPR 2019]




- Dense comparison module (DCM) concatenates prototypes to each spatial location in query feature map
- Rough segmented maps are produced after comparing with mask-pooled feature prototypes
- The final result is optimized in an iterative manner

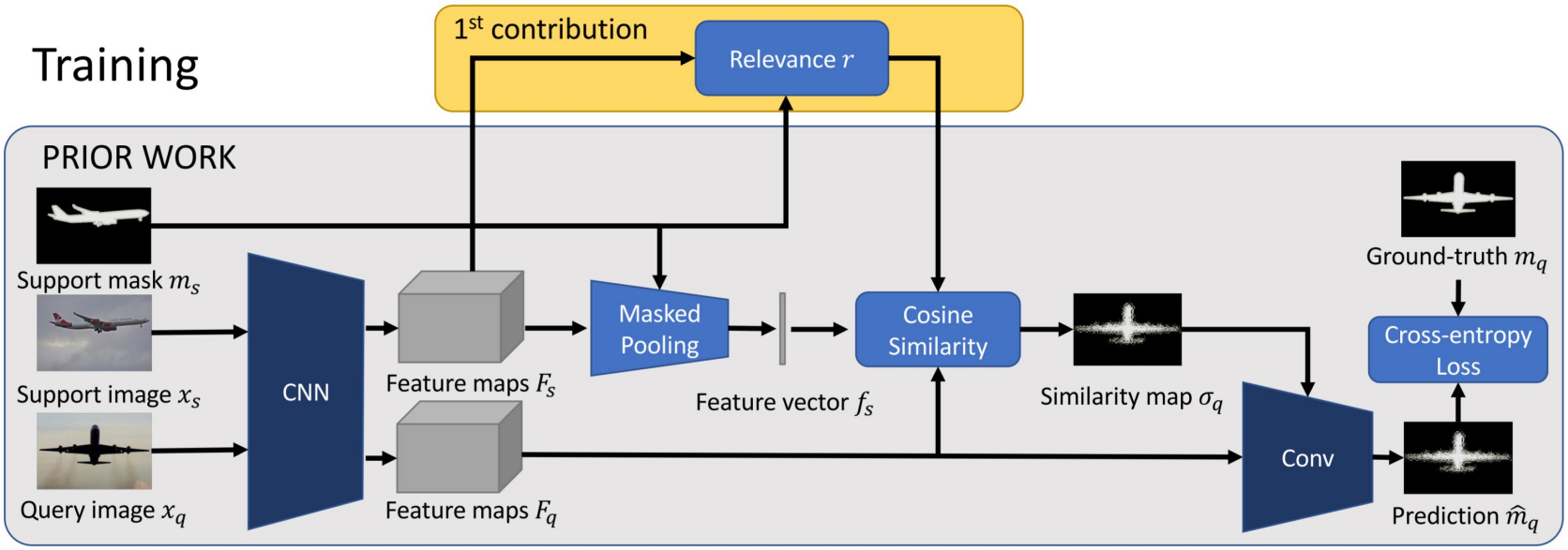
Chi, et al. CANet: Class-agnostic segmentation networks with iterative refinement and attentive FSL, CVPR 2019.

# CANet [CVPR 2019]




Chi, et al. CANet: Class-agnostic segmentation networks with iterative refinement and attentive FSL, CVPR 2019.

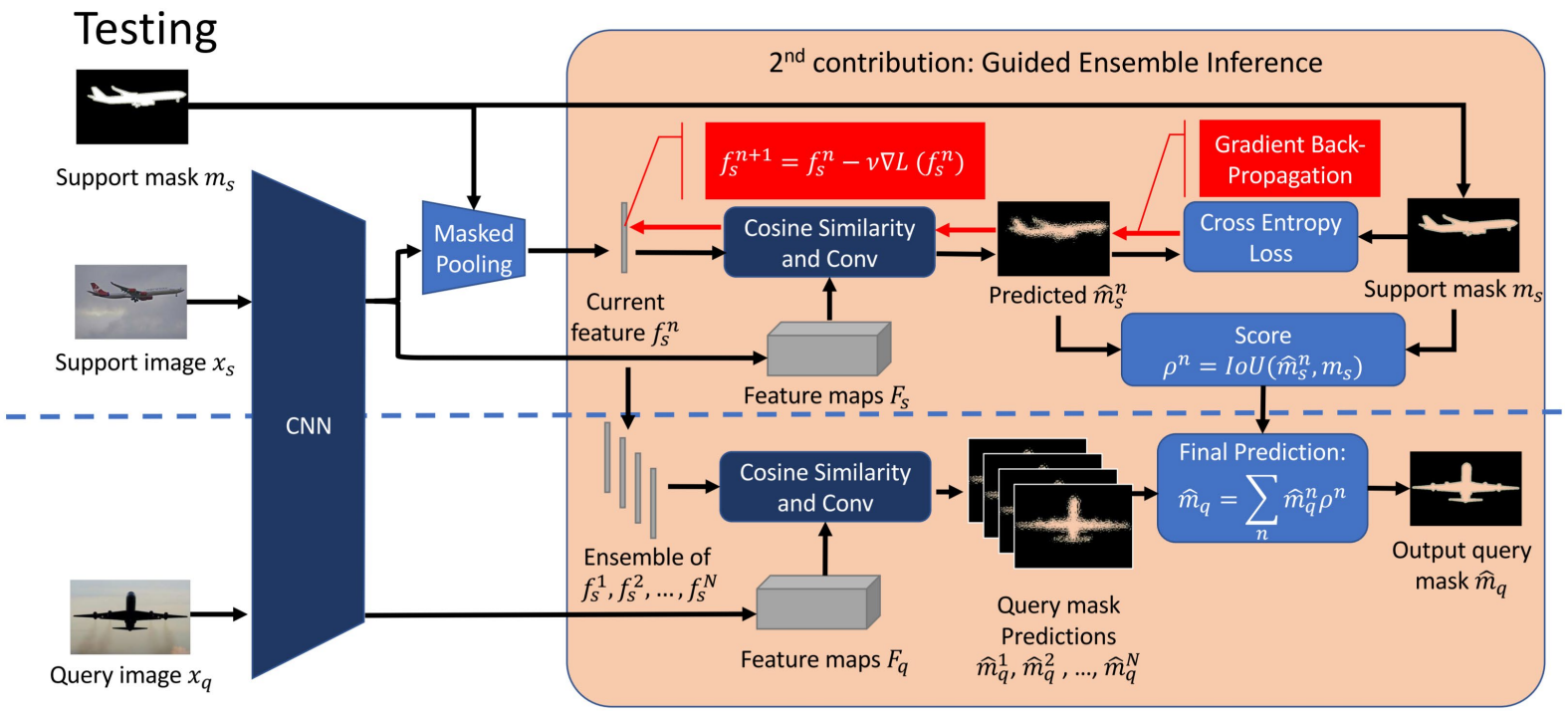
# FWB [ICCV 2019]




- Standard FSL methods (e.g., shared backbone, masked pooling…) are used during training.
- A ‘relevance’ factor is added and taken into account during cosine similarity computation.

Todorovic et al., Feature weighting and boosting for few-shot segmentation, ICCV 2019.

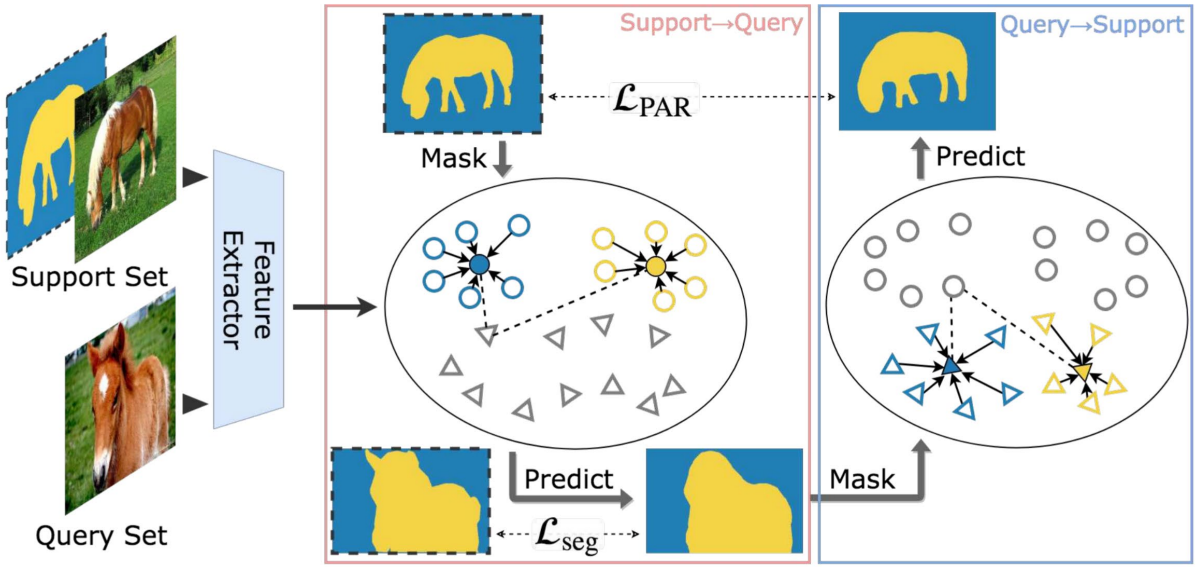
# FWB [ICCV 2019]




- During inference, ensemble is utilized to select the best set of parameters
- Prototypes are used to predict the support masks reversely, which can be compared to the ground truth.

Todorovic et al., Feature weighting and boosting for few-shot segmentation, ICCV 2019.

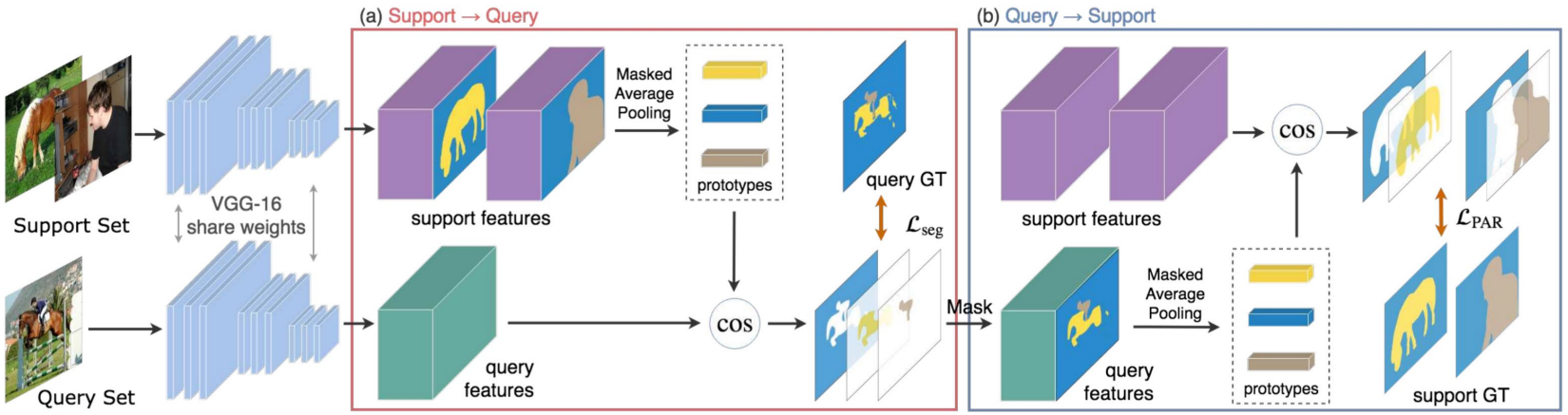
# PANet [ICCV 2019]




- Extracted prototypes are first used to predict query masks, as standard FSL methods do.
- Predicted query masks are used to generate new prototypes and reversely predict support masks
- Similar concept to that of the “cycle consistency” (support→query; query→support)

Kaixin, et al., Panet: Few-shot image semantic segmentation with prototype alignment, ICCV 2019.

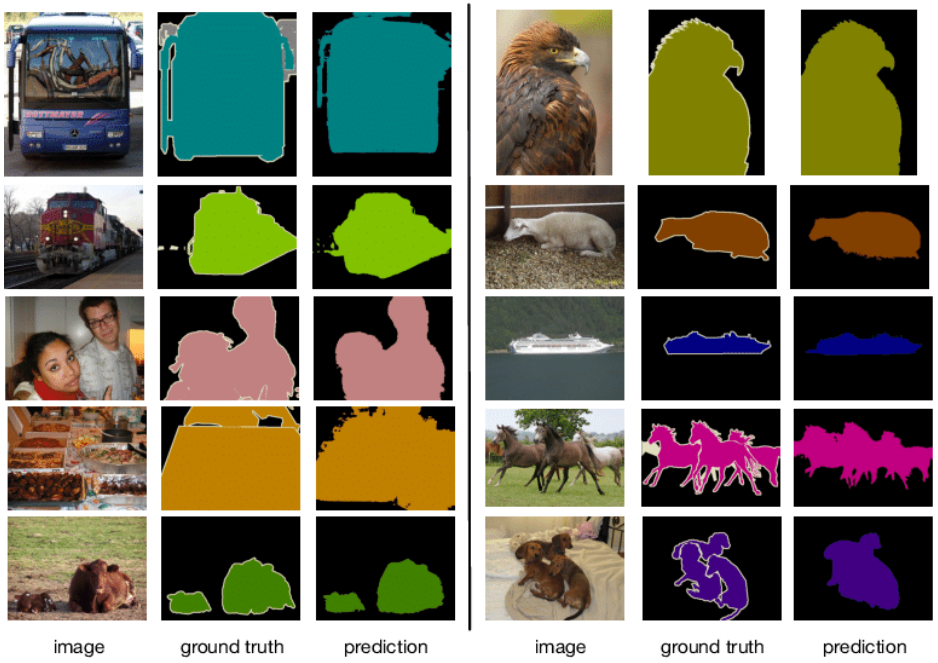
# PANet [ICCV 2019]




Kaixin, et al., Panet: Few-shot image semantic segmentation with prototype alignment, ICCV 2019.

# Dataset & Evaluation Metric

- **Datasets**
  - **PASCAL VOC 2012** (main)
    - 20 classes
    - Split: (15 *base* + 5 *novel*)
  - coco (secondary)
- **Evaluation Metrics**
  - **Binary-mIoU** (difficult)
  - FB-mIoU (easy)
    - Foreground/Background IoU



# Performance Comparisons

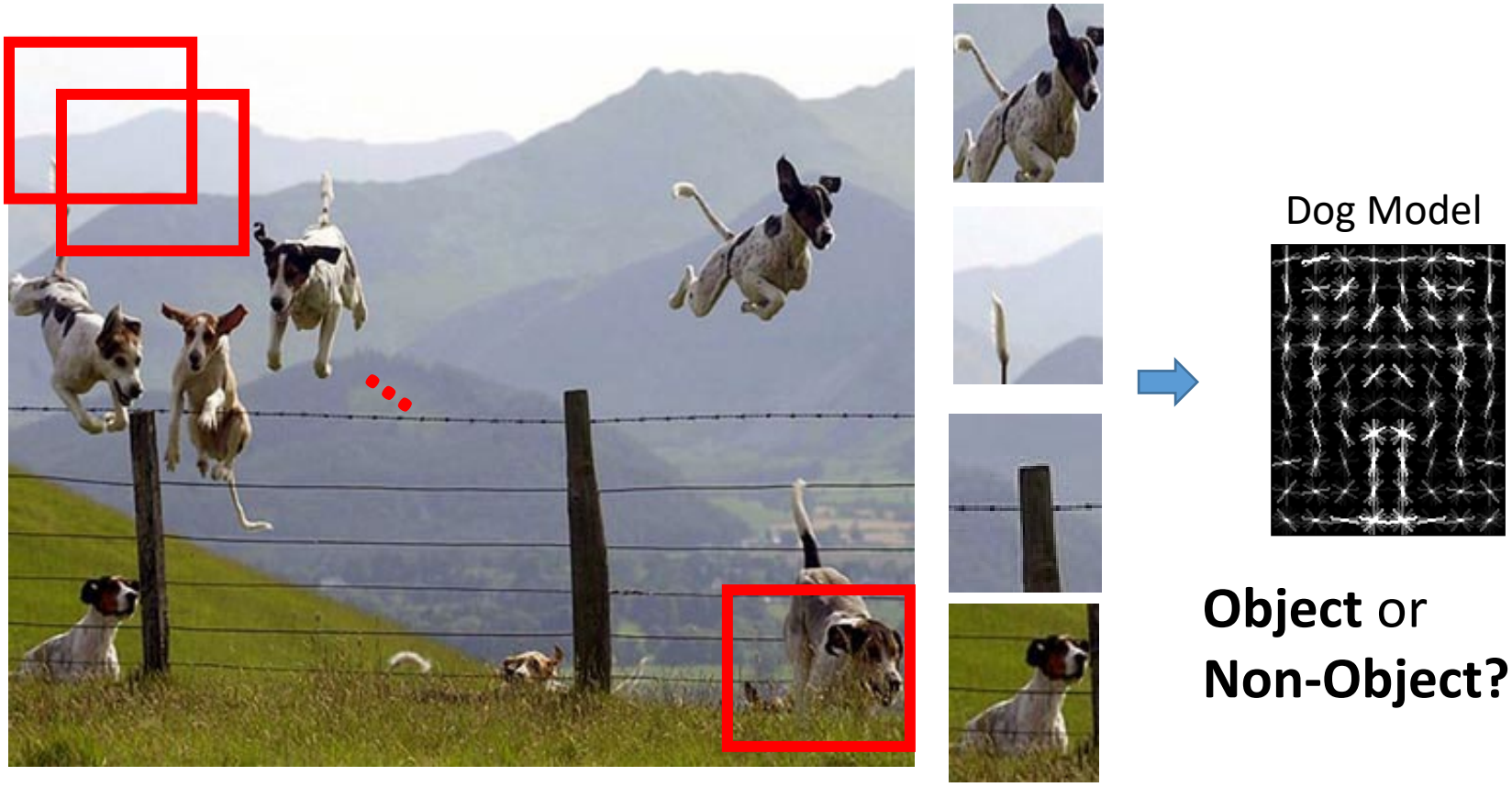
| Method | | Split-0 | Split-1 | Split-2 | Split-3 | Mean |
|---|---|---|---|---|---|---|
| Reduced-DFCN8s | | 39.2 | 48.0 | 39.3 | 34.2 | 40.2 |
| OSLSM | BMVC 2017 | 33.6 | 55.3 | 40.9 | 33.5 | 40.8 |
| co-FCN | ICLRW 2018 | 36.7 | 50.6 | 44.9 | 32.4 | 41.2 |
| AMP | ICCV 2019 | 41.9 | 50.2 | 46.7 | 34.7 | 43.4 |
| SG-One | | 40.2 | 58.4 | 48.4 | 38.4 | 46.4 |
| PANet | ICCV 2019 | 42.3 | 58.0 | 51.1 | 41.2 | 48.1 |
| PRNet | | 51.6 | 61.3 | 53.1 | 47.6 | 53.4 |
| Co-att | | 49.5 | 65.5 | 50.0 | 49.2 | 53.5 |
| CANet | CVPR 2019 | 52.5 | 65.9 | 51.3 | 51.9 | 55.4 |
| PGNet | ICCV 2019 | 56.0 | 66.9 | 50.6 | 50.4 | 56.0 |
| FWB | ICCV 2019 | 51.3 | 64.5 | 56.7 | 52.2 | **56.2** |

# What to Cover Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# Object Detection

- Focus on object search: “Where is it?”
- Build templates that quickly differentiate object patch from background patch



# Two-Stage vs. One-Stage Object Detection

**Methods**

- Sliding Windows
- **Two−stage Frameworks**
  - R−CNN
  - Fast R−CNN
  - Mask R−CNN
  - ⋮
- **One−stage Frameworks**
  - YOLO
  - YOLOv2
  - YOLOv3
  - ⋮

# Few-Shot Object Detection

- What if one cannot collect a sufficient amount of training data for the objects of interest? → **Small Data Problem!**
- Applications: defect detection, medical image analysis, etc.

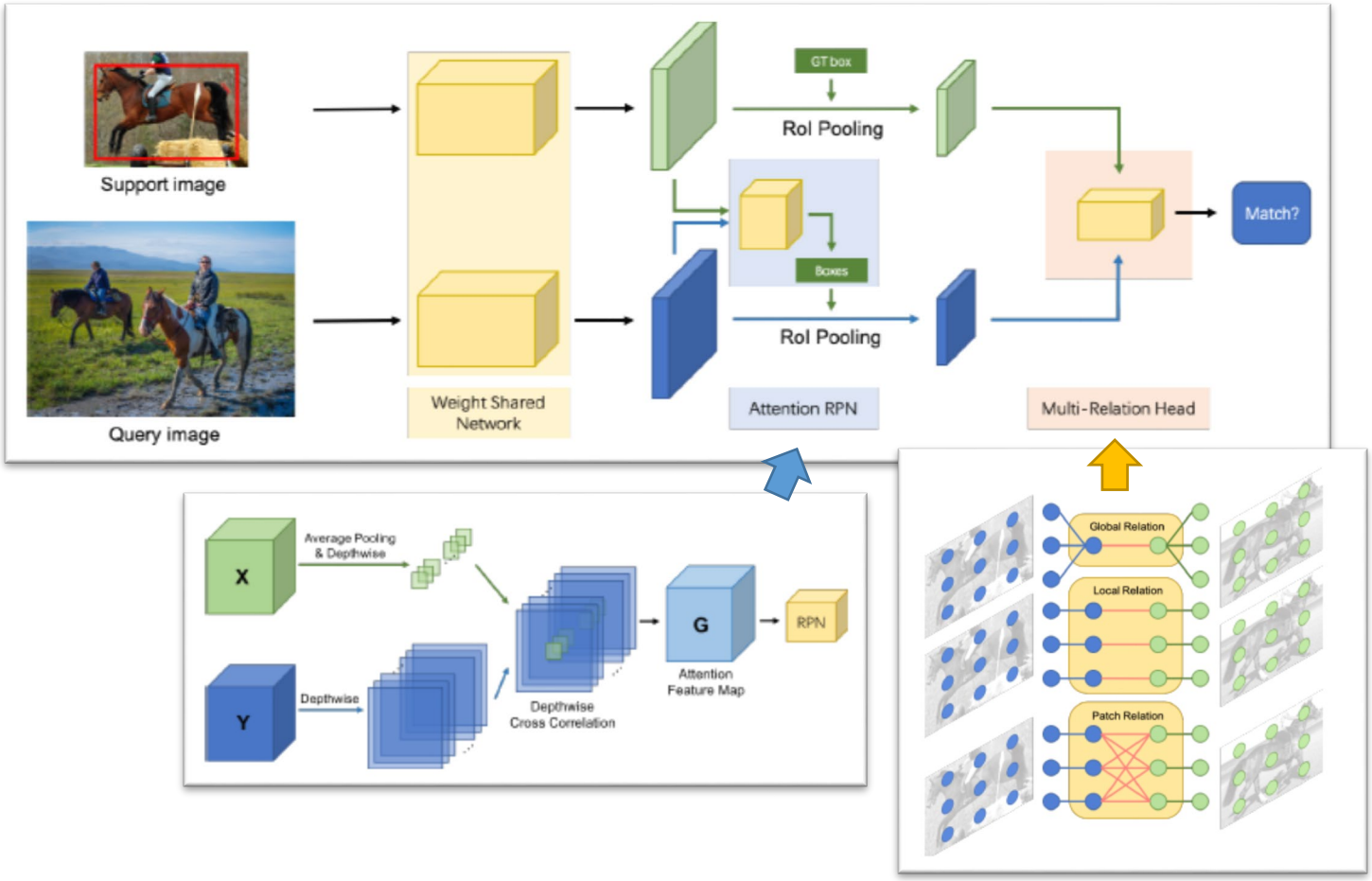
Support #1


Support #2

Example of 1-shot object detection

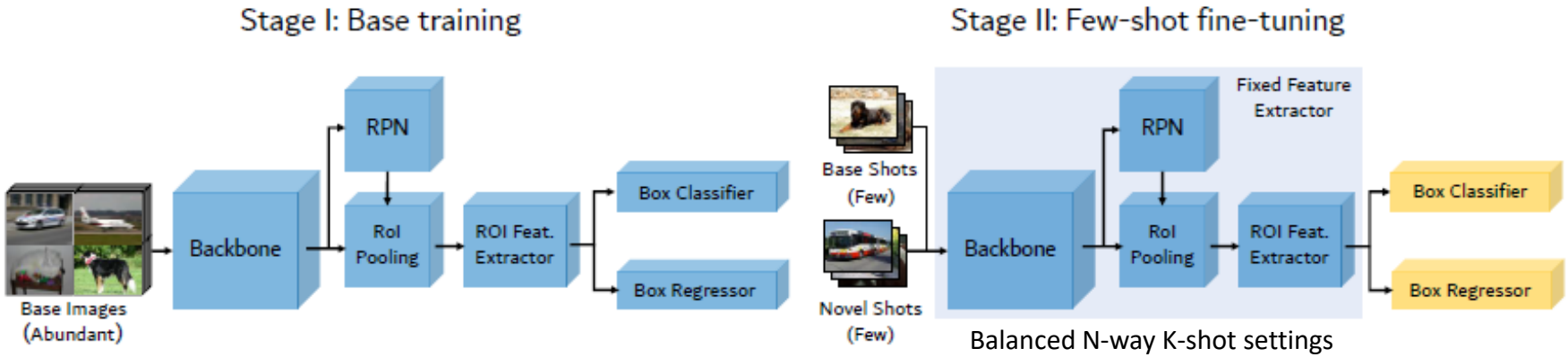
# Few-Shot Object Detection with Attention-RPN & Multi-Relation Detector [CVPR'20]

- Possible solution: meta-learning + object detection
- Network architecture (applicable for N-way K-shot setting)



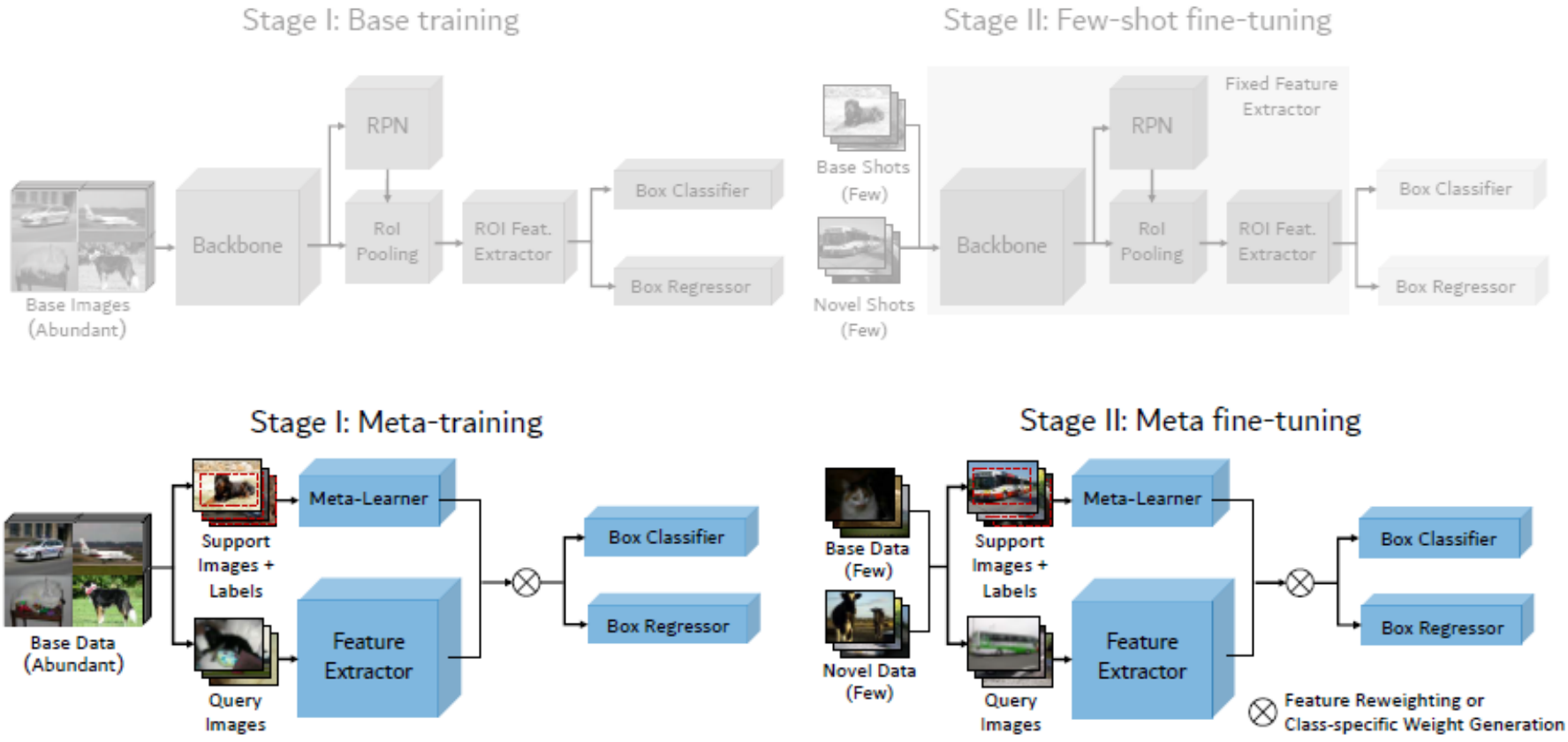
# **Frustratingly Simple Few-Shot Object Detection** [ICML'20]

- Possible solution: object detection + fine tuning ~~or meta-learning~~
- Network architecture



# **Frustratingly Simple Few-Shot Object Detection** [ICML'20]

- Possible solution: object detection + fine tuning + meta-learning (not so preferable)
- Network architecture

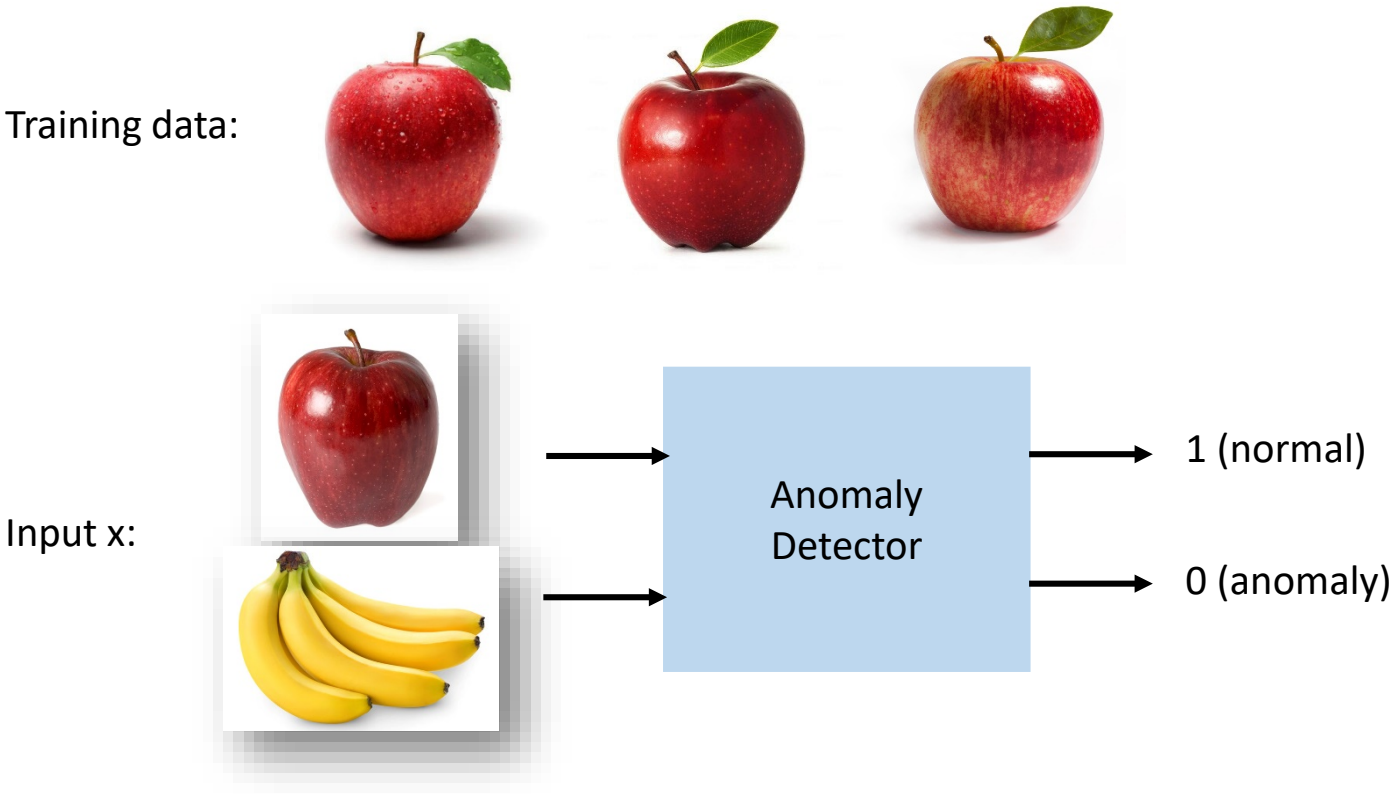


# What to Cover Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# (Unsupervised) Anomaly Detection

- Motivation
  - Given a set of training data (typically normal),
    find a function which determines whether the input *x* is normal or not.



# Applications of Anomaly Detection

- Video Surveillance
  - Determine abnormal events of interest

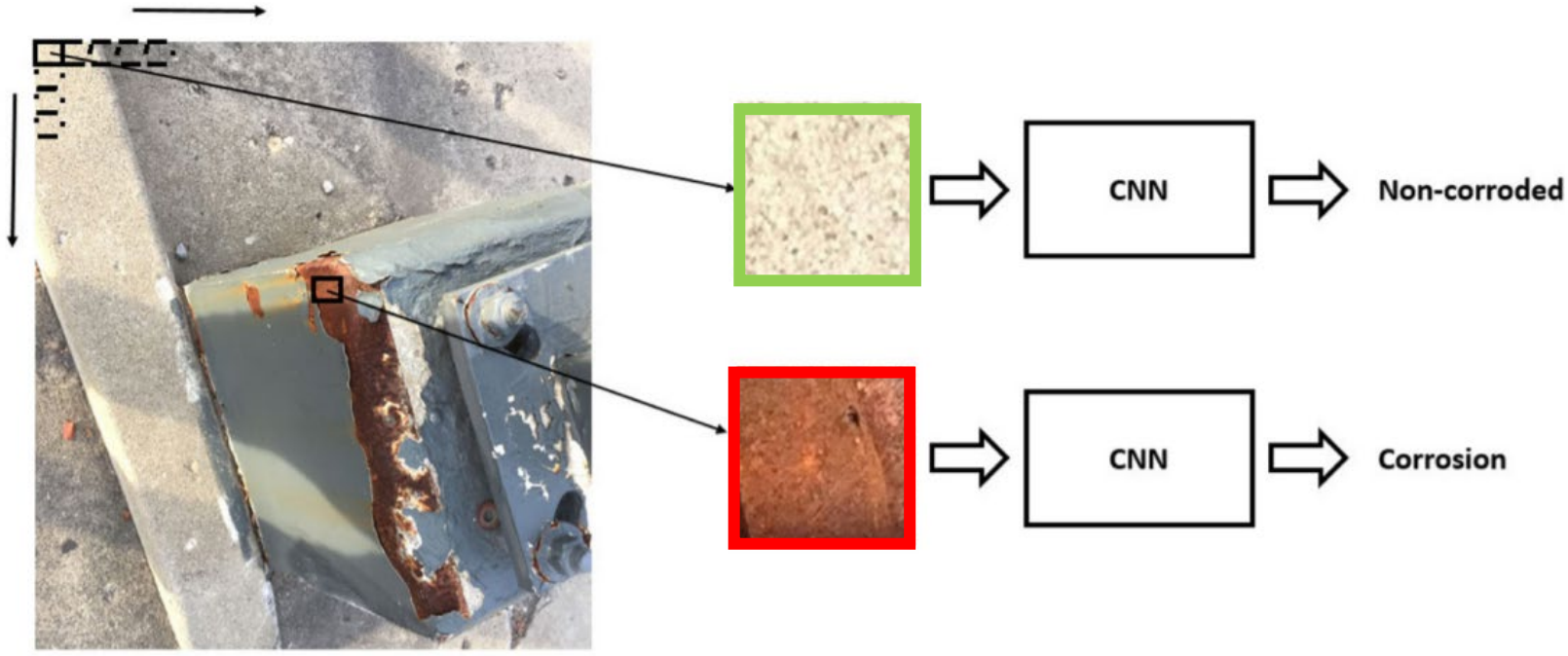
Normal frame

Abnormal frame

A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework (ICCV 2017)
Future Frame Prediction for Anomaly Detection -- A New Baseline (CVPR 2018)

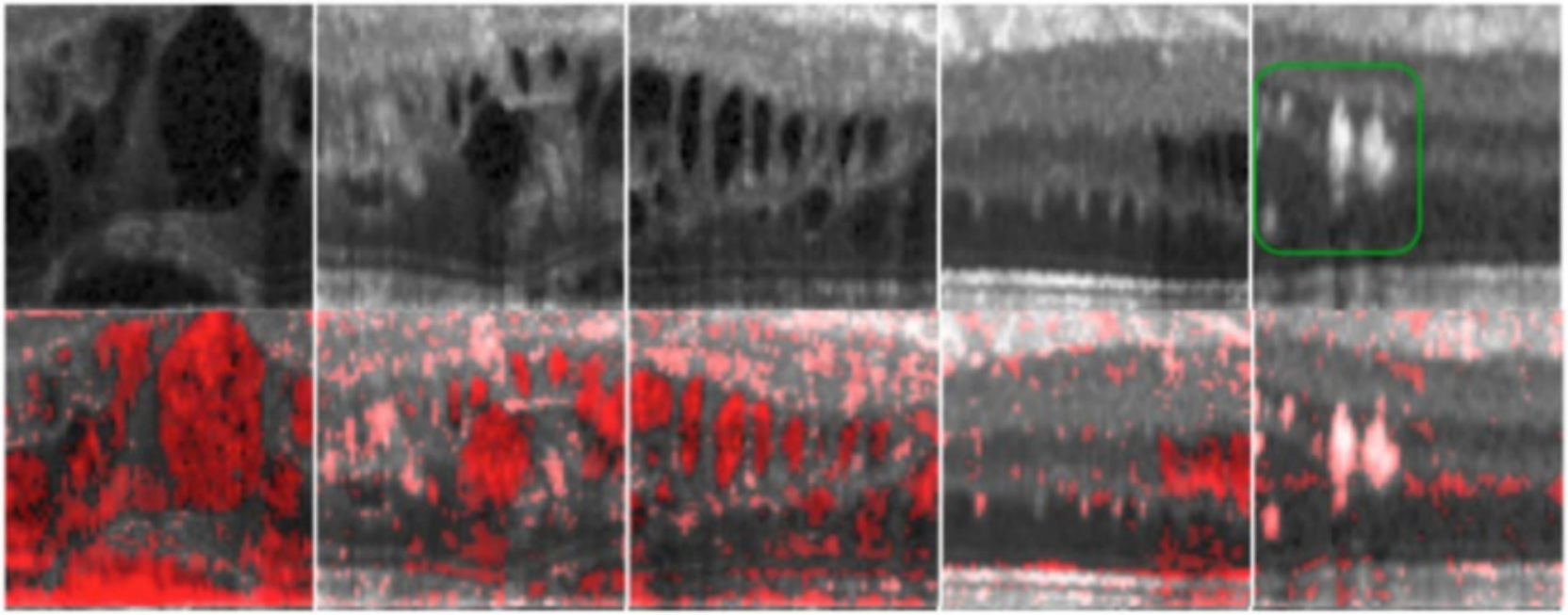
# Applications of Anomaly Detection

- Defect Detection
  - Find abnormal patterns/defects on or within particular objects or products
  - Can be applied to transportation safety as well (e.g., 斷軌, 異物入侵, etc.)



# Applications of Anomaly Detection

- Medical Image Analysis
  - Detect infected or damaged regions or organs
  - E.g., retinal damage, liver tumor, etc.



# Applications of Anomaly Detection

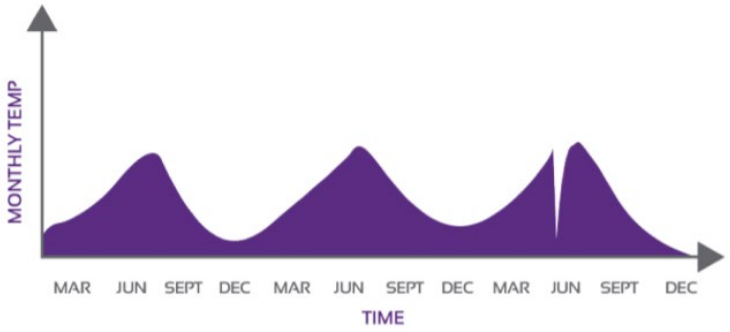
- Fraud Detection
  - Detect individual points which exhibit unusual behaviors or characteristics
  - If sequential data, can be applied to 機台數據分析與檢測, etc.
  - Will not address this type of problem in this lecture

| | | | | |
|---|---|---|---|---|
| May-22 | 1:14 pm | FOOD | Monaco Café | $1,127.80 |
| May-22 | 2:14 pm | WINE | Wine Bistro | $28.00 |
| ... | | | | |
| Jun-14 | 2:14 pm | MISC | Mobil Mart | $75.00 |
| Jun-14 | 2:05 pm | MISC | Mobil Mart | $75.00 |
| Jun-15 | 2:06 pm | MISC | Mobil Mart | $75.00 |
| Jun-15 | 11:49 pm | MISC | Mobil Mart | $75.00 |
| May-28 | 6:14 pm | WINE | Acton shop | $31.00 |
| May-29 | 8:39 pm | FOOD | Crossroads | $128.00 |
| Jun-16 | 11:14 am | MISC | Mobil Mart | $75.00 |
| Jun-16 | 11:49 am | MISC | Mobil Mart | $75.00 |

Point Anomaly

Collective Anomaly

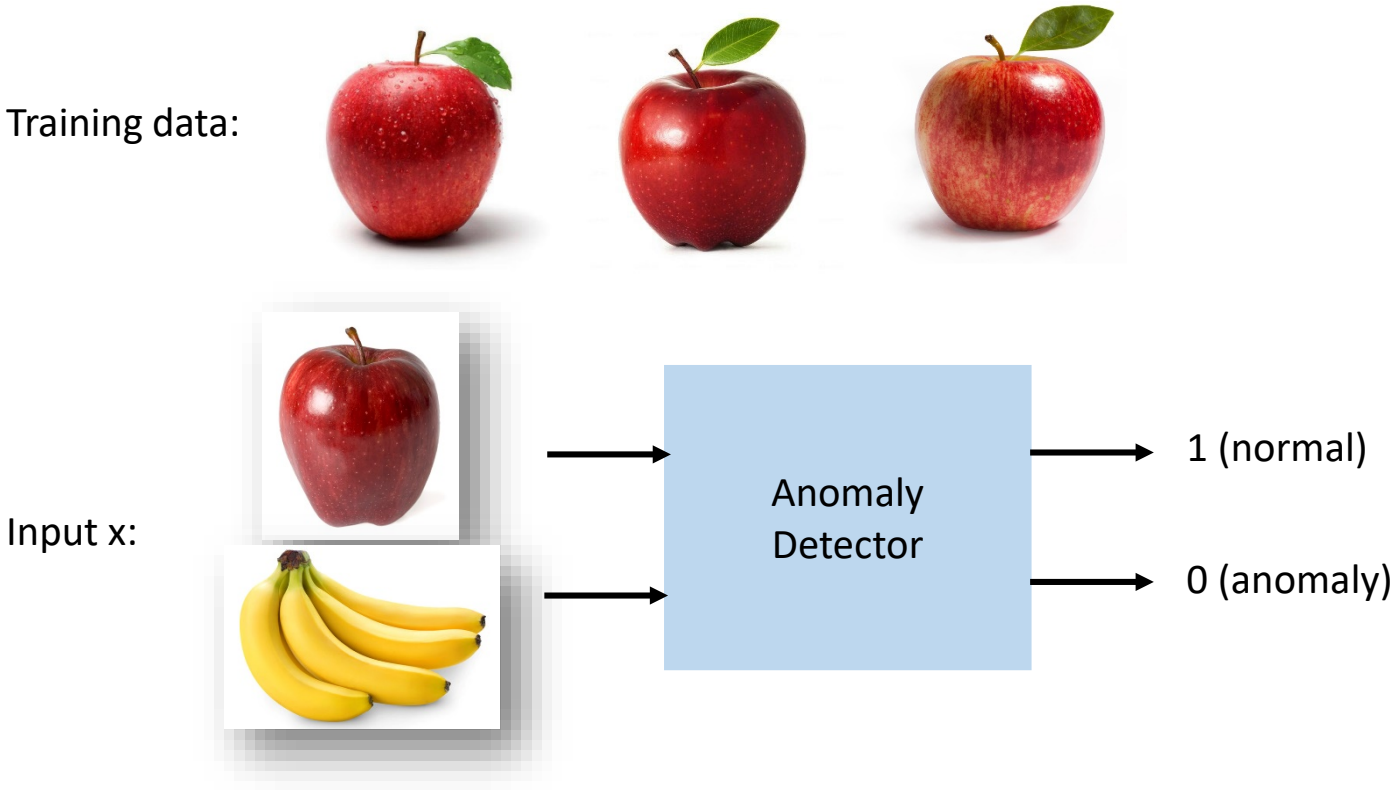
Credit card fraud detection




Temperature data

# Anomaly Detection (cont'd)

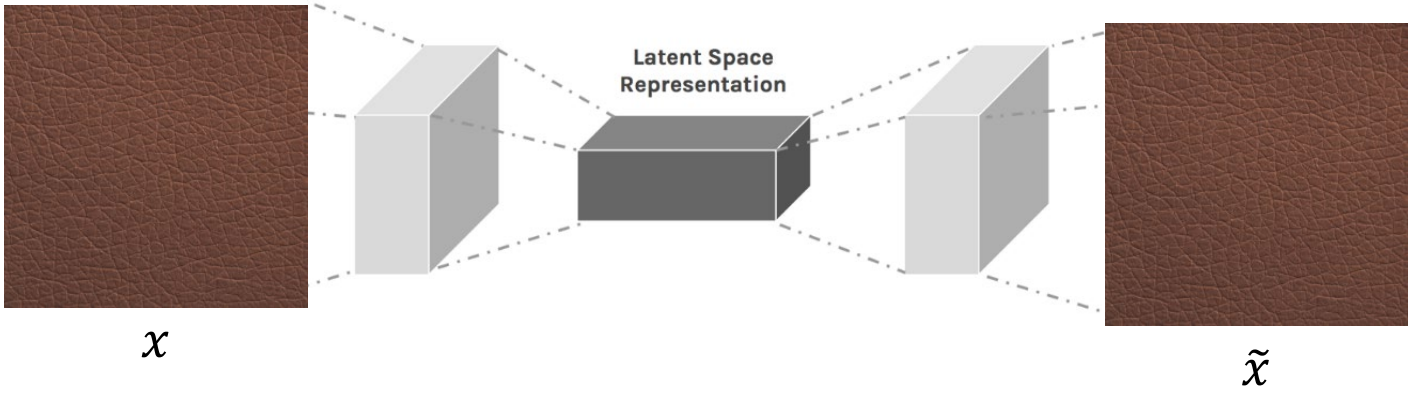
- Remarks
  - *Typically*, we do not train a classifier to solve this task, since one cannot possibly collect abnormal data in advance. → **Small Data Problem!**
  - Thus, **reconstruction-based DL models** are generally preferable.
  - **Unsupervised learning** for anomaly detection



# Autoencoder for Anomaly Detection

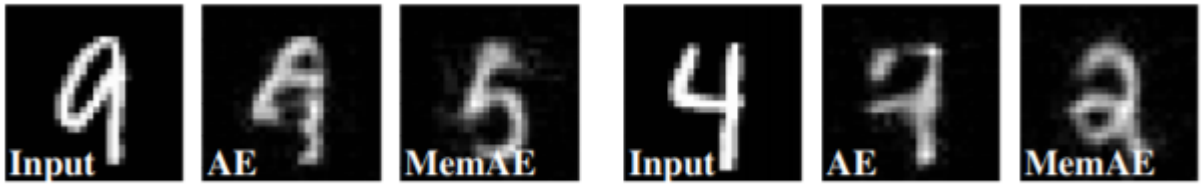
- Remarks
  - Learn an autoencoder which is trained to reconstruct only normal images
  - Use reconstruction loss $L_R$ as evaluation metric
    - Large $L_R$ → abnormal images
    - Small $L_R$ → normal images

$$L_R = |x - \tilde{x}|$$



# Autoencoder for Anomaly Detection (cont'd)

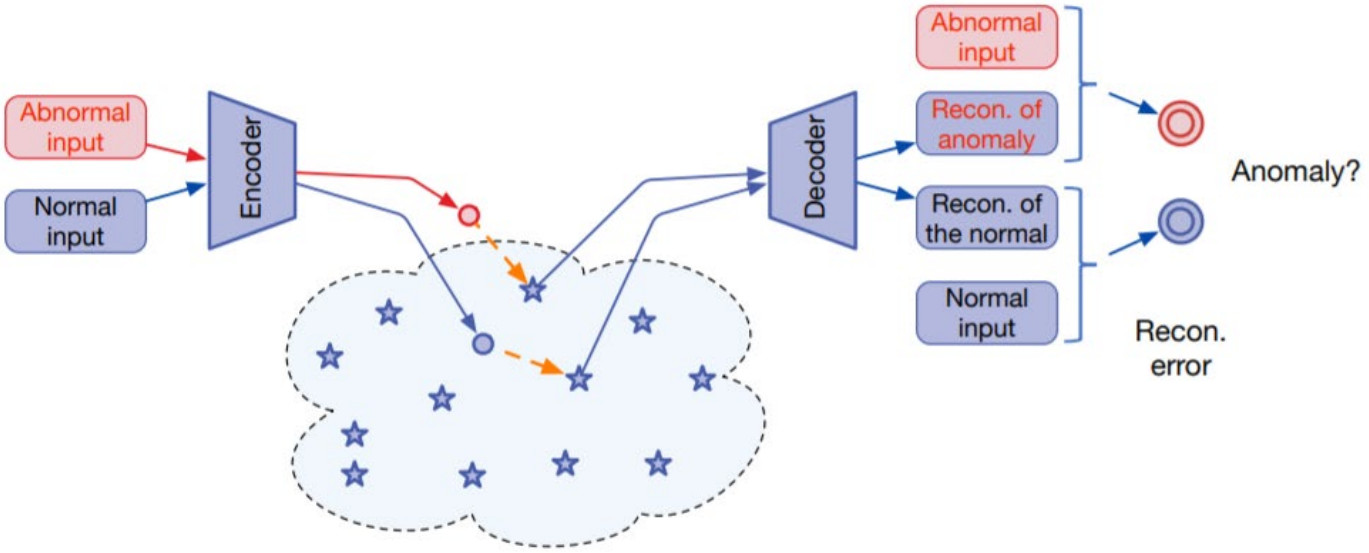
- Limitation
  - The learned autoencoder may sufficiently generalize, so that even the distorted/abnormal ones can also be recovered
  - Possible alternatives…
    - Regularization
    - Sparsity Constraints
    - Memory-augmented autoencoder




(a) Training on the normal "5" (b) Training on the normal "2"

# **Memorizing Normality to Detect Anomaly:** Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

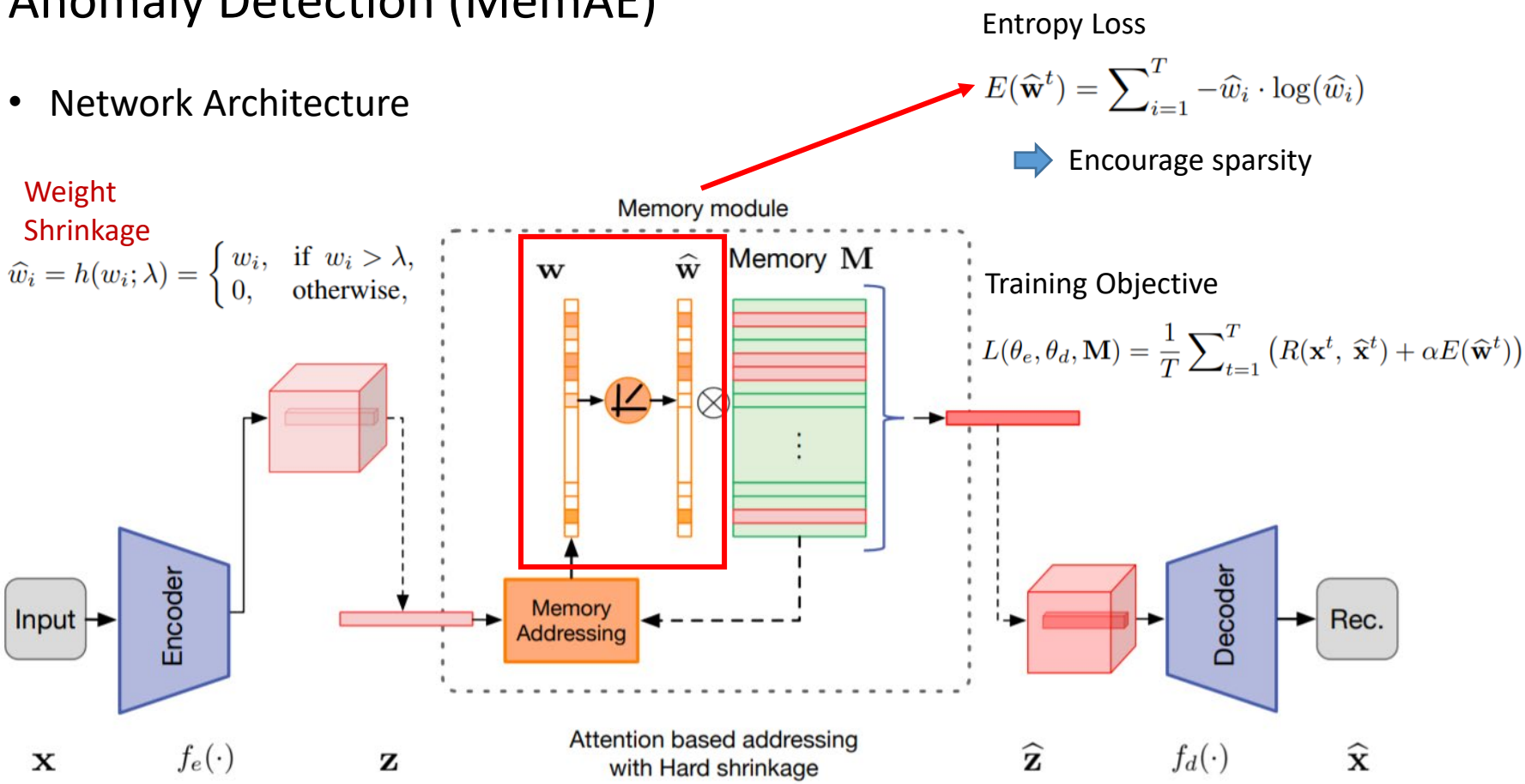
- Remarks
  - Store normal patterns in a memory bank
  - Reconstruct input instances using memory bank items only




Gong et al., Memorizing Normality to Detect Anomaly: Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE), ICCV 2019

# **Memorizing Normality to Detect Anomaly:** Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

- Network Architecture



# **Memorizing Normality to Detect Anomaly:** Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

- Network Architecture

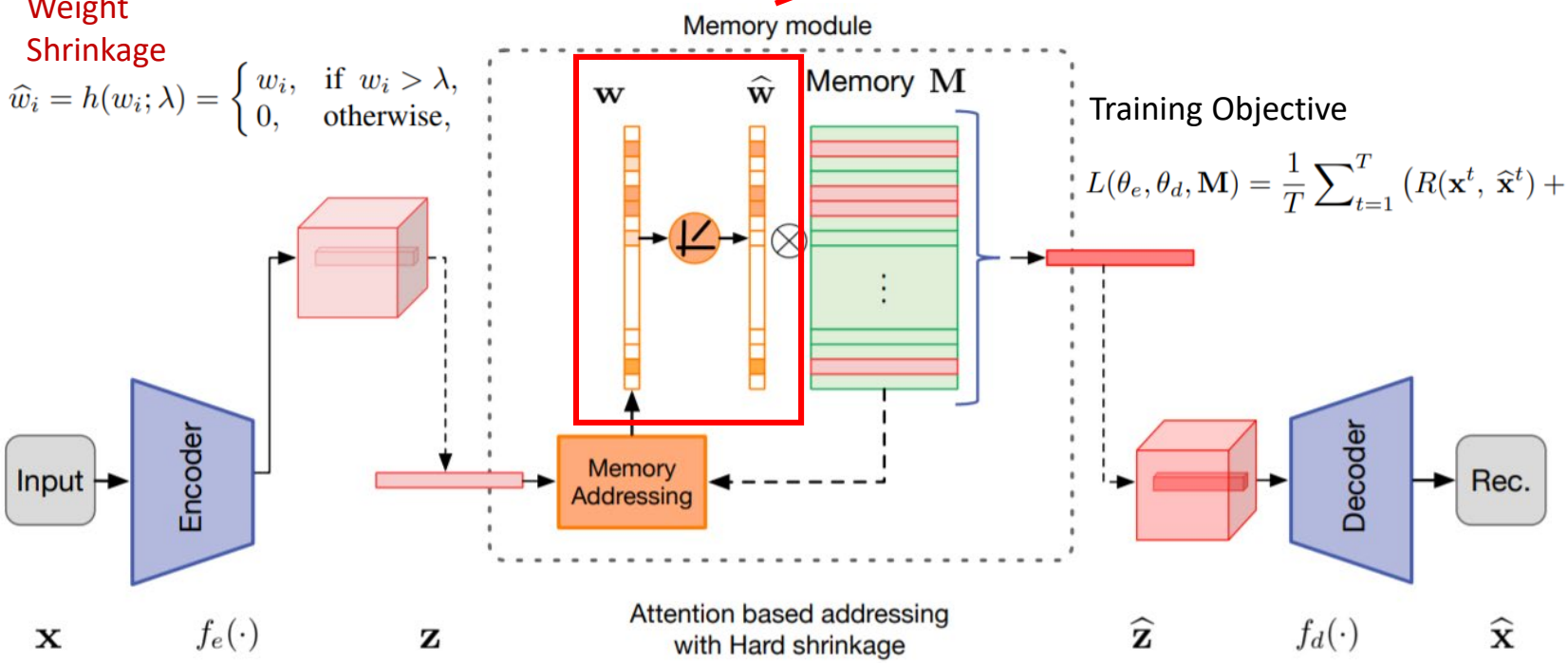
Entropy Loss

$$E(\hat{\mathbf{w}}^t) = \sum_{i=1}^{T} -\hat{w}_i \cdot \log(\hat{w}_i)$$

Encourage sparsity

Weight Shrinkage

$$\hat{w}_i = h(w_i; \lambda) = \begin{cases} w_i, & \text{if } w_i > \lambda, \\ 0, & \text{otherwise,} \end{cases}$$




Training Objective

$$L(\theta_e, \theta_d, \mathbf{M}) = \frac{1}{T} \sum_{t=1}^{T} \left( R(\mathbf{x}^t, \hat{\mathbf{x}}^t) + \alpha E(\hat{\mathbf{w}}^t) \right)$$

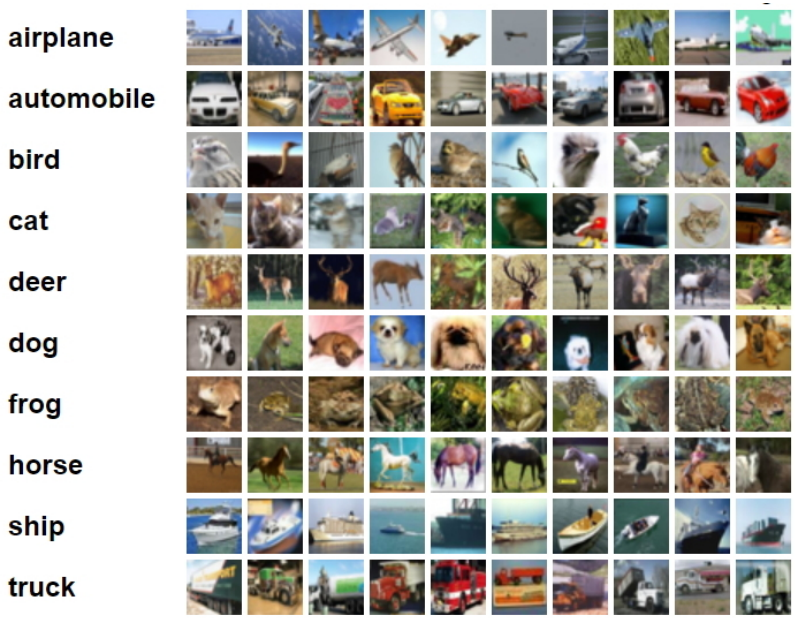
**Memorizing Normality to Detect Anomaly:**
Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

- Datasets

MNIST, Cifar10

CUHK Avenue

UCSD Ped2




**Memorizing Normality to Detect Anomaly:**
Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

- Encouraging sparsity (shrinkage/entropy) brings improvement.

| | Method\Dataset | UCSD-Ped2 | CUHK | SH.Tech |
|---|---|---|---|---|
| Non-Recon. | MPPCA [13] | 0.693 | - | - |
| | MPPCA+SFA [25] | 0.613 | - | - |
| | MDT [25] | 0.829 | - | - |
| | AMDN [40] | 0.908 | - | - |
| | Unmasking [36] | 0.822 | 0.806 | - |
| | MT-FRCN [10] | 0.922 | - | - |
| | Frame-Pred [24] | **0.954** | **0.849** | **0.728** |
| Recon. | AE-Conv2D [9] | 0.850 | 0.800 | 0.609 |
| | AE-Conv3D [44] | 0.912 | 0.771 | - |
| | TSC [24] | 0.910 | 0.806 | 0.679 |
| | StackRNN [24] | 0.922 | 0.817 | 0.680 |
| | AE | 0.917 | 0.810 | 0.697 |
| | MemAE-nonSpar | 0.929 | 0.821 | 0.688 |
| | MemAE | **0.941** | **0.833** | **0.712** |

Table 4. Ablation studies based on UCSD-Ped2 dataset.

| Method | AUC |
|---|---|
| AE | 0.9170 |
| AE-$\ell_1$ | 0.9286 |
| MemAE-nonSpar | 0.9293 |
| MemAE w/o Shrinkage | 0.9324 |
| MemAE w/o Entropy loss | 0.9372 |
| MemAE | **0.9410** |

**Memorizing Normality to Detect Anomaly:**
Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection (MemAE)

- Memory Size
  - Given a sufficient amount of memory size,
    MemAE is able to robustly produce plausible results.

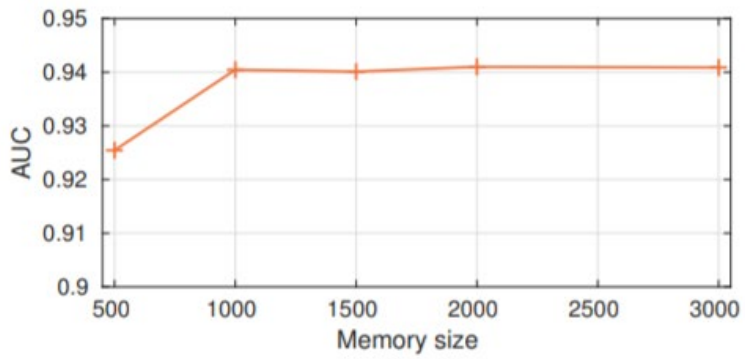
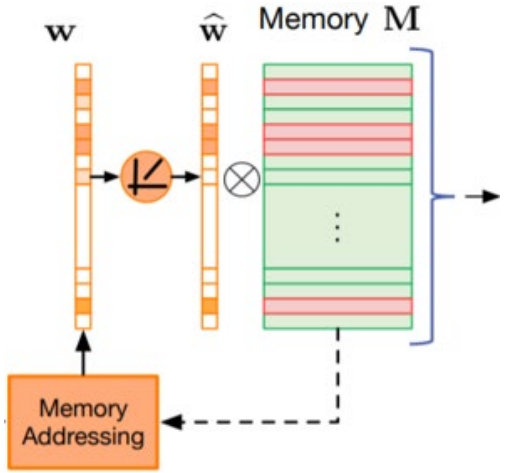



Figure 7. Robustness to the setting of memory size. AUC values of MemAE with different memory size on UCSD-Ped2 are shown.

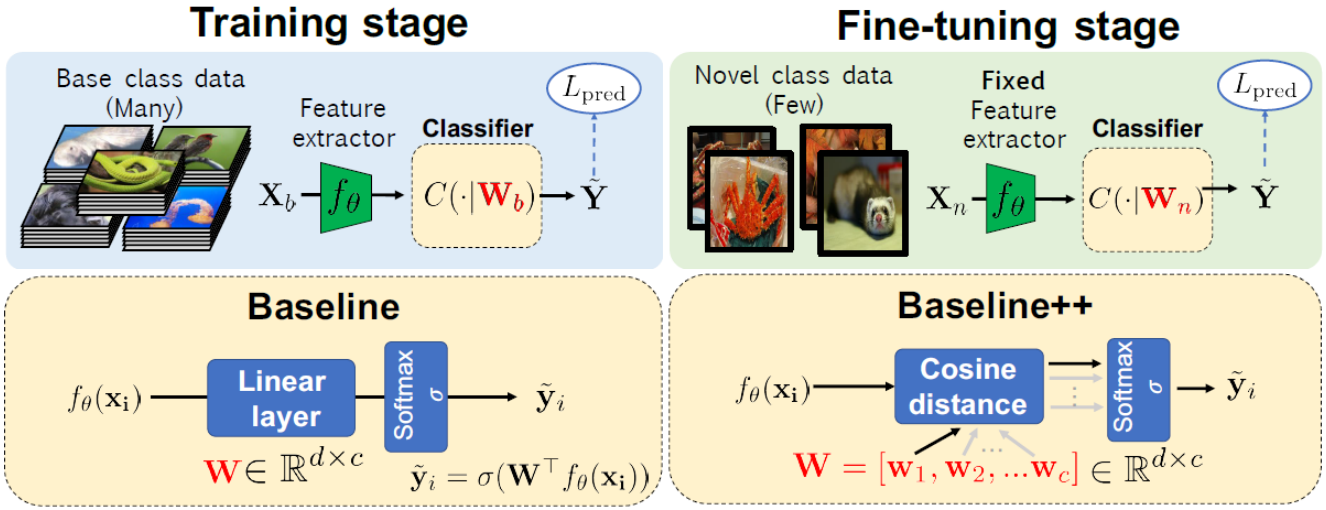


# What to Cover Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# A Closer Look at FSL (1/3)

- Observation/Motivation
  - **Deeper backbones** significantly reduce the gap across existing FSL methods. (with decreased domain shifts between base and novel classes)
  - A slightly modified baseline method (**baseline++**) surprisingly achieves competitive performance.
  - Simple baselines (**baseline** and **baseline++**: trained on base and fine-tuned on novel) outperform representative FSL methods when the domain shift grows larger.



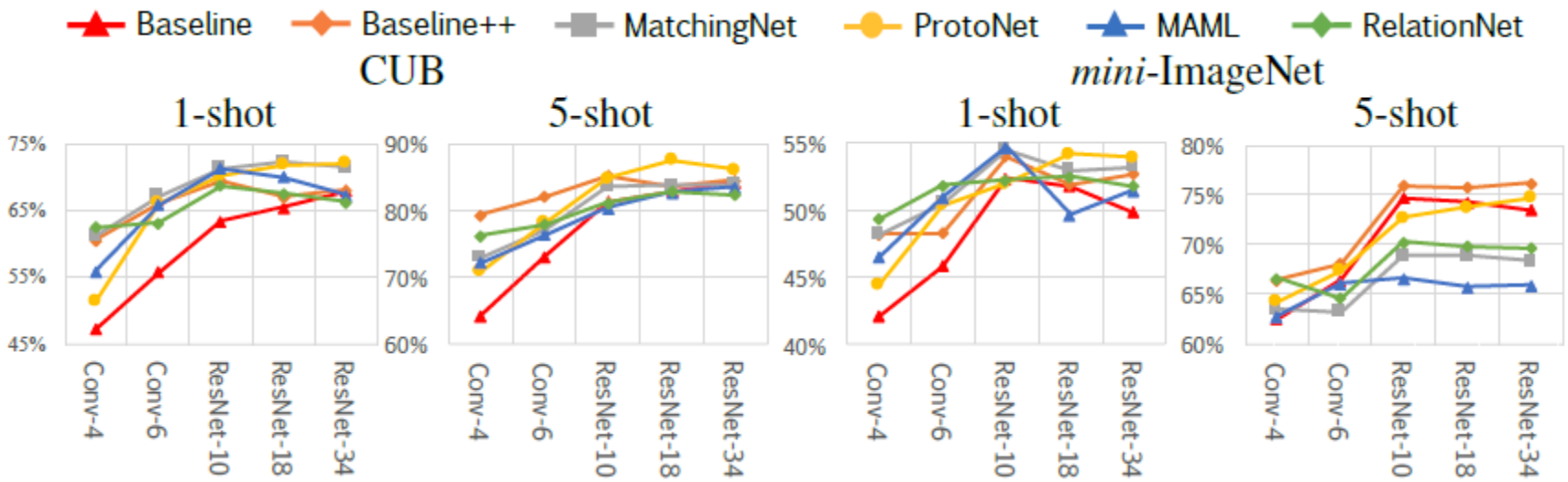

use **cosine distances** between the input feature and the weight vector for each class to reduce intra-class variations



Chen et al., A Closer Look at Few-shot Classification, *ICLR*, 2019

# A Closer Look at FSL (2/3)

- Performance with deeper backbones
  - For CUB, gaps among different methods diminish as the backbone gets deeper.
  - For mini-ImageNet, some meta-learning methods are even beaten by baselines with a deeper backbone.

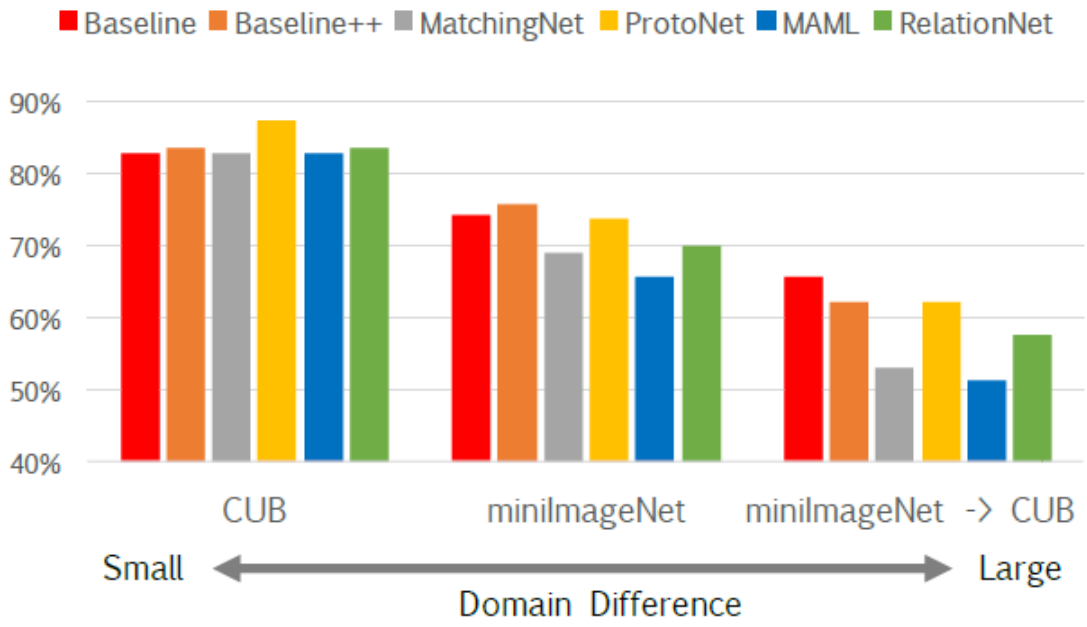





Chen et al., A Closer Look at Few-shot Classification, *ICLR*, 2019

# A Closer Look at FSL (3/3)

- Performance with domain shifts (using ResNet-18)
  - Existing FSL methods fail to address large domain shifts (e.g., mini-ImageNet → CUB) and are inferior to the baseline methods.
  - This highlights the importance of learning to adapt to domain differences in FSL.

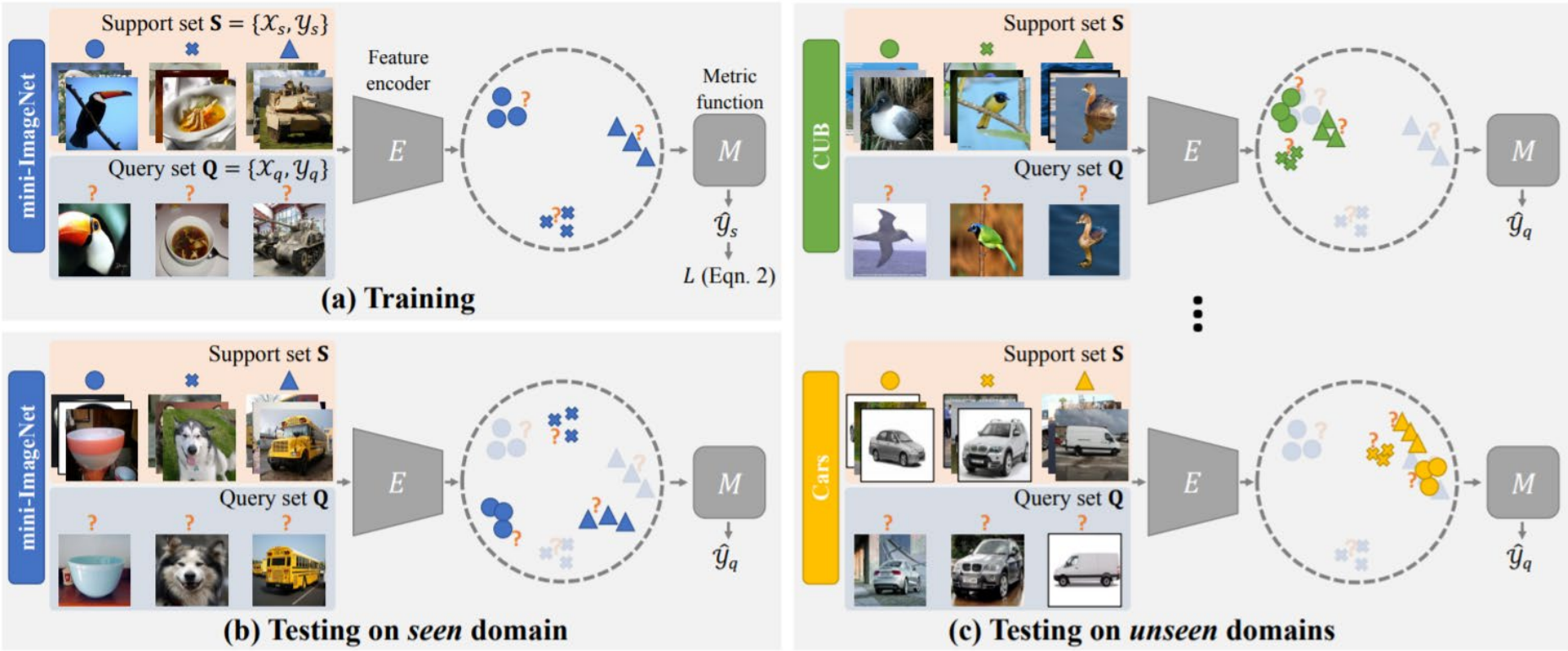





Chen et al., A Closer Look at Few-shot Classification, *ICLR*, 2019

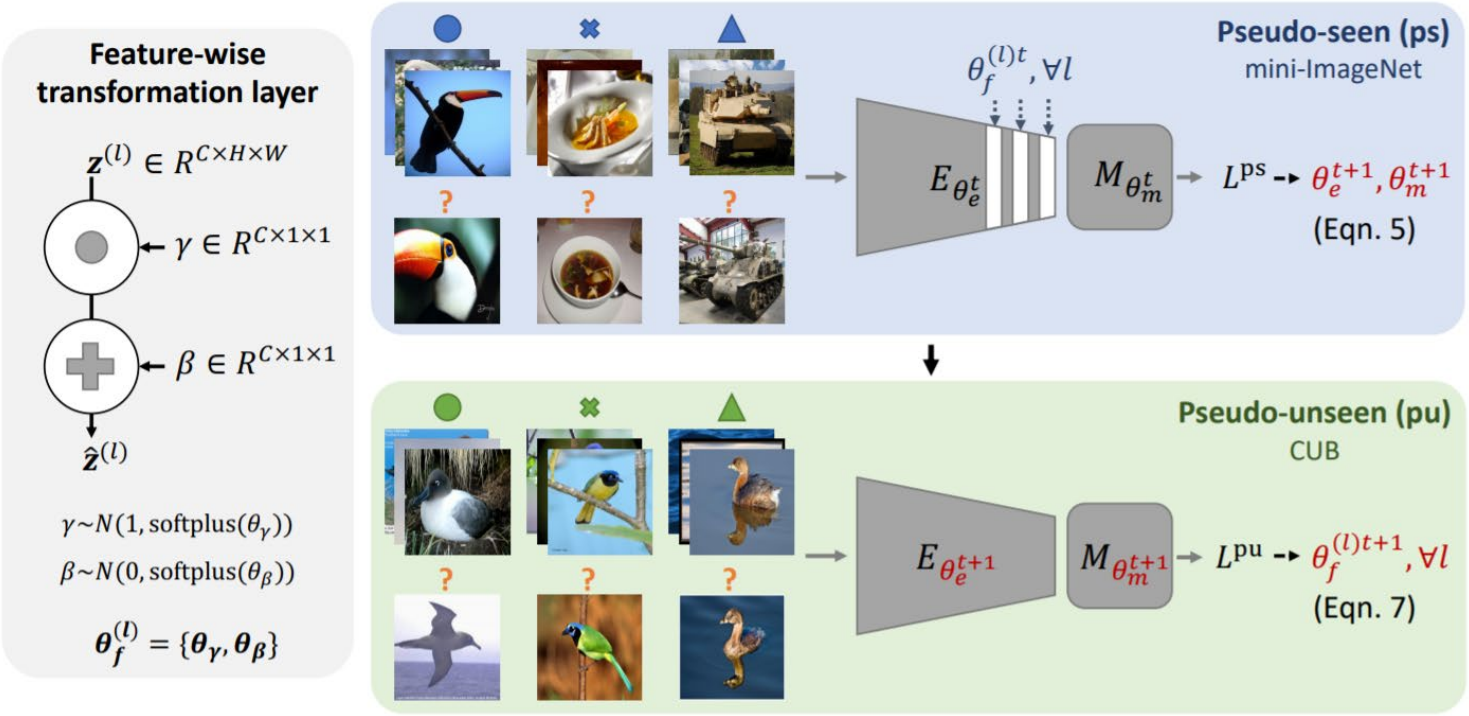
# Cross-Domain (or Domain Generalized) FSL

- Tackle N-way K-shot problems on **unseen domains**
- Novel classes (few-shot) and domains (cross-domain)



# Cross-Domain FSL

- **Cross-domain Few-shot Classification via Learned Feature-wise Transformation**
- Feature-wise transformation layer for feature level **augmentation**
- **Meta-learning** is applied.




Tseng, H. Y., Lee, H. Y., Huang, J. B., & Yang, M. H.: Cross-domain few-shot classification via learned feature-wise transformation. In ICLR (2020)

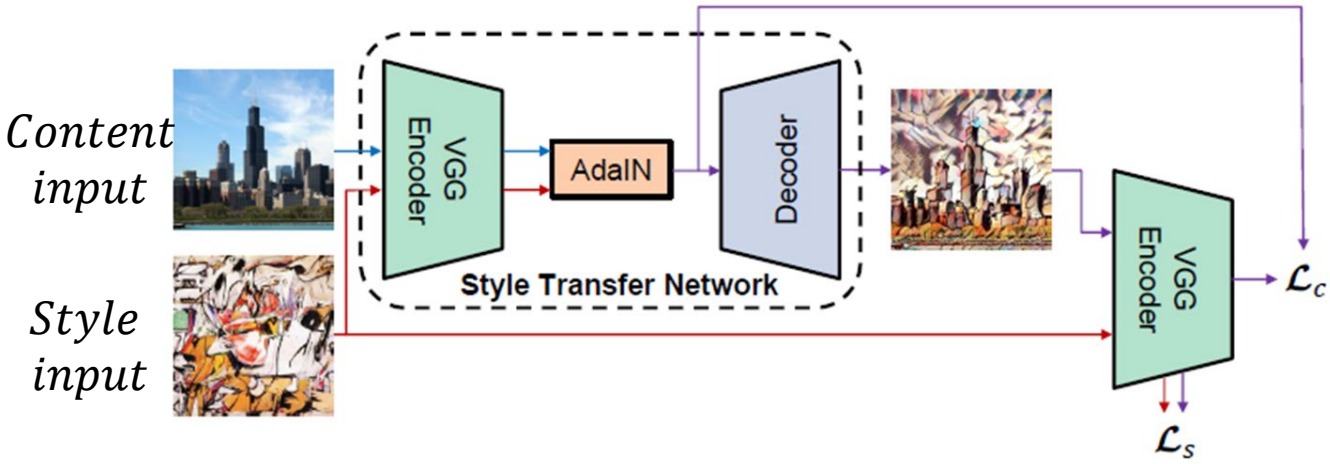
# Recap: AdaIN

- We've talked about style transfer methods like Pix2Pix or CycleGAN…
- Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization (ICCV'17)
  - Adaptive instance normalization (AdaIN) for arbitrary style transfer in real-time



Huang *et al.* " Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization . " *ICCV* 2017

# AdaIN




- Adaptive Instance Normalization

$$\text{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

- $x$: content input, $y$: style input
- No learnable affine parameters
- Perform style transfer in the feature space

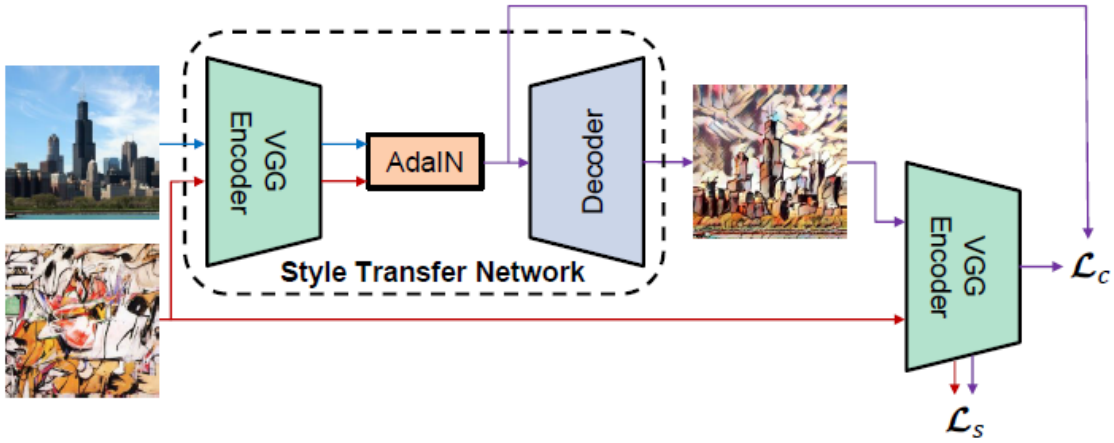


Huang *et al.* " Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization . " *ICCV* 2017

# AdaIN

- $f$: Encoder, $g$: Decoder

$$t = \text{AdaIN}(f(c), f(s))$$




$$\mathcal{L}_c = \|f(g(t)) - t\|_2$$ *Content loss: content consistency*

$$\mathcal{L}_s = \sum_{i=1}^{L} \|\mu(\phi_i(g(t))) - \mu(\phi_i(s))\|_2 + \sum_{i=1}^{L} \|\sigma(\phi_i(g(t))) - \sigma(\phi_i(s))\|_2$$

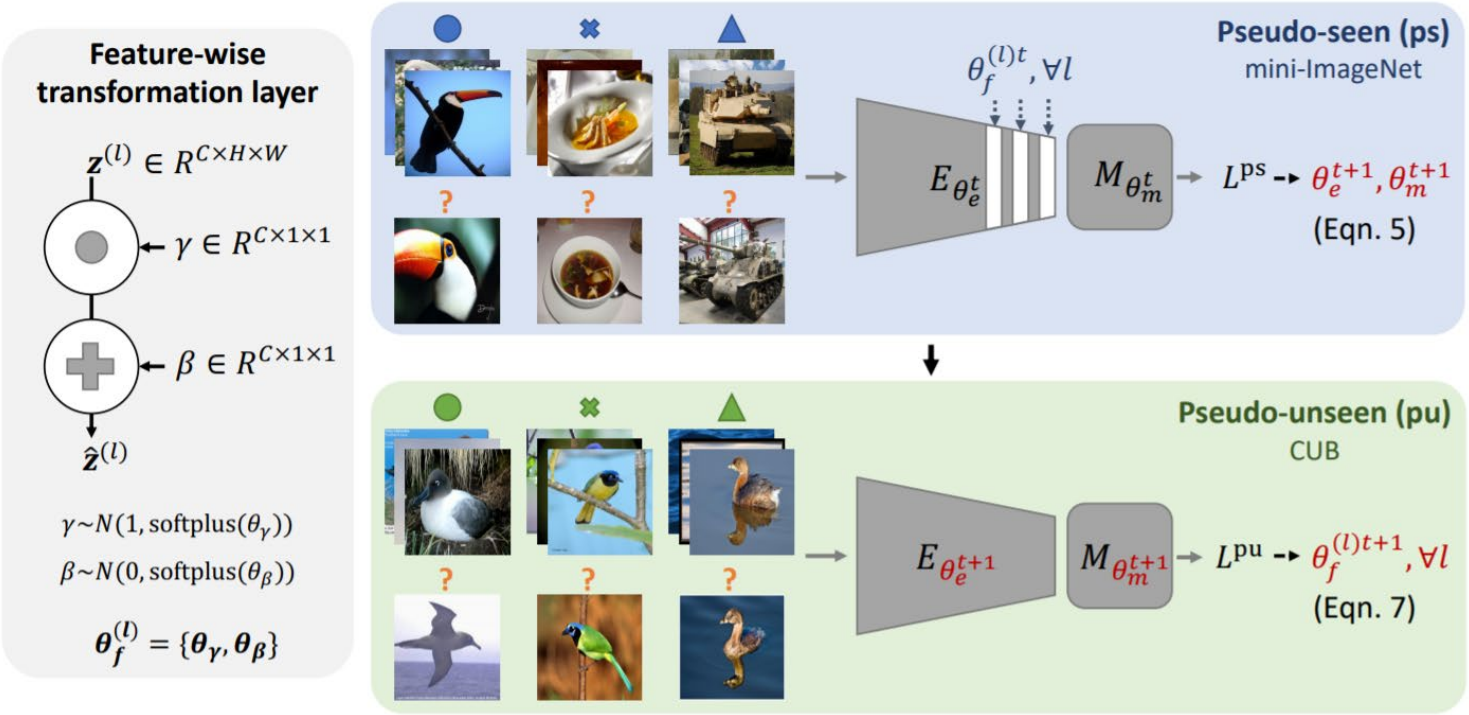
*Style loss: Gram matrix loss*
*($\emptyset_i$ denotes a layer in $VGG-19$ used to compute style loss)*



Huang *et al.* "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization ." *ICCV* 2017

# Cross-Domain FSL

- Cross-domain Few-shot Classification via Learned Feature-wise Transformation
- Feature-wise transformation layer for feature level **augmentation**
- **Meta-learning** is applied.



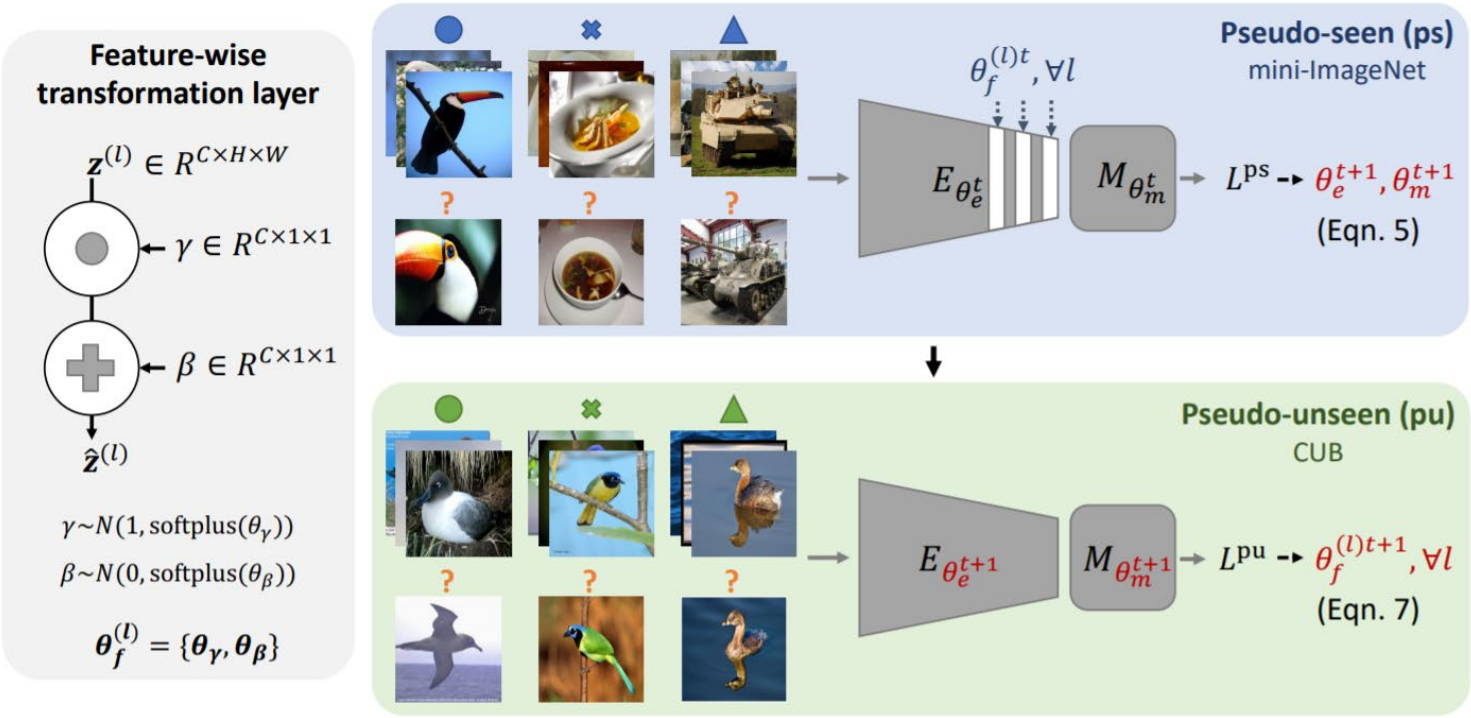

Tseng, H. Y., Lee, H. Y., Huang, J. B., & Yang, M. H.: Cross-domain few-shot classification via learned feature-wise transformation. In ICLR (2020)

# Cross-Domain FSL

- **Alternately** update
  - backbone model (E, M)
  - transformation layer

**// Update metric-based model with pseudo-seen task:**
Obtain $\theta_e^{t+1}, \theta_m^{t+1}$ using equation 5

**// Update feature-wise transformation layers with pseudo-unseen task:**
Obtain $\theta_f^{t+1}$ using equation 6 and equation 7




Tseng, H. Y., Lee, H. Y., Huang, J. B., & Yang, M. H.: Cross-domain few-shot classification via learned feature-wise transformation. In ICLR (2020)

# Cross-Domain FSL

- Experiments
  - FT: hand-tuned feature transformer ($\gamma$ and $\beta$)
  - LFT: learned feature transformer

| 5-way 5-Shot | | CUB | Cars | Places | Plantae |
|---|---|---|---|---|---|
| MatchingNet | - | $51.92 \pm 0.80\%$ | $39.87 \pm 0.51\%$ | $61.82 \pm 0.57\%$ | $47.29 \pm 0.51\%$ |
| | FT | $56.29 \pm 0.80\%$ | $39.58 \pm 0.54\%$ | $62.32 \pm 0.58\%$ | $46.48 \pm 0.52\%$ |
| | LFT | $\mathbf{61.41 \pm 0.57\%}$ | $\mathbf{43.08 \pm 0.55\%}$ | $\mathbf{64.99 \pm 0.59\%}$ | $\mathbf{48.32 \pm 0.57\%}$ |
| RelationNet | - | $62.13 \pm 0.74\%$ | $40.64 \pm 0.54\%$ | $64.34 \pm 0.57\%$ | $46.29 \pm 0.56\%$ |
| | FT | $63.64 \pm 0.77\%$ | $42.24 \pm 0.57\%$ | $65.42 \pm 0.58\%$ | $47.81 \pm 0.51\%$ |
| | LFT | $\mathbf{64.99 \pm 0.54\%}$ | $\mathbf{43.44 \pm 0.59\%}$ | $\mathbf{67.35 \pm 0.54\%}$ | $\mathbf{50.39 \pm 0.52\%}$ |
| GNN | - | $69.26 \pm 0.68\%$ | $48.91 \pm 0.67\%$ | $72.59 \pm 0.67\%$ | $58.36 \pm 0.68\%$ |
| | FT | $70.37 \pm 0.68\%$ | $47.68 \pm 0.63\%$ | $74.48 \pm 0.70\%$ | $57.85 \pm 0.68\%$ |
| | LFT | $\mathbf{73.11 \pm 0.68\%}$ | $\mathbf{49.88 \pm 0.67\%}$ | $\mathbf{77.05 \pm 0.65\%}$ | $\mathbf{58.84 \pm 0.66\%}$ |

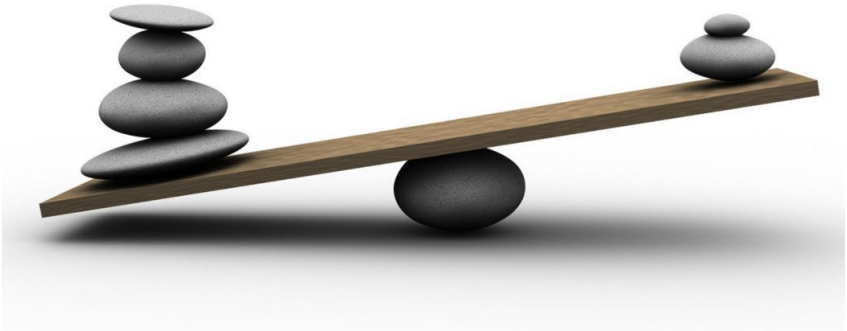
Tseng, H. Y., Lee, H. Y., Huang, J. B., & Yang, M. H.: Cross-domain few-shot classification via learned feature-wise transformation. In ICLR (2020)

# What to Cover Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL (w/ Q&A)
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
  - Discussions

# Challenges & Opportunities in Small-Data Problems

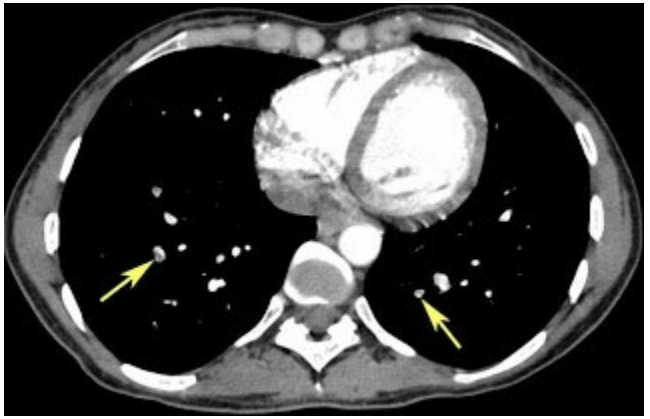
- Imbalanced Data Learning
  - Some categories with a sufficient # of data, while others are not → **Small Data Problem!**
  - E.g., medical image analysis, defect detection, etc.
- Possible Solutions
  - Reweighting instances, loss functions accordingly
  - Data augmentation
  - Data hallucination
  - However, augmenting/hallucinating data requires **domain knowledge**!



# Challenges & Opportunities in Small-Data Problems

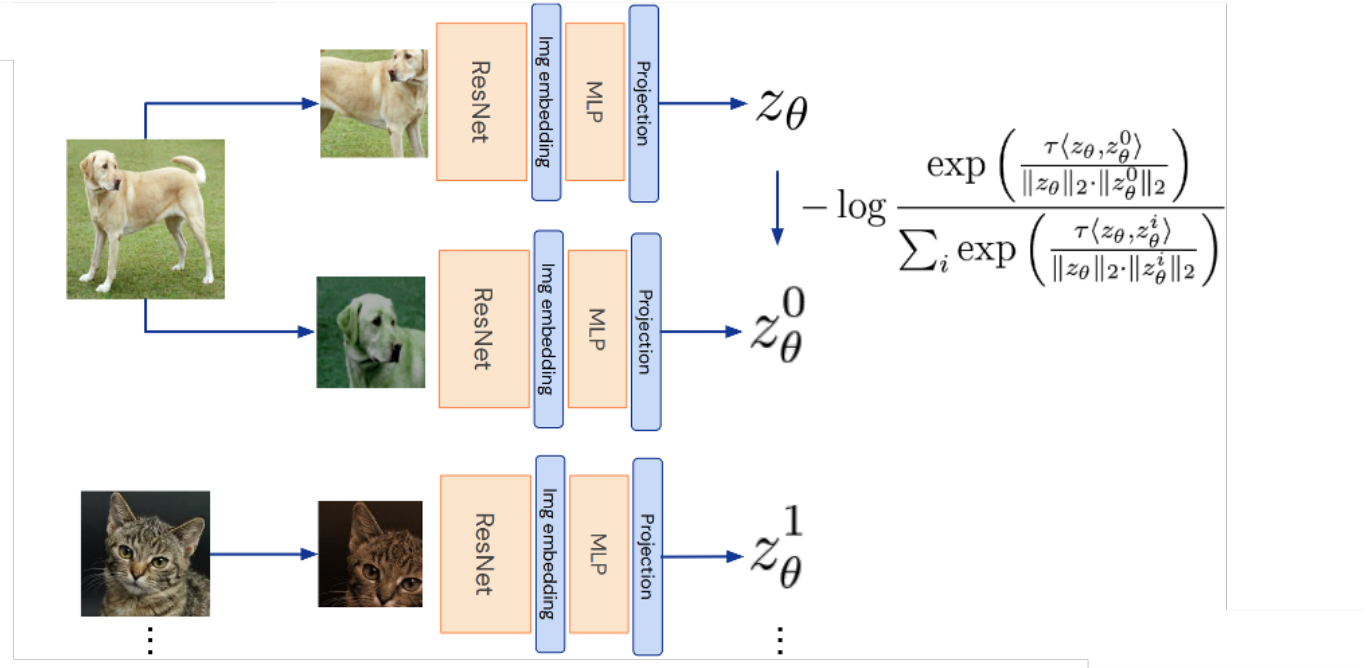
- Learning with Partial Supervision
  - No pixel-level ground truth but only image-level labels available → **Small Data Problem!**
  - E.g., medical image detection or segmentation
- Possible Solutions
  - Active Learning (human-in-the-loop)
  - Semi-Supervised Learning (at least collect few images with pixel-wise labels)
  - Weakly Supervised Learning (e.g., multiple instance learning)
  - Can be guided with auxiliary info (e.g., location or number of objects in an image)


Tumor Yes/No?



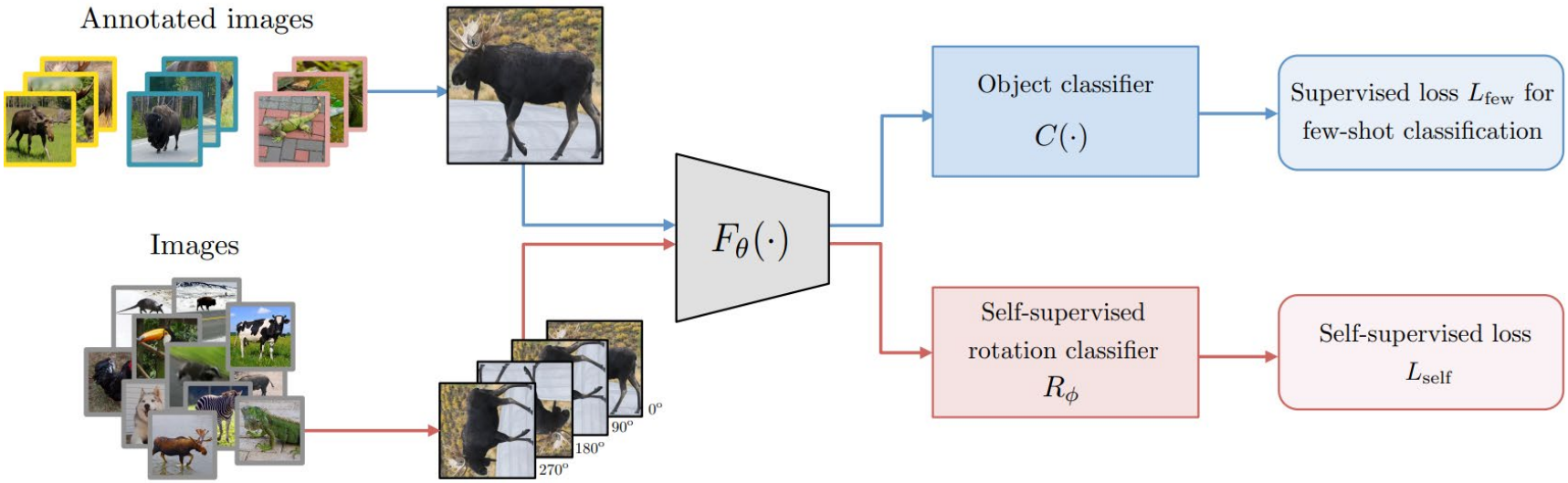
# More Opportunities in Small-Data Problems

- Self-Supervised Learning (SSL)
  - A **properly trained network backbone** is the King!
    (recall examples in transfer learning, domain adaptation or few-shot learning)
  - Typically implemented in an unsupervised manner (e.g., via contrastive learning)
  - SimCLR [ICML'20]
    - Use augmented images as positive pairs, and other images as negative pairs



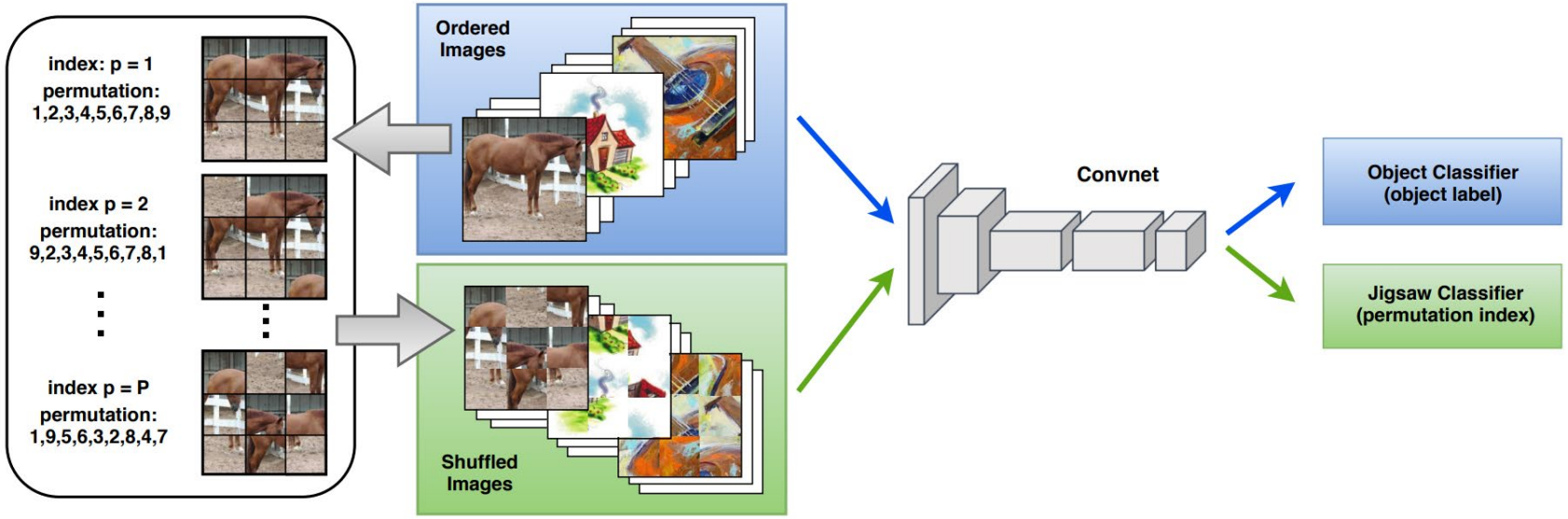
# Challenges & Opportunities in Small-Data Problems

- SSL in FSL
  - Gidaris *et al. "* Boosting Few-Shot Visual Learning with Self-Supervision. *" ICCV* 2019
  - Use the *rotation prediction* as self-supervision
  - Multi-task learning
    - Supervised loss for few-shot learning
    - ***Self-supervised loss (rotation prediction)***
  - Self-supervision as an auxiliary task in FSL, which enables feature extractor *F* to learn richer and more transferable visual features while still using few annotated samples.



# Challenges & Opportunities in Small-Data Problems

- SSL in FSL
  - Carlucci *et al. ”* Domain Generalization by Solving Jigsaw Puzzles. *” CVPR* 2019
  - Use the *jigsaw puzzles* as self-supervision
  - Multi-task learning
    - Supervised loss for few-shot learning
    - ***Self-supervised loss (jigsaw puzzle order prediction)***



# What We've Covered Today…
# Few-Shot Learning & Its Applications

- (A Brief Review of) Meta-Learning
  - Definition
  - Parametric & Non-Parametric based Approaches
- Few-Shot Learning (via Meta-Learning)
  - Few-Shot Classification
    - Metric Learning vs. Data Hallucination
  - Few-Shot Image Segmentation
  - Few-Shot Object Detection
- Applications & Challenges in FSL
  - Anomaly/Defect Detection
  - Domain Bias in FSL
  - Imbalanced, Weakly Supervised, Self-Supervised Learning in FSL
- Q&A