One-shot Learning(2017-2018)

One-shot Learning (2017-2018)

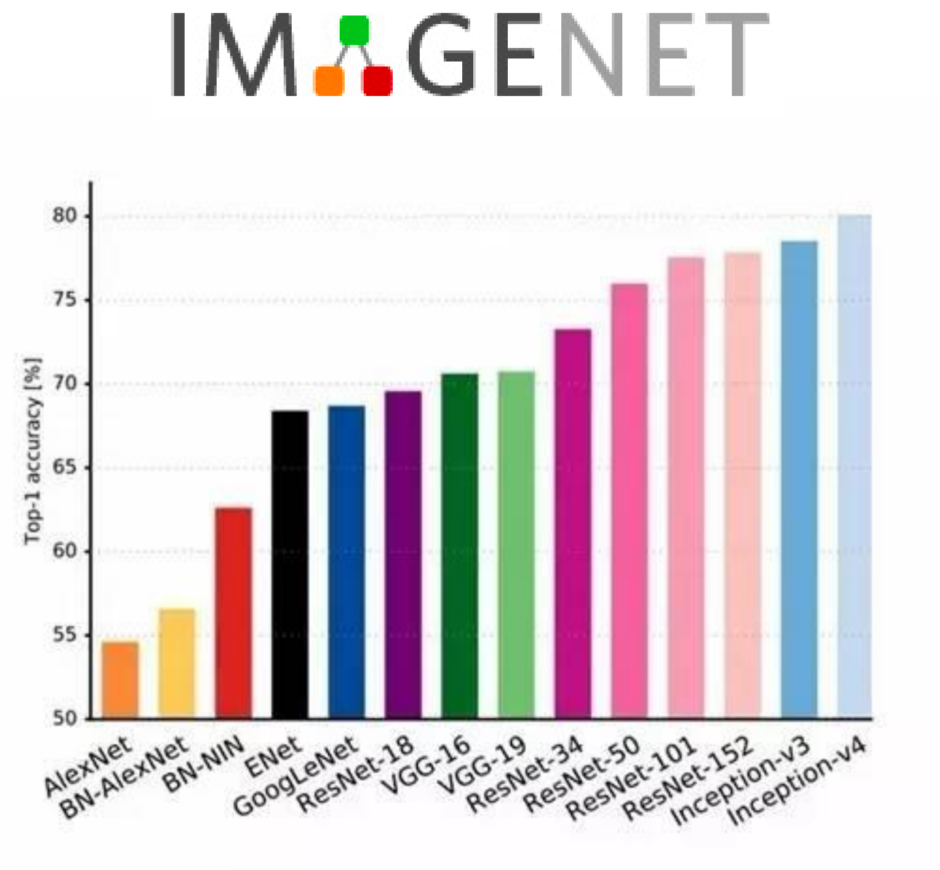
- Introduction&Motivation
- Survey of One-shot Learning
- Transfer Learning-based Approaches
- Data Augmentation Approaches
- Beyond One-shot Learning

付彦伟（复旦大学）

Introduction

0

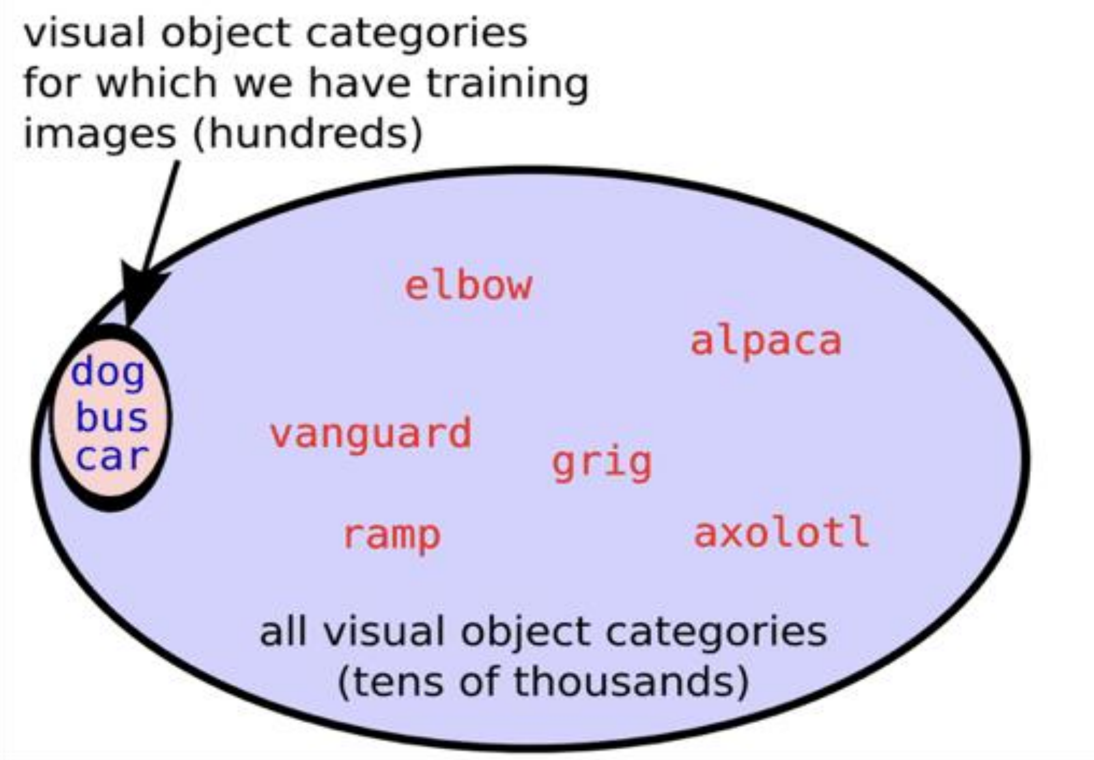
Success of the Large-scale Learning



Face Detection

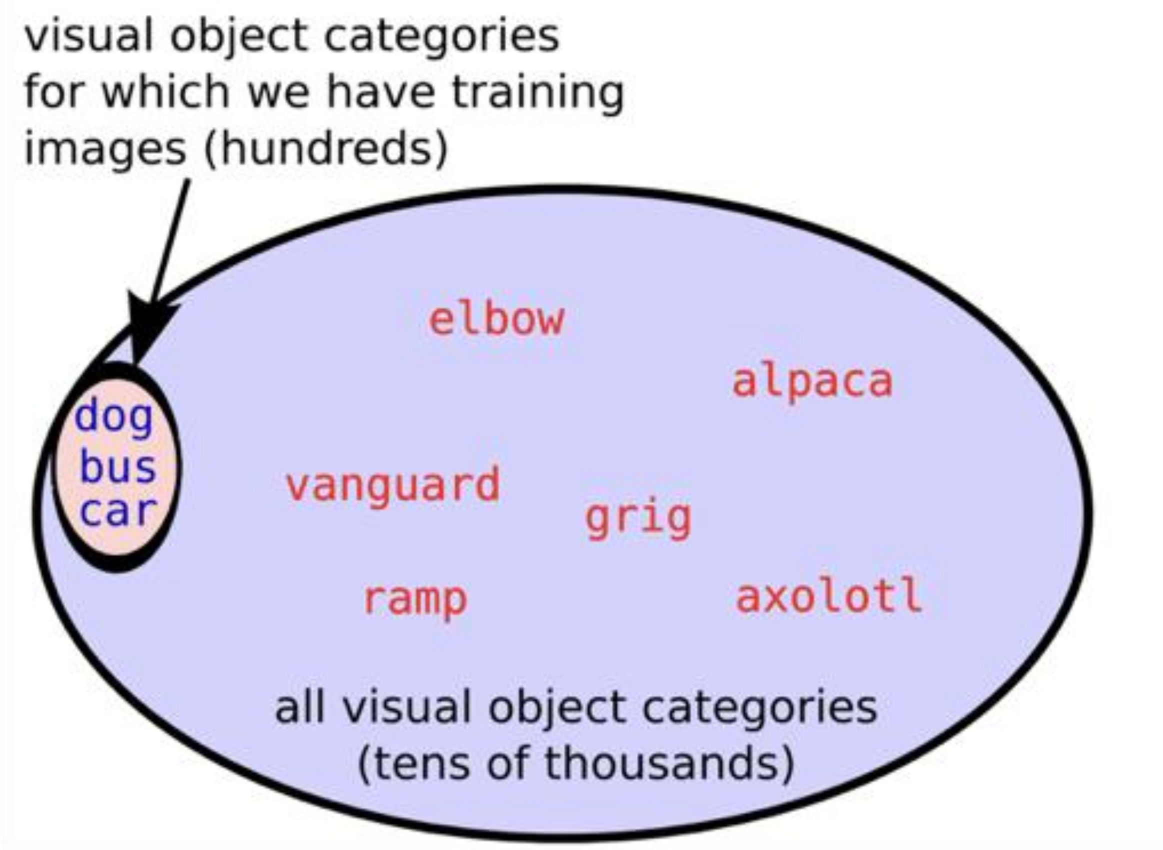
Dilemma of Large-scale Supervised Recognition

- **Problems:**
  1. No memory: Knowledge learned is not retained
     - Knowledge is not cumulative.
     - Cannot learn by leveraging past learned knowledge
  2. Needs a large number of training examples.
     - Humans can learn effectively from a few examples.
     - Humans can learn to learn.



We never have enough training data to classify all the categories!

Dilemma of Large-scale Supervised Recognition



We never have enough training data to classify all the categories!

What we want? Learn as humans do.

1. Humans have the ability to recognize without seeing examples (zero-shot learning);
2. Retain learned knowledge from previous tasks & use it to help future learning (transfer learning);

One-shot Learning

One-shot learning aims to learn information about object categories from ***one, or only a few***, training images.

Survey of One-shot Learning

1

Outlines of One-shot Learning Methods

1.Directly supervised learning-based approaches

- do not use auxiliary data;
- directly learn one-shot classier;

2.Transfer learning-based approaches:

- Use knowledge from auxiliary data
- The paradigm of *learning to learn* or *meta-learning*

Fu et al. Recent Advances in Zero-Shot Recognition: Toward Data-Efficient Understanding of Visual Content. IEEE SPM 2018

Directly supervised learning-based approaches

- Instance-based learning
  - K-nearest neighbor
- Non-parameteric methods
  - Fei-Fei et al. A Bayesian approach to unsupervised one-shot learning of object categories, CVPR 2003
  - Fei-Fei, L., Fergus, R., Perona, P.: One-shot learning of object categories. TPAMI 2006

Fu et al. Recent Advances in Zero-Shot Recognition: Toward Data-Efficient Understanding of Visual Content. IEEE SPM 2018

Transfer learning-based approaches

- Attribute-based algorithms (M2LATM, TMV-HLP, and so on)
- Meta-learning algorithms:
  - MAML
  - META-LEARN LSTM
  - Meta-Net
- Metric-learning algorithms
  - Matching Nets
  - PROTO-NET
  - RELATION NET

Fu et al. Recent Advances in Zero-Shot Recognition: Toward Data-Efficient Understanding of Visual Content. IEEE SPM 2018

Data Augmentation for One-shot Learning

1. Learning one-shot models by utilizing the manifold information of large amount of unlabelled data in a semi-supervised or transductive setting
2. Adaptively learning the one-shot classifiers from off-shelf trained models
3. Borrowing examples from relevant categories or semantic vocabularies to augment the training set;
4. Synthesizing new labelled training data by rendering virtual examples or composing synthesized representations or distorting existing training examples;
5. Generating new examples using Generative Adversarial Networks (GANs);
6. Attribute-guided augmentation (AGA) to synthesize samples at desired values or strength.

Chen et al. Semantic Feature Augmentation in Few-shot Learning. https://arxiv.org/abs/1804.05298, submitted to ECCV2018

1

Transfer learning-based approaches

Most of previous methods cannot beat ResNet-18

Methods	*mini*ImageNet (%)		CUB-200(%)	
	1-shot	5-shot	1-shot	5-shot
MAML [38]	48.70±1.84	63.11±0.92	38.43	59.15
Meta-SGD [39]	50.47±1.87	64.03±0.94	-	-
DEML+Meta-SGD [40]	**58.49**±0.91	71.28±0.69	-	-
META-LEARN LSTM [41]	43.44±0.77	60.60±0.71	40.43	49.65
Meta-Net [42]	49.21±0.96	-	-	-
Matching Nets [36]	43.56±0.84	55.31±0.73	49.34	59.31
PROTO-NET [37]	49.42±0.78	68.20±0.66	45.27	56.35
RELATION NET [43]	57.02±0.92	71.07±0.69	-	-
MACO [44]	41.09±0.32	58.32±0.21	60.76	74.96
ResNet-18	52.73±1.44	73.31±0.81	66.54±0.53	82.38±0.43
Ours: ResNet-18+Dual TriNet	58.12±1.37	**76.92**±0.69	**69.61**±0.46	**84.10**±0.35

Table 1. Results on *mini*ImageNet and CUB-200. The "±" indicates 95% confidence intervals over tasks. Note that "±" is not reported on CUB-200 in previous works.

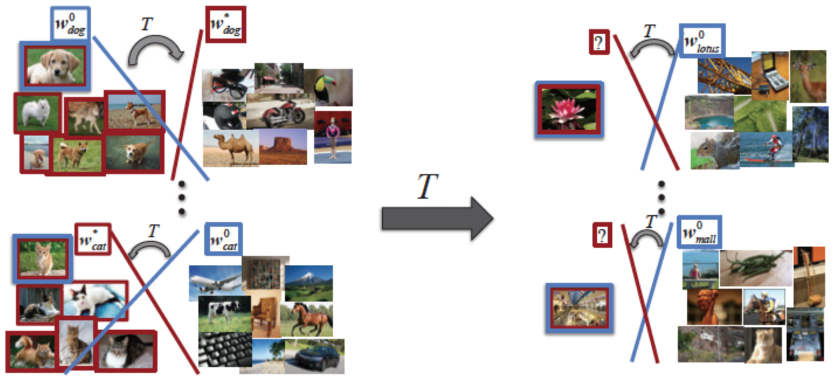
Codes: https://github.com/tankche1/Semantic-Feature-Augmentation-in-Few-shot-Learning

Chen et al. Semantic Feature Augmentation in Few-shot Learning. Axiv:1804.05298, submitted to ECCV2018

Learning to Learn: Model Regression Networks for Easy Small Sample Learning (ECCV2016)

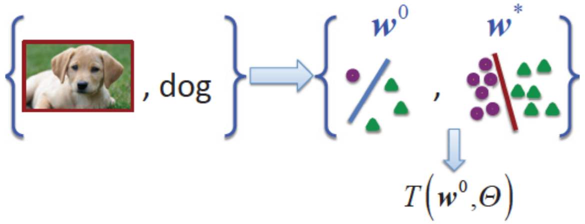
MOTIVATION

- **Inter-class knowledge**: A generic transformation T: small-sample models $\boldsymbol{w}^0 \rightarrow$ large-sample models $\boldsymbol{w}^*$
- **Recognition of novel categories from few examples**: Predict the target models $\boldsymbol{w}$ by transferring T



LEARNING MODEL TRANSFORMATION

- T learns **predictive structures** in the model space
  - Discriminative representation of natural intra-class variability: Sparse samples $\rightarrow$ a category cluster
  - Duality perspective: Feature $\leftrightarrow$ classifier spaces
  - Alternative parametric way of model distillation
- T can be learned
  - On a large collection of model pairs $\{(\boldsymbol{w}^0, \boldsymbol{w}^*)\}$
  - By a high-capacity regression function $T(\boldsymbol{w}^0, \Theta)$



Matching Networks for One Shot Learning (NIPS2016)

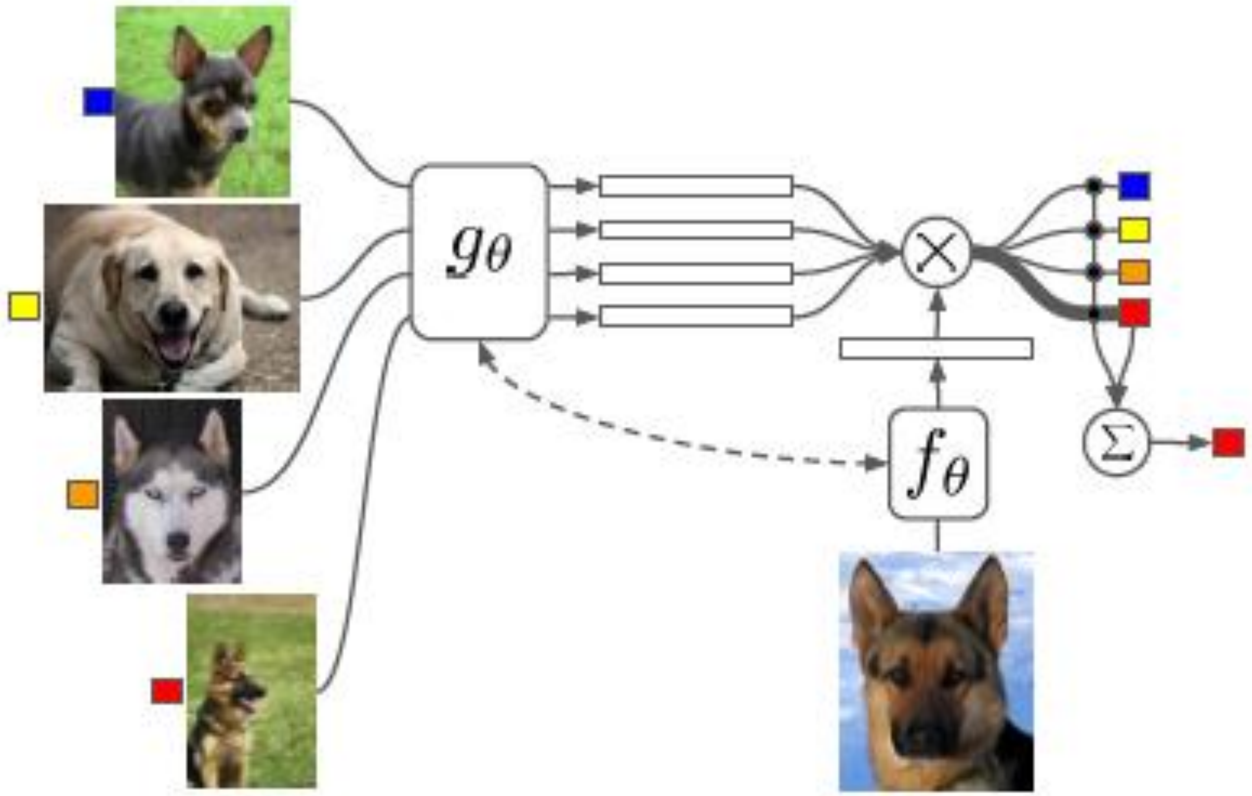


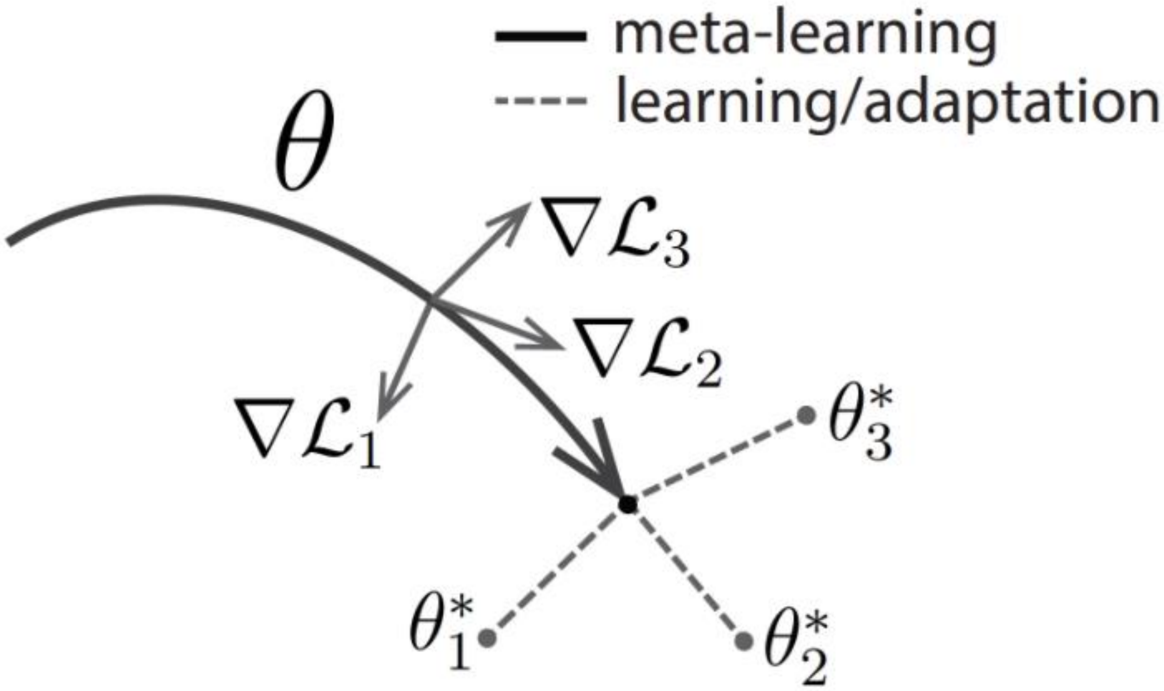
Figure 1: Matching Networks architecture

Metric learning based on deep neural features:

The training procedure is chosen carefully so as to match inference at test time.

The network maps a small labelled support set and an unlabelled example to its label, obviating the need for fine-tuning to adapt to new class types.

Model-Agnostic Meta-Learning (MAML) for Fast Adaptation of Deep Networks (ICML2017)



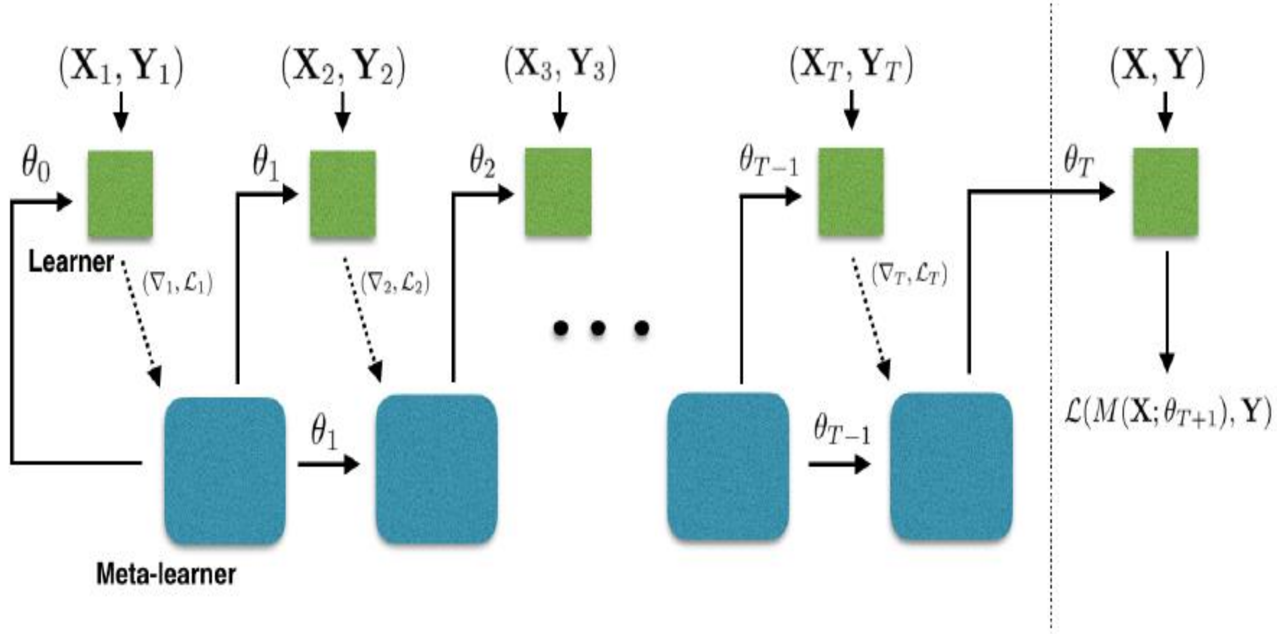
Model parameters are explicitly trained such that a small number of gradient steps with a small amount of training data of new task will make good generalization.

Goal: train a model on a variety of learning tasks, such that it can solve new learning tasks using only a small number of training samples.

MAML can learn good initial neural network weights which can be easily fine-tuned for unseen categories.

Optimization as a model for few-shot learning (ICLR2017)

META-LEARN LSTM learn a general initialization of the learner (classifier) network that allows for quick convergence of training.



Problem: Gradient-based optimization in high capacity classifiers requires many iterative steps over many examples to perform well.

Solution: an LSTM-based meta-learner model to learn the exact optimization algorithm to train another learner neural network classifier in the few-shot learning.

Meta networks (ICML2017)

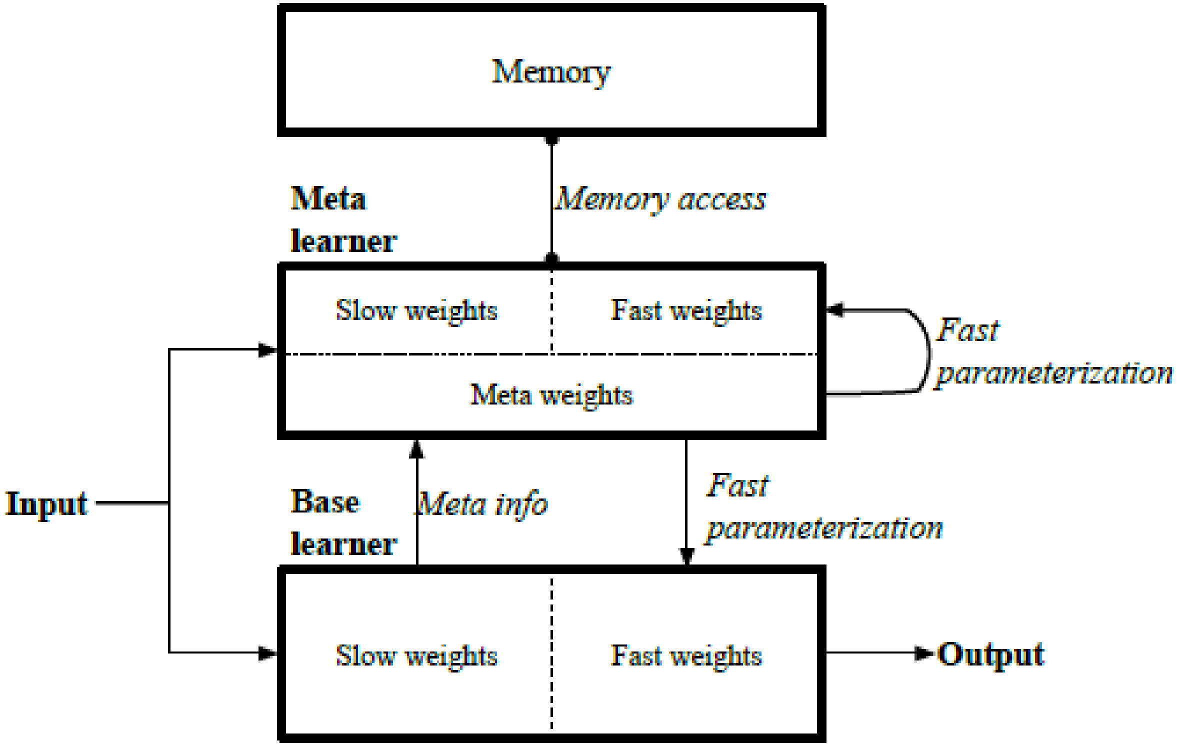


Figure 1. Overall architecture of Meta Networks.

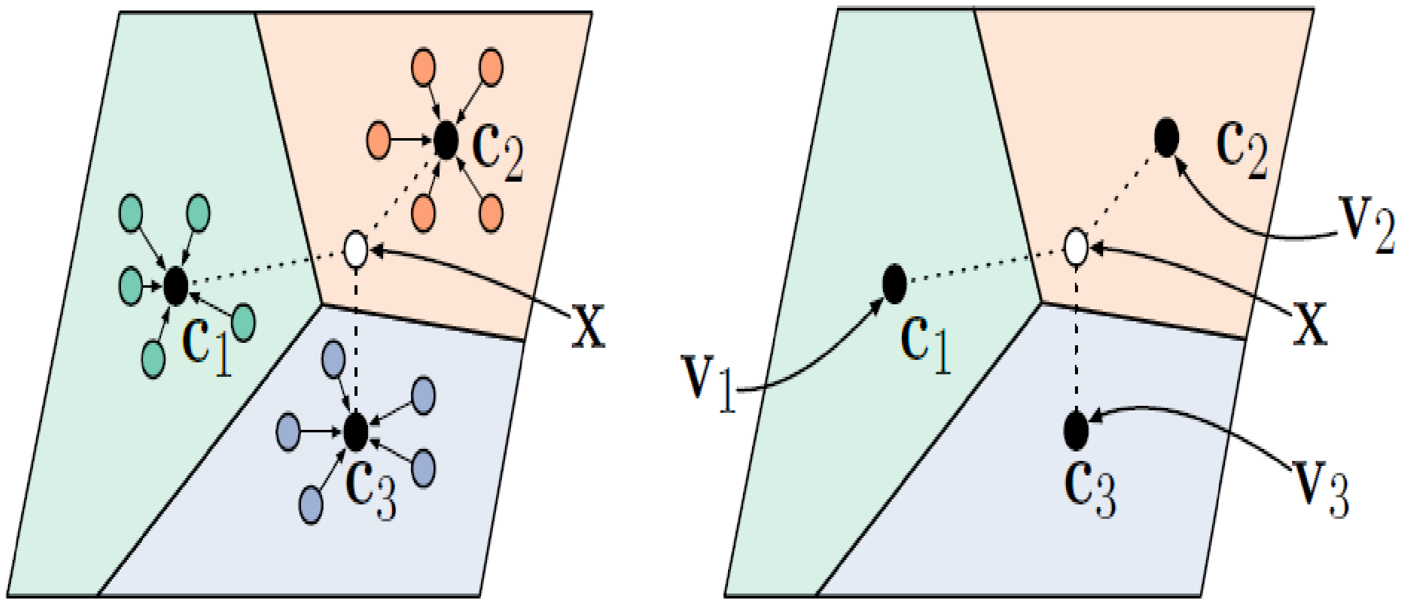
Meta-Net: learns a meta-level knowledge across tasks and shifts its inductive biases via fast parameterization for rapid generalization.

The loss of MetaNet:

(1) a representation (i.e. embedding) loss defined for the good representation learner criteria

(2) a main (task) loss used for the input task objective

Prototypical Networks for Few-shot Learning (NIPS2017)



(a) Few-shot (b) Zero-shot

Metric-learning algorithm.

Prototypical networks learn a metric space in which classification can be performed by computing distances to prototype representations of each class.

Learning to compare: Relation network for few-shot learning (CVPR2018)

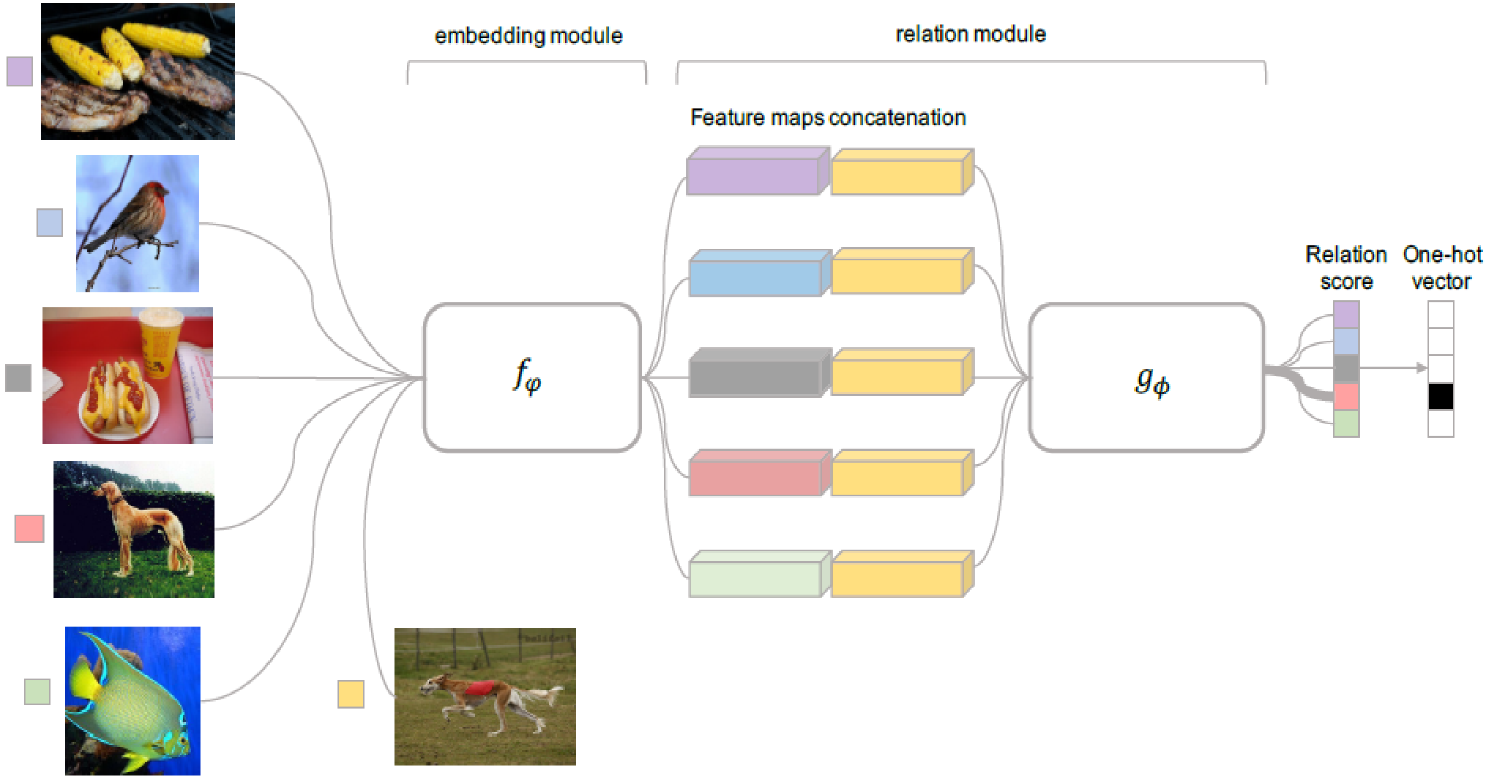


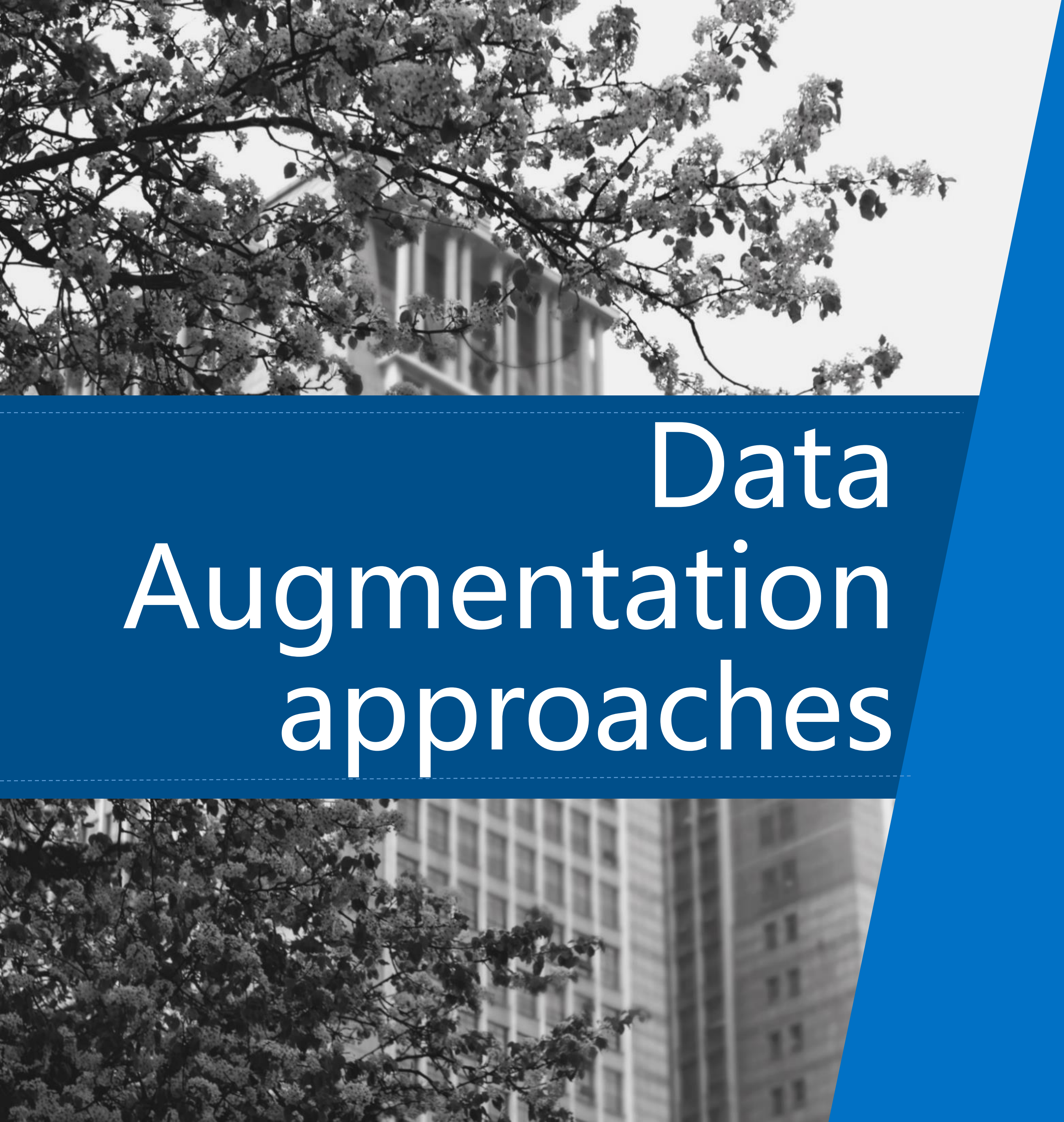
Figure 1: Relation Network architecture with a 5-way 1-shot 1-query example.

Relation Network (RN):
learns a deep distance metric to compare a small number of images within episodes, simulating few-shot setting.

Once trained, a RN is able to classify images of new classes by computing relation scores between query images and the few examples of each new class without further updating the network.

Other References

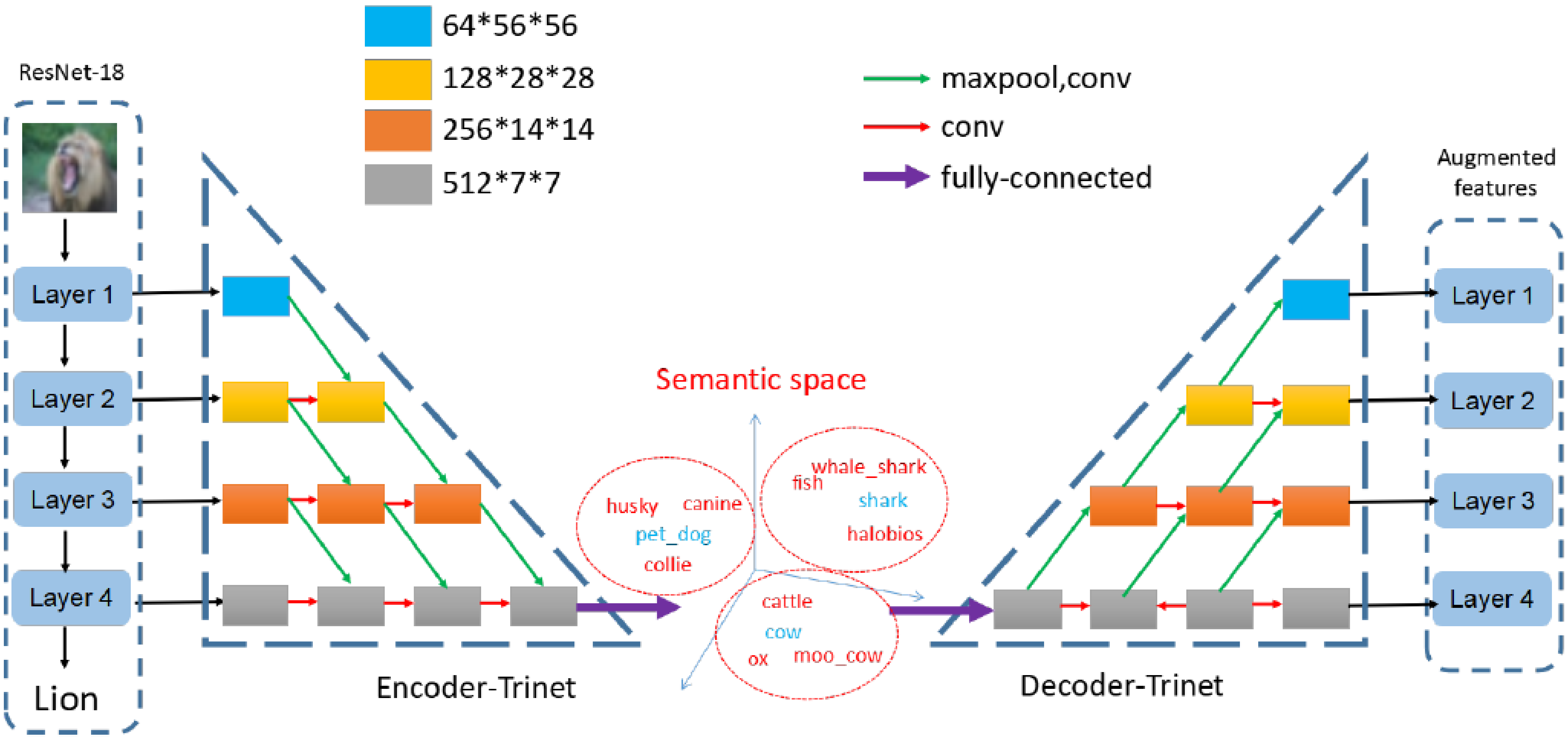
- Siamese Neural Networks for One-shot Image Recognition(ICML15 Deep Learning Workshop)
- Learning Structure and Strength of CNN Filters for Small Sample Size Training (CVPR2018)
- FEW-SHOT LEARNING WITH GRAPH NEURAL NETWORKS(ICLR 2018)
- Multi-attention Network for One Shot Learning (CVPR2017)
- META-LEARNING FOR SEMI-SUPERVISED FEW-SHOT CLASSIFICATION (ICLR 2018)
- One-shot Learning with Memory-Augmented Neural Networks (arxiv2016)
- Generative Adversarial Residual Pairwise Networks for One Shot Learning (arxiv: 2017)



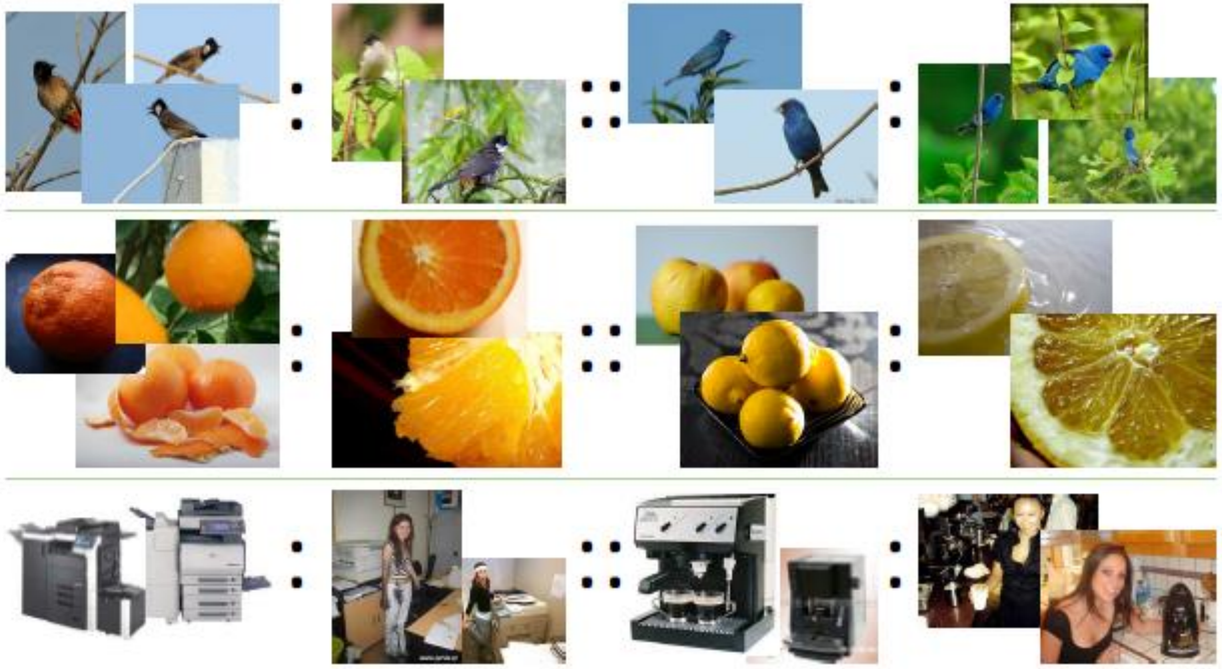
Data Augmentation approaches

2

Semantic Feature Augmentation in Few-shot Learning (submitted to ECCV18)

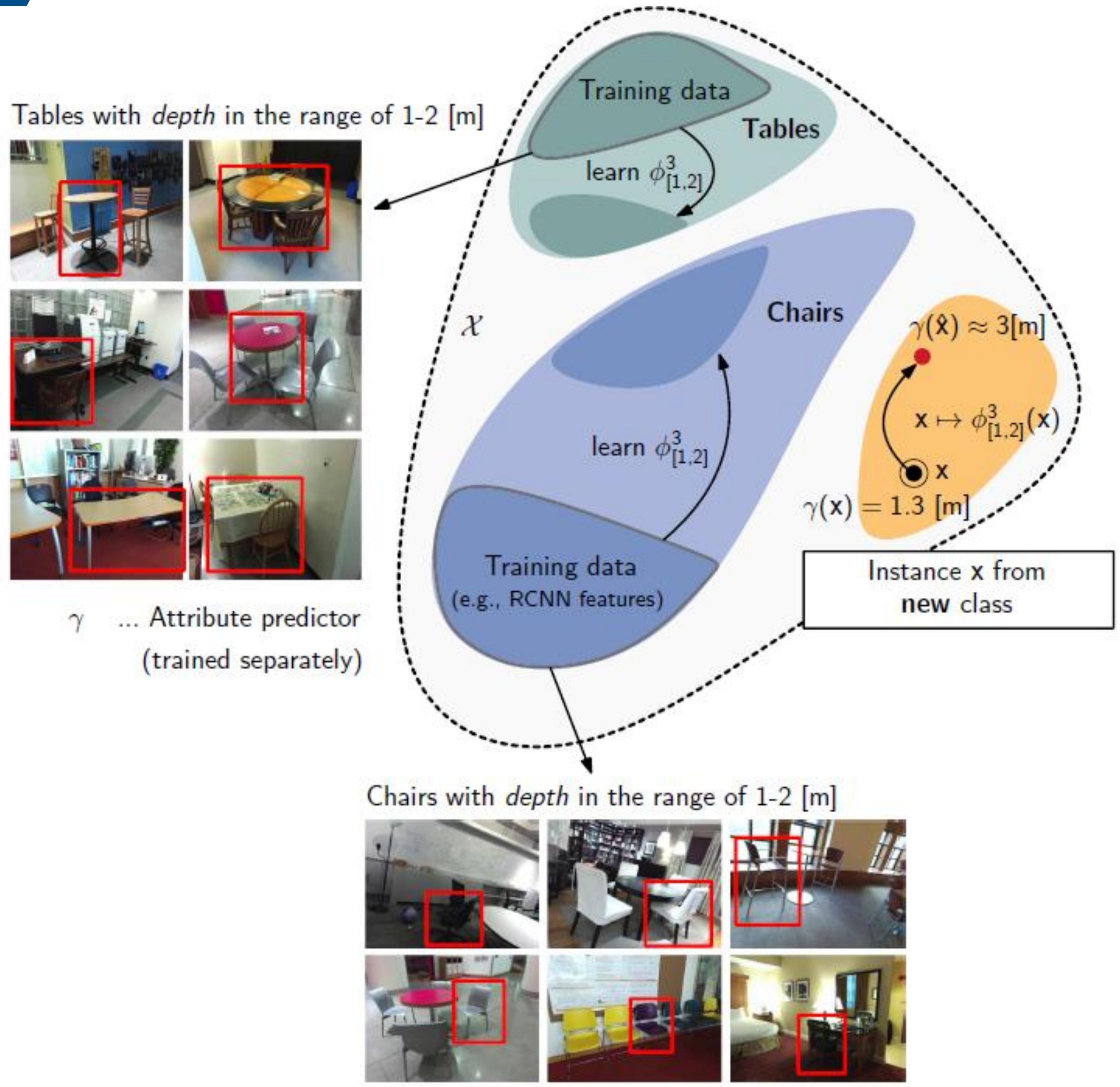


Low-shot Visual Recognition by Shrinking and Hallucinating Features (ICCV2017)



We present a low-shot learning benchmark on complex images that mimics challenges faced by recognition systems in the wild. We then propose (1) representation regularization techniques, and (2) techniques to hallucinate additional training examples for data-starved classes.

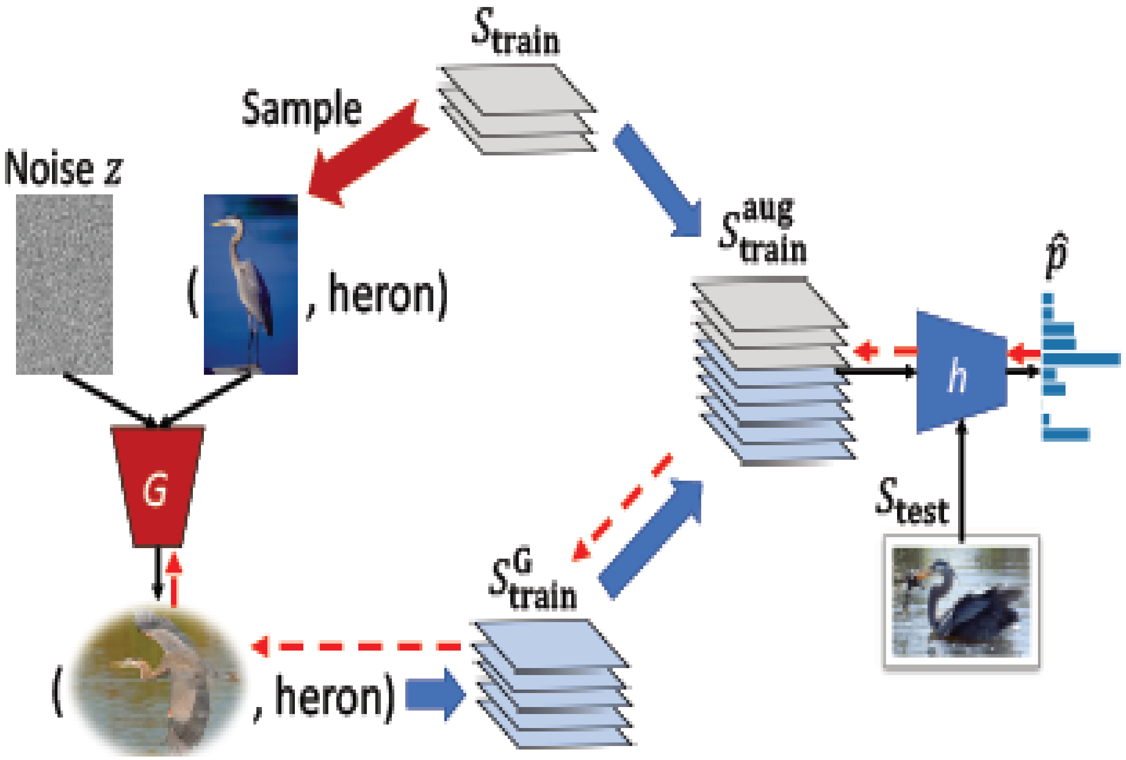
AGA : Attribute-Guided Augmentation (CVPR2017)



Attributed-guided augmentation (AGA) learns a mapping that allows to synthesize data such that an attribute of a synthesized sample is at a desired value or strength. The data is limited with no attribute annotations.

They propose to perform augmentation in feature space instead. The network is implemented as an encoder-decoder architecure.

Low-Shot Learning from Imaginary Data (CVPR2018)



Combining a meta-learner with a "hallucinator" that produces additional training examples, and optimizing both models jointly.

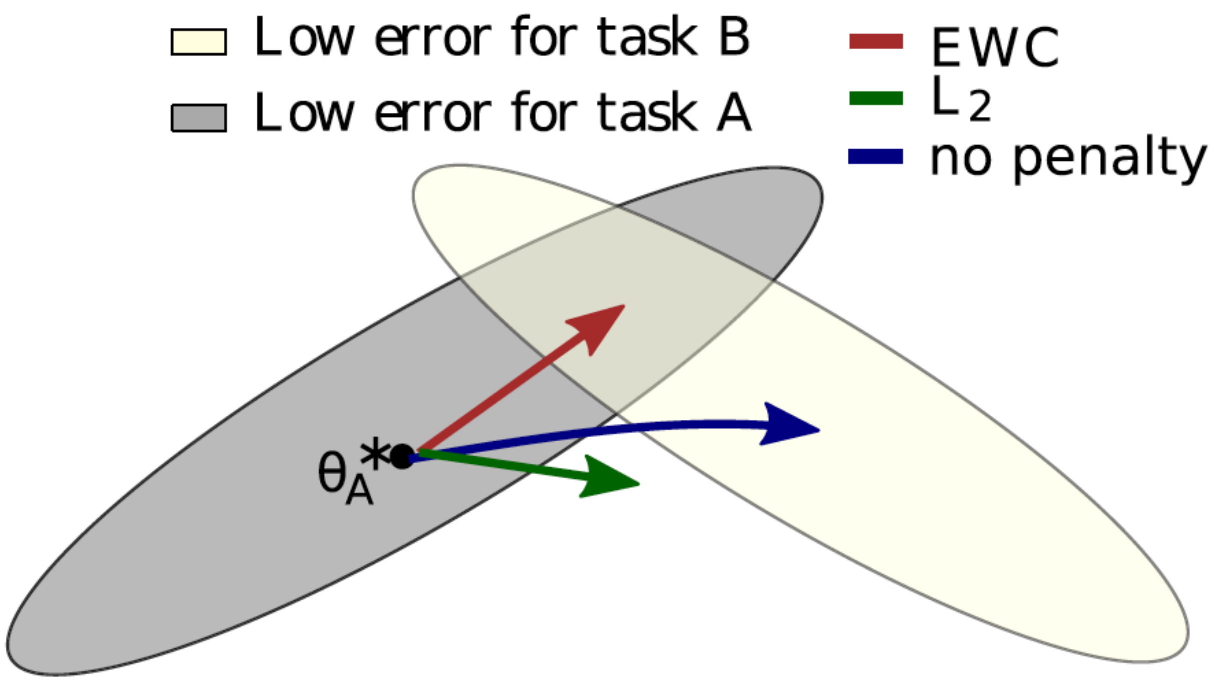
Our hallucinator can be incorporated into a variety of meta-learners and provides significant gains.

3

Beyond one-shot learning

Overcoming catastrophic forgetting in neural networks (PNAS, 2017)

One weakness of deep models: unable to learn multiple tasks sequentially.
Catastrophic forgetting!!



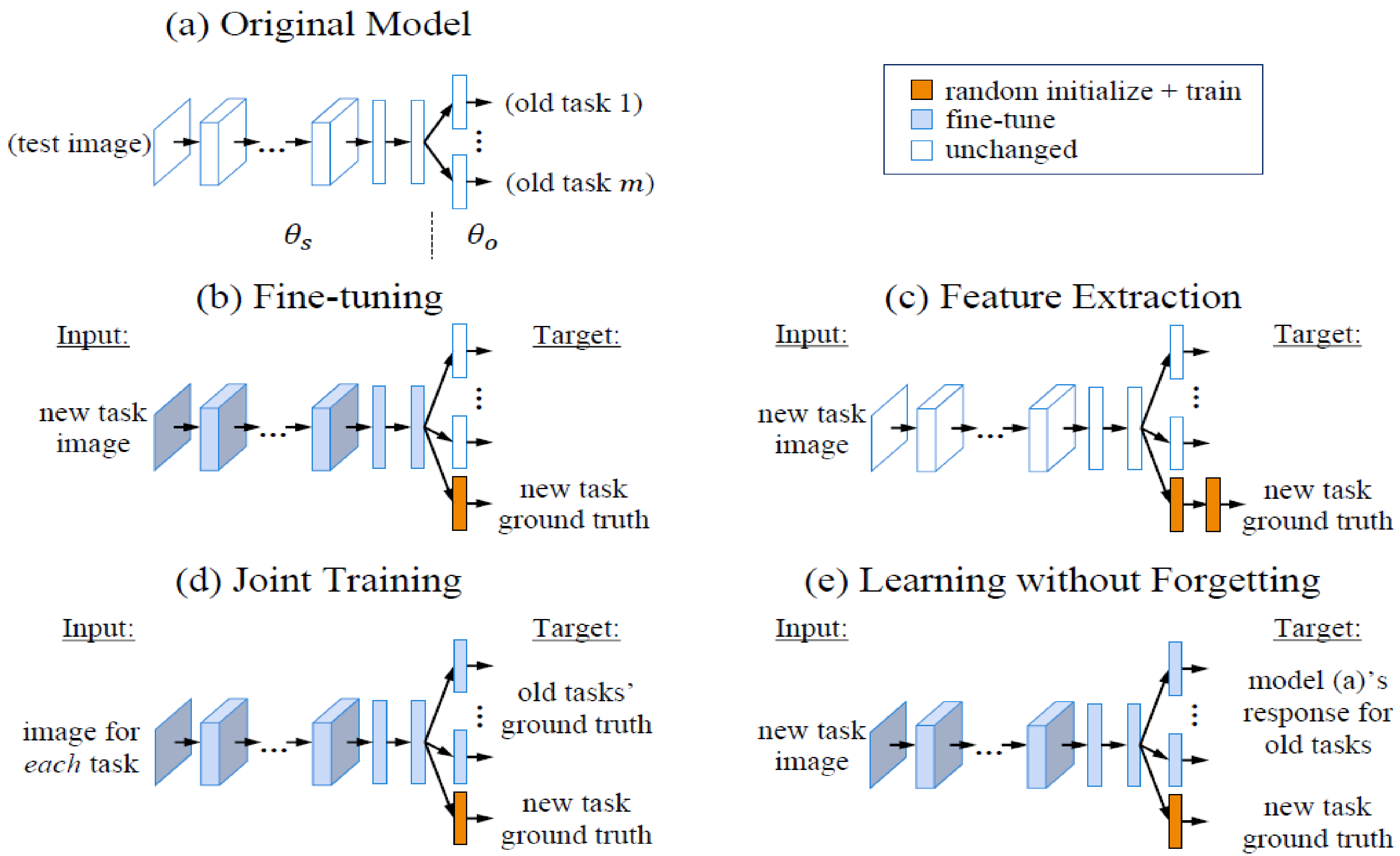
If we constrain each weight with the same coefficient (green arrow), the restriction imposed is too severe and we can remember task A only at the expense of not learning task B.

EWC: finds a solution for task B without incurring a significant loss on task A (red arrow) by explicitly computing how important weights are for task A.

The proposed elastic weight consolidation (EWC) ensures task A is remembered while training on task B.

Learning without Forgetting (Li et al. ECCV16, TPAMI 17)

The proposed method uses only new task data to train the network while preserving the original capabilities.



References

[M2LATM]   Learning Multi-modal Latent Attributes, IEEE TPAMI 2014
[TMV-HLP] Transductive Multi-View Zero-Shot Learning, IEEE TPAMI 2015

THANKS