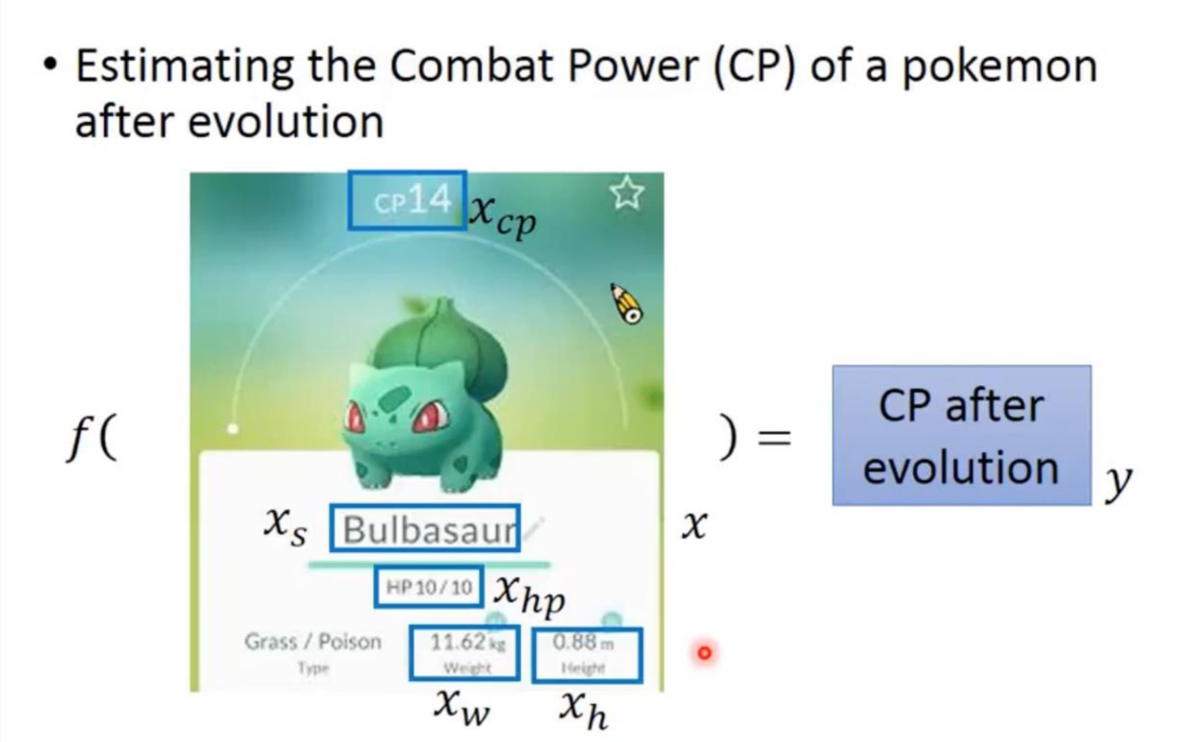
TRANSFER LEARNING: AN INTRODUCTION

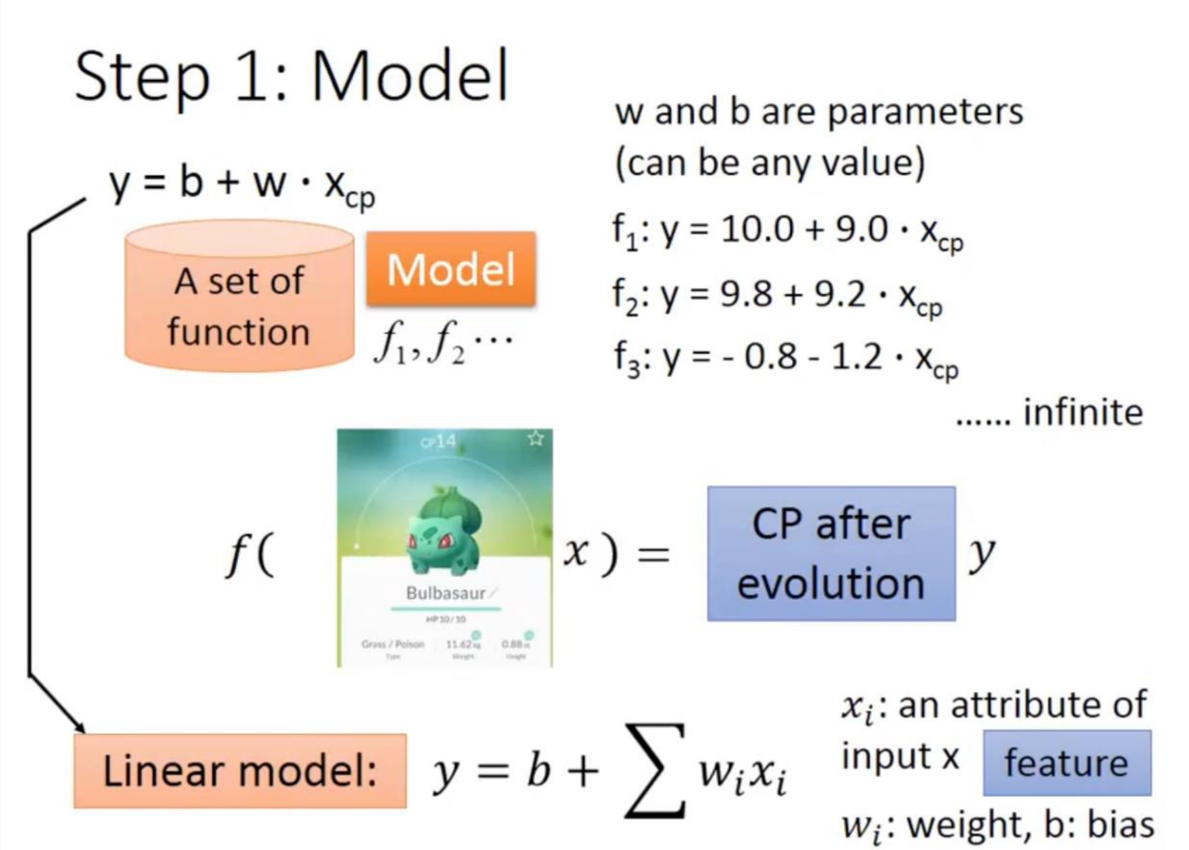
Zhangkai Wu
Data Science Institute
University of Technology Sydney

LINEAR REGRESSION

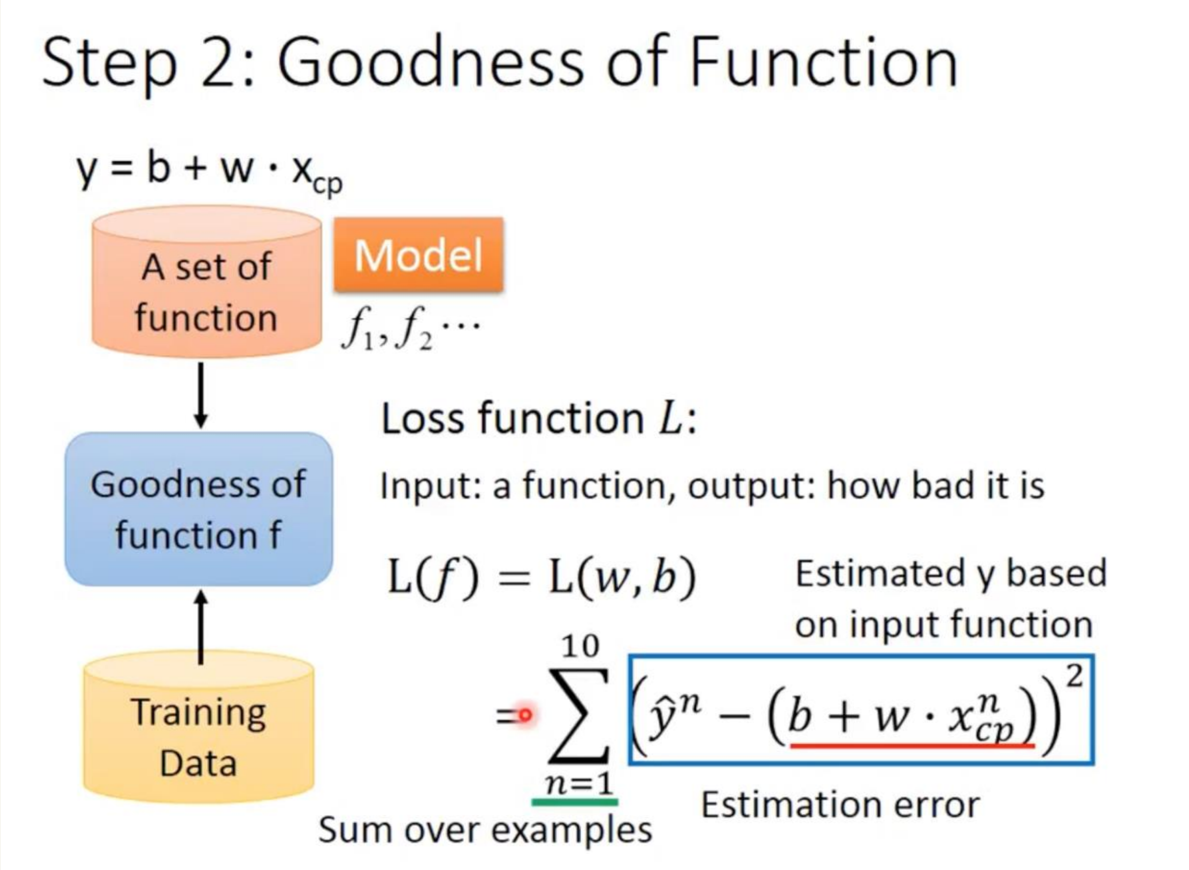
- Estimating the Combat Power (CP) of a pokemon after evolution



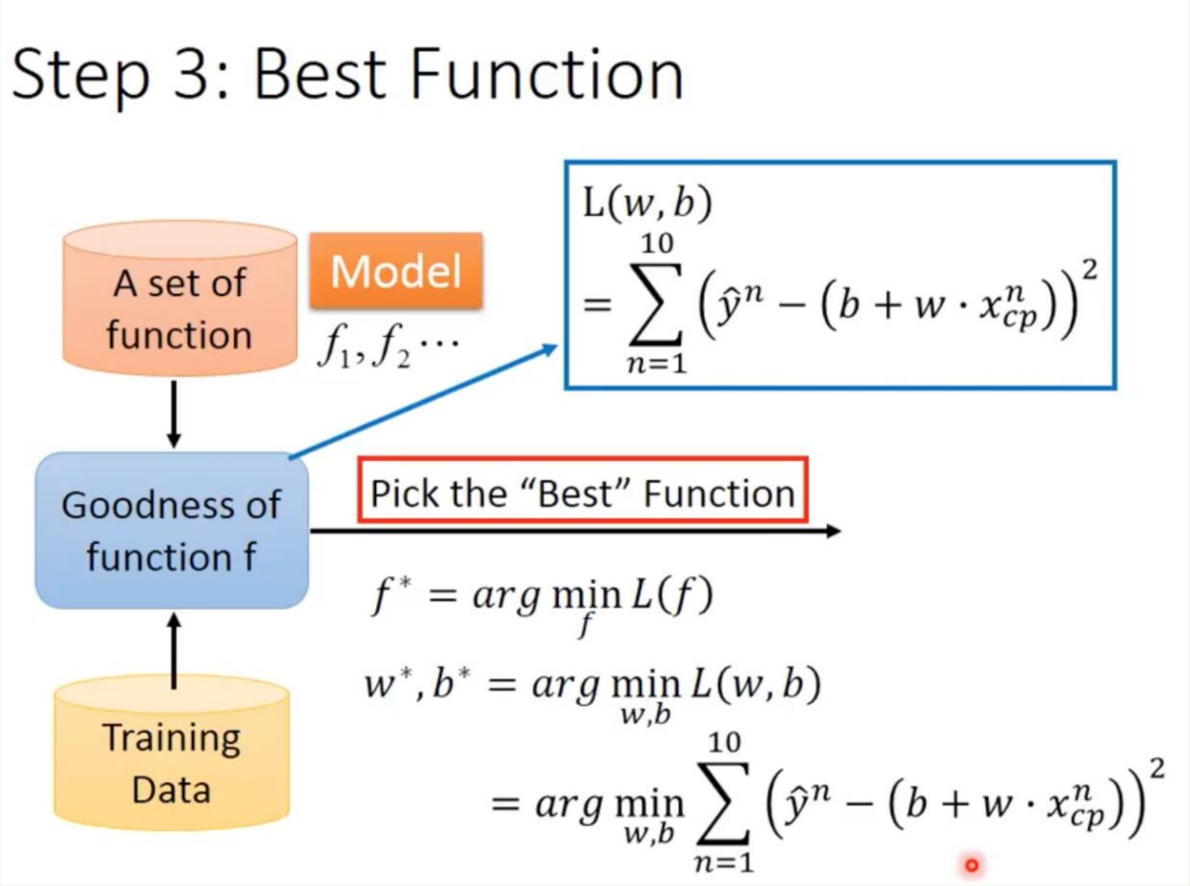
LINEAR REGRESSION



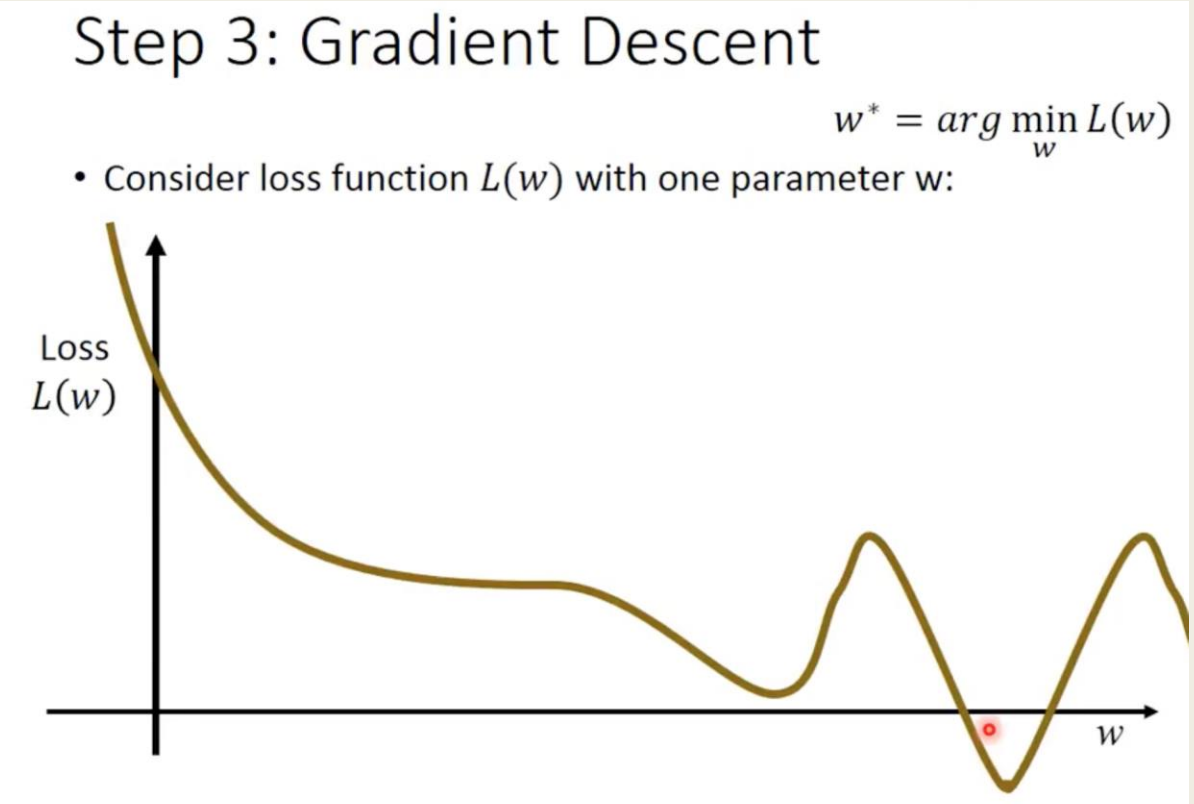
LINEAR REGRESSION



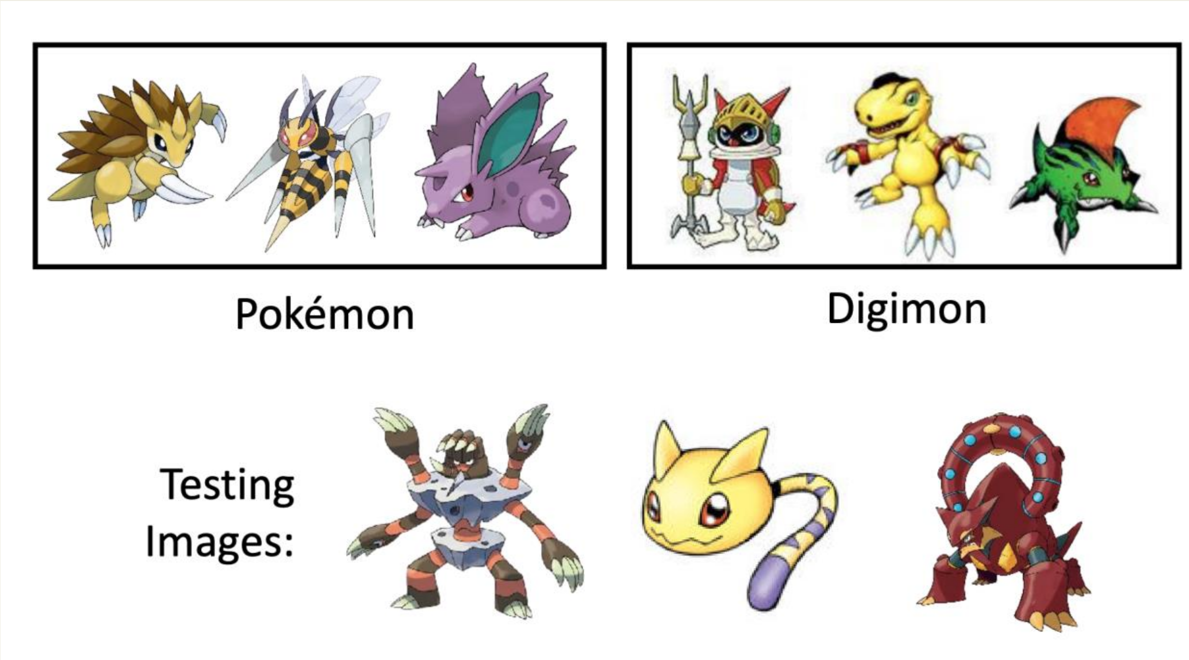
LINEAR REGRESSION



LINEAR REGRESSION



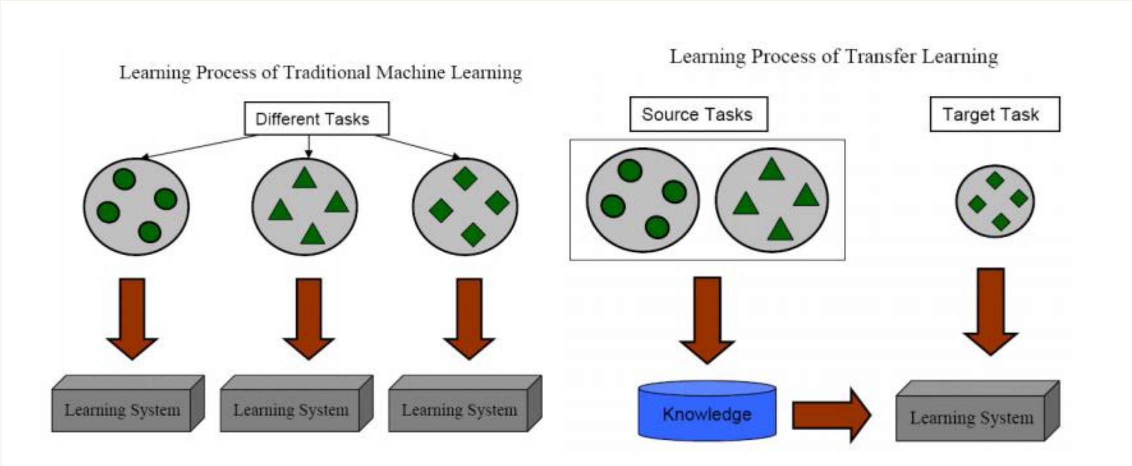
LINEAR REGRESSION



LINEAR REGRESSION

Background

- What is Transfer learning?
  - Transferring the knowledge of one model to perform a new task.



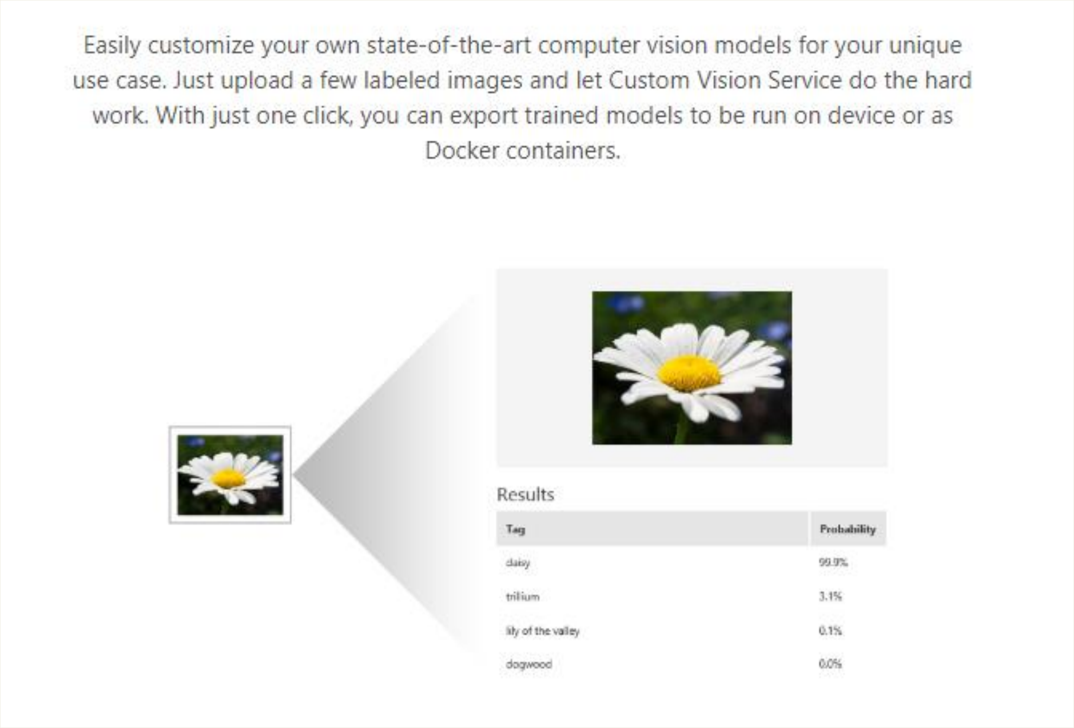
(a) Traditional Machine Learning (b) Transfer Learning

Background

- *Motivation*
  - Cheap
  - Time-consuming
  - More realistic

Applications

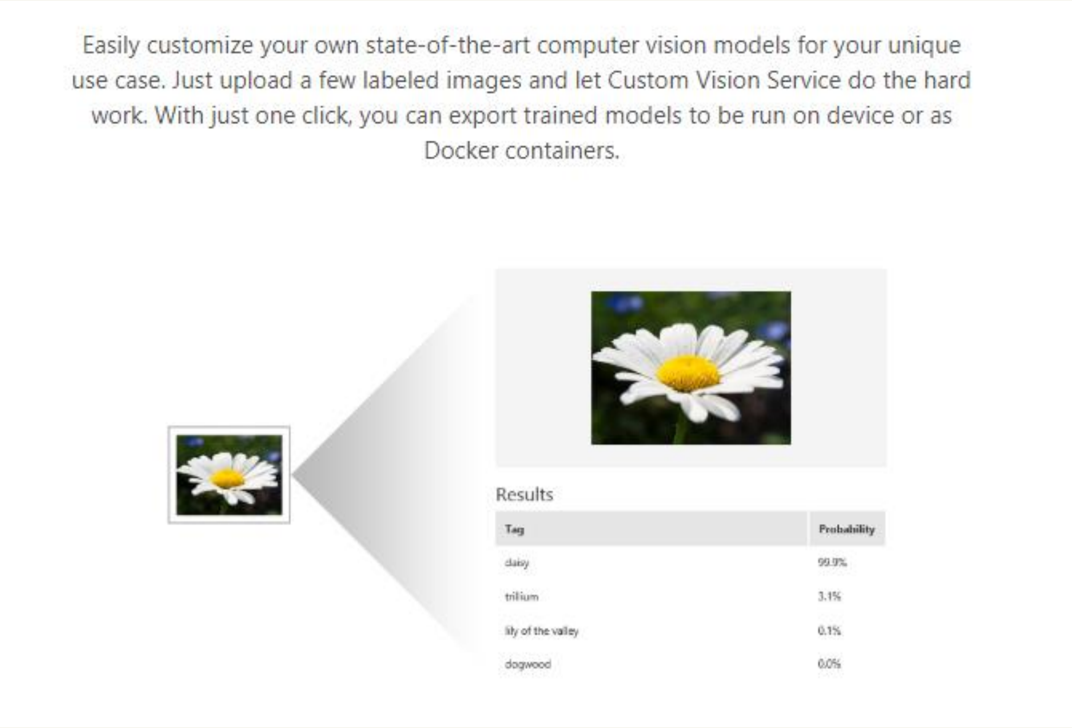
- *ML/DM/CV/NLP*
  - Image classification (most common): learn new image classes
  - Text sentiment classification
  - Text translation to new languages
  - Speaker adaptation in speech recognition
  - Question answering



Applications

– *Supervised/Unsupervised/semi-supervised*

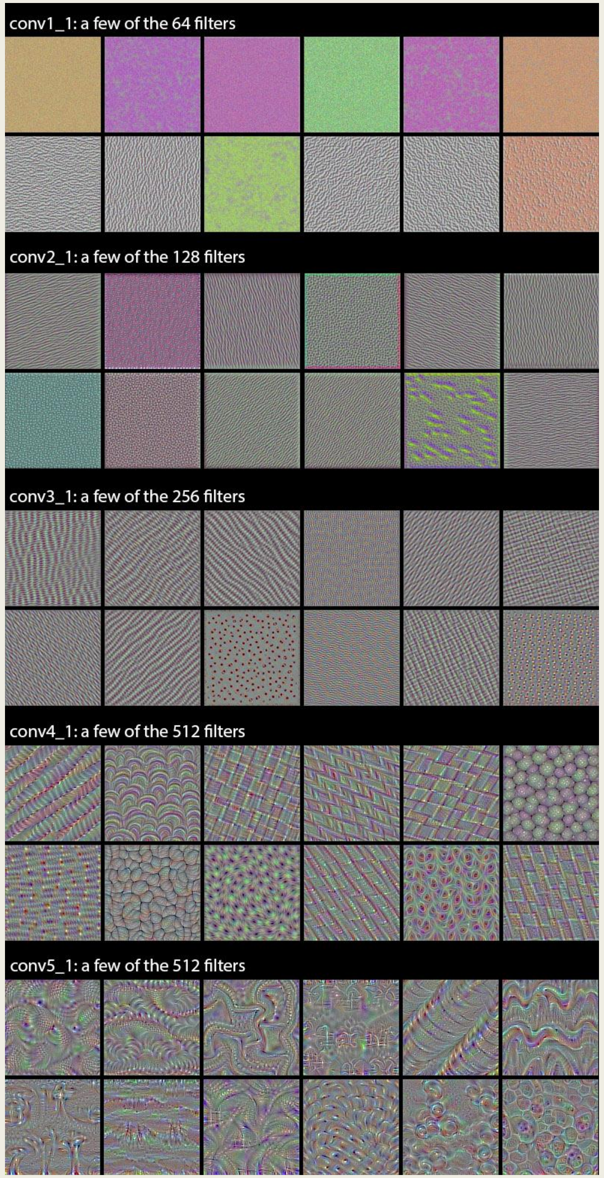
1. Classification, Regression
2. Clustering, Dimensionality Reduction, *Generation*

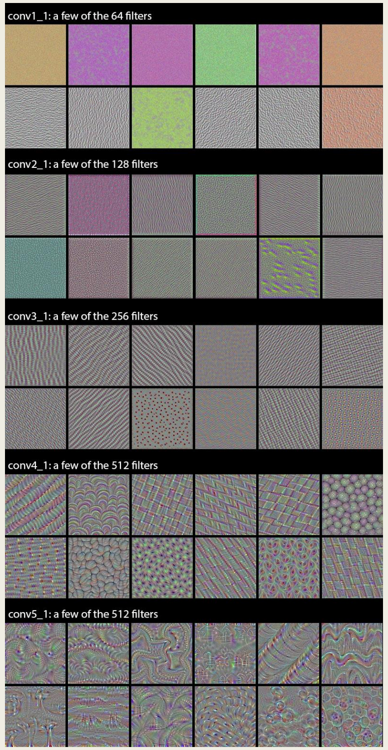


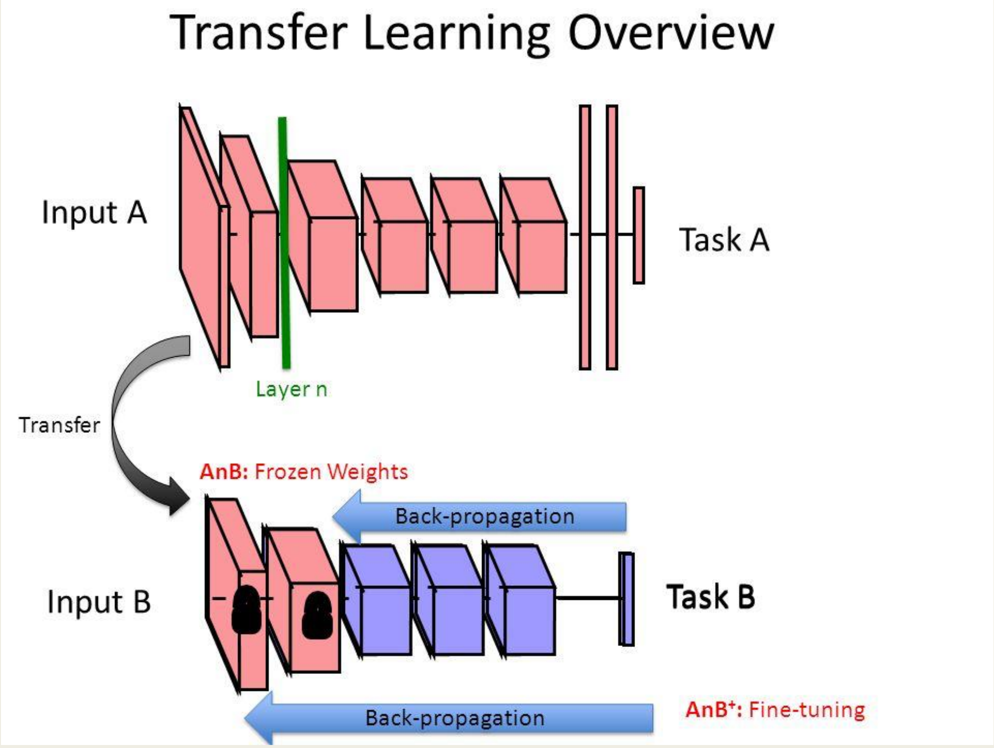
Transfer Learning in Neural Networks

- **Neural Network Layers: General to Specific**
  - Bottom/first/earlier layers: general learners
  - Low-level notions of edges, visual shapes
  - Top/last/later layers: specific learners
  - High-level features such as eyes, feathers

- **Example: VGG 16 Filters**





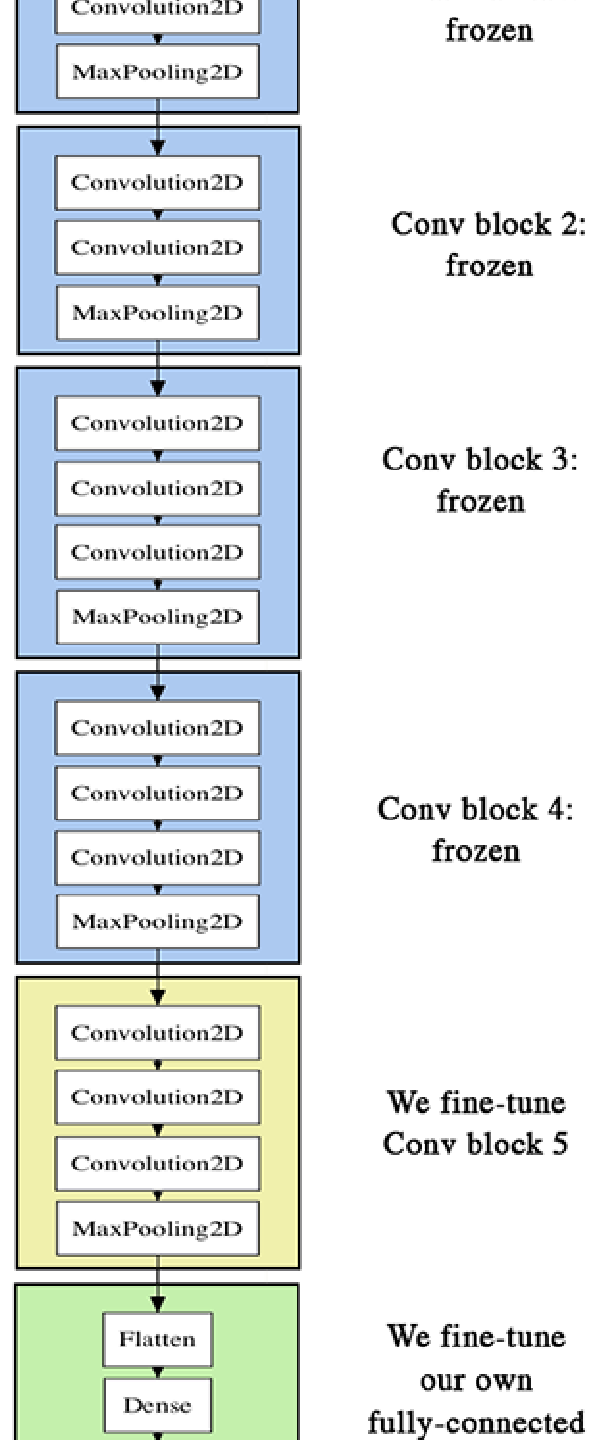


TRANSFER LEARNING IN NEURAL NETWORKS

Transfer Learning in Neural Networks

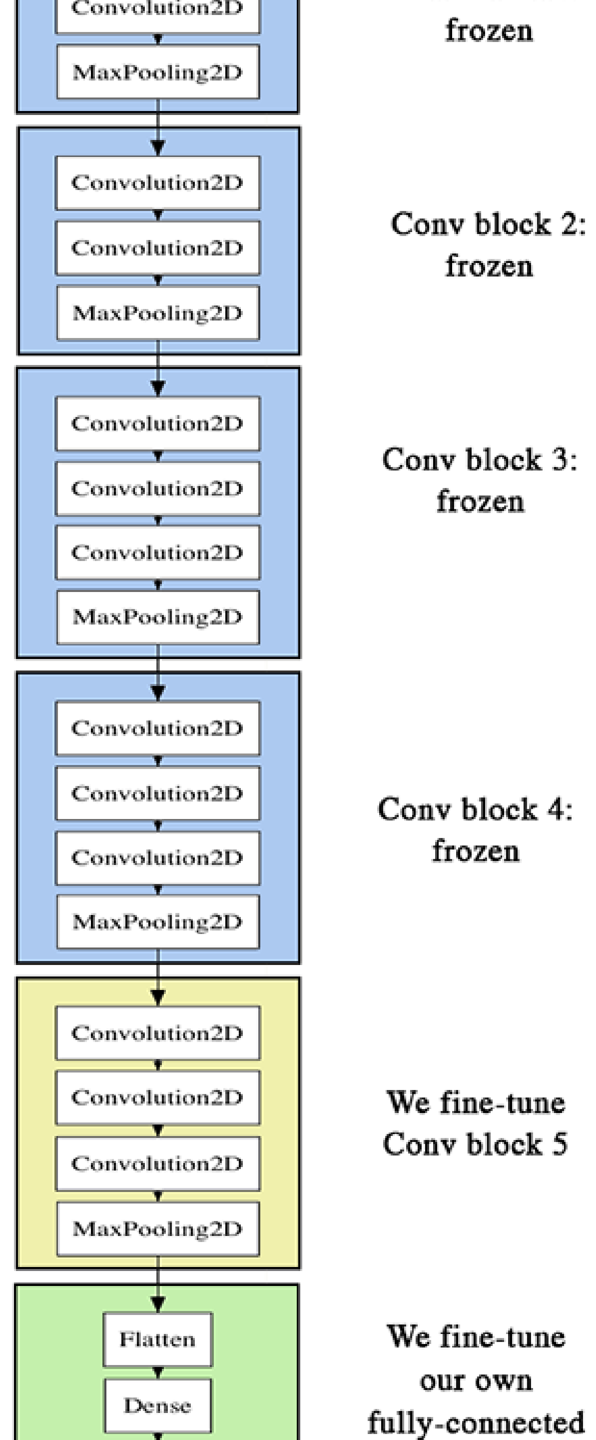
- **Process**

1. Start with a pre-trained network
2. Partition network into:
    1. Features: identify which layers to keep
    2. Classifiers: identify which layers to replace
3. Re-train classifier layers with new data
4. Unfreeze weights and fine-tune the whole network with lower learning rate



Transfer Learning in Neural Networks

- **Freezing and Fine-tuning**
- **Which layers to re-train?**
  - Depends on the domain
  - Start by re-training the last layers (last full-connected and last convolutional)
  - work backwards if performance is not satisfactory



Transfer learning is high-order complexity

Neural Networks

- RNN based, CNN based, VGG, AlexNet, KAN

Deep Model

- Generative model(vaes, gans, diffusion, flows)

Modality

- Image, sequence, tabular

Training methods

- Supervised, unsupervised, semi-supervised

Dive into Transfer Learning

Domain Generalization

perform well on any unseen domains. [2]

Transfer learning (Domain Adaption)[1]

Target data is exploited in training

Meta-Learning(Zero-shot Learning, Multi-Task Learning)

Learn to learn:

Lifelong Learning(Continual Learning)

Learn to not forget: [3]

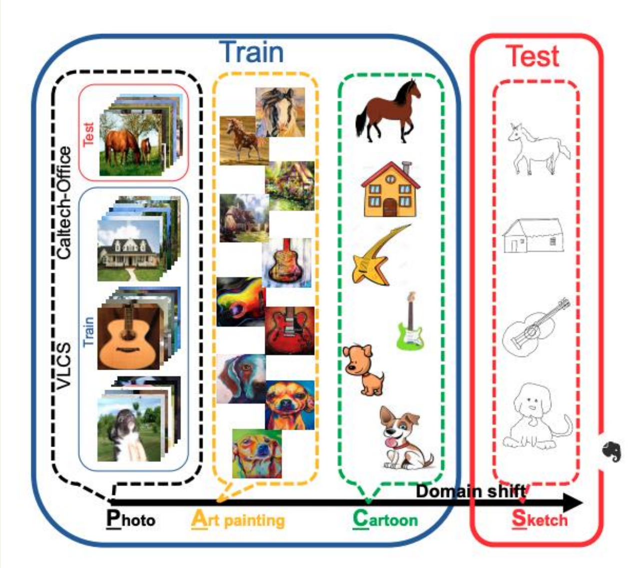
Test Time Adaptation

DA in test phrase

[1] A Survey on Transfer Learning
[2] Generalizing to Unseen Domains: A Survey on Domain Generalization
[3] Lifelong machine learning 2016

Domain Generalization in classification

- **PACS** : consists of ***Art painting, Cartoon, Photo*** and ***Sketch*** domains, which so far considers the largest domain shift as it is from the different image style depictions. [1]
- Access K similar but distinct source domain and predict on an unseen target domain[2]



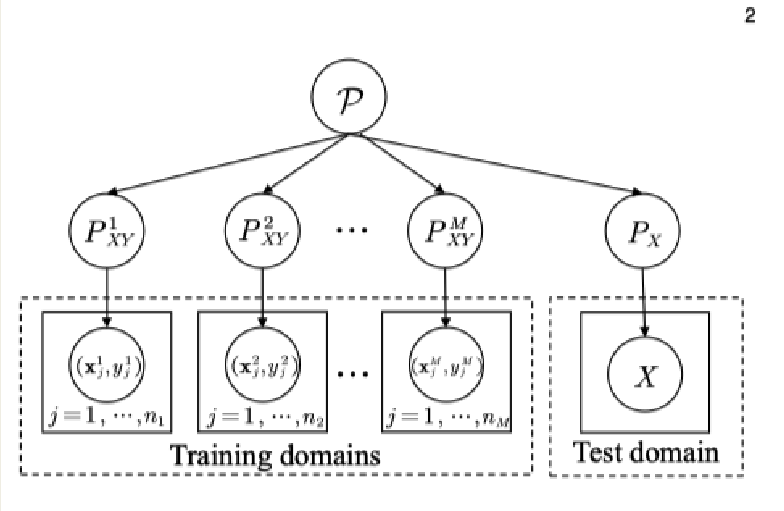
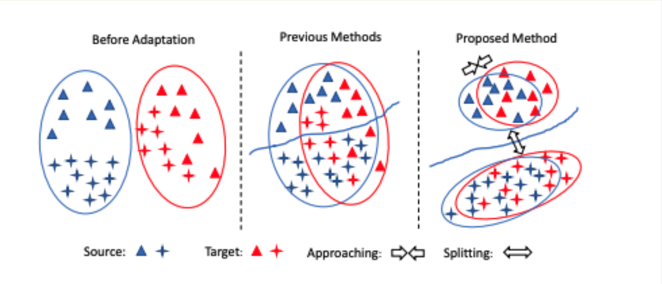


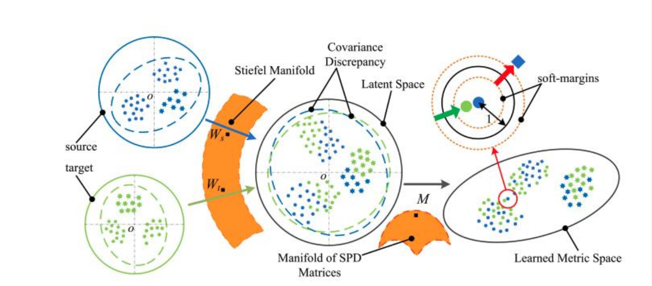
Fig. 2. Illustration of domain generalization. Adapted from [6].

[1] Deeper, Broader and Artier Domain Generalization ICCV2017
[2] Generalizing to Unseen Domains: A Survey on Domain Generalization TKDE 22

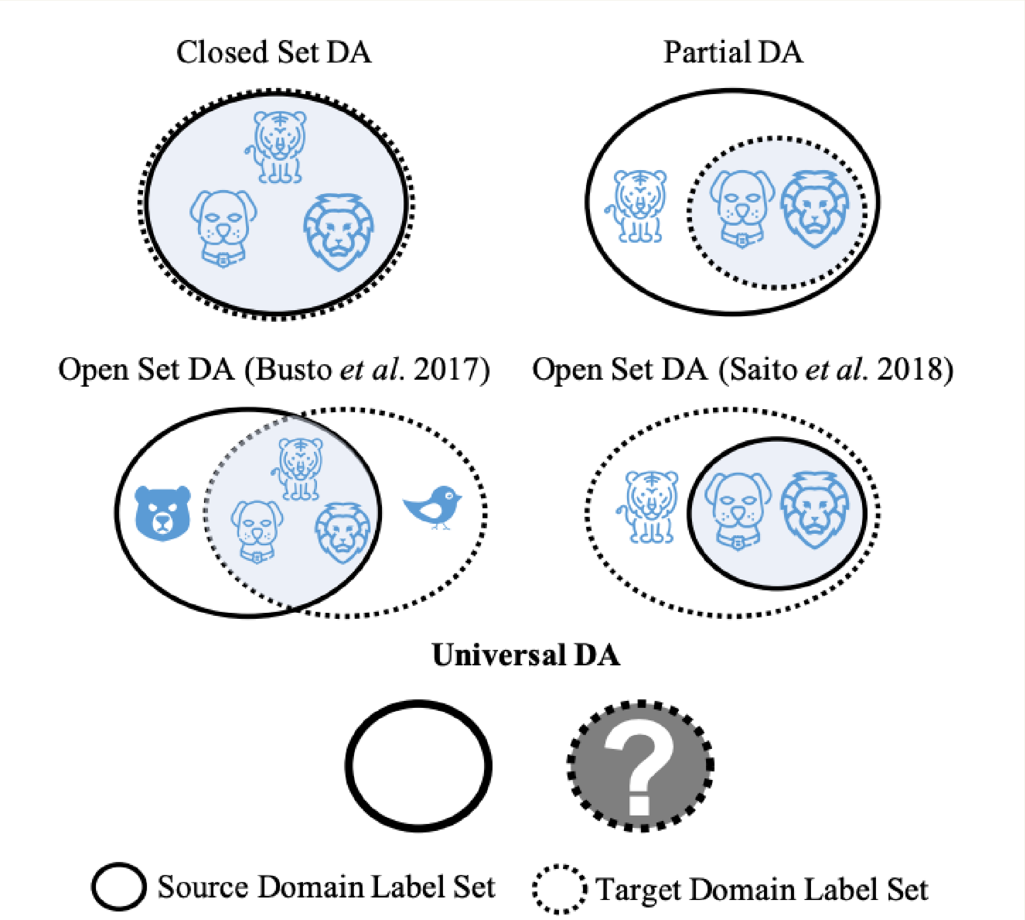
UDA benchmarks

- **Office-31:** contains 31 object categories in three domains: Amazon, DSLR and Webcam[1]
  - ***Amazon domain captured from online merchants with clean background and unified scale***
  - ***DSLR domain low noise and high resolution images***
  - ***Webcam domain significant noise and color as well as white balance artifacts***
- the entire data of the target domain is used for training and testing[2,3]





[1] Adapting Visual Category Models to New Domains ECCV 2010
[2] Learning an Invariant Hilbert Space for Domain Adaptation CVPR17
[2] Contrastive Adaptation Network for Unsupervised Domain Adaptation, CVPR19



DOMAIN ADAPTATION

Target data is exploited in training

[1] Universal Domain Adaptation CVPR2019

Continual Learning on ASC

- Learning to not forget

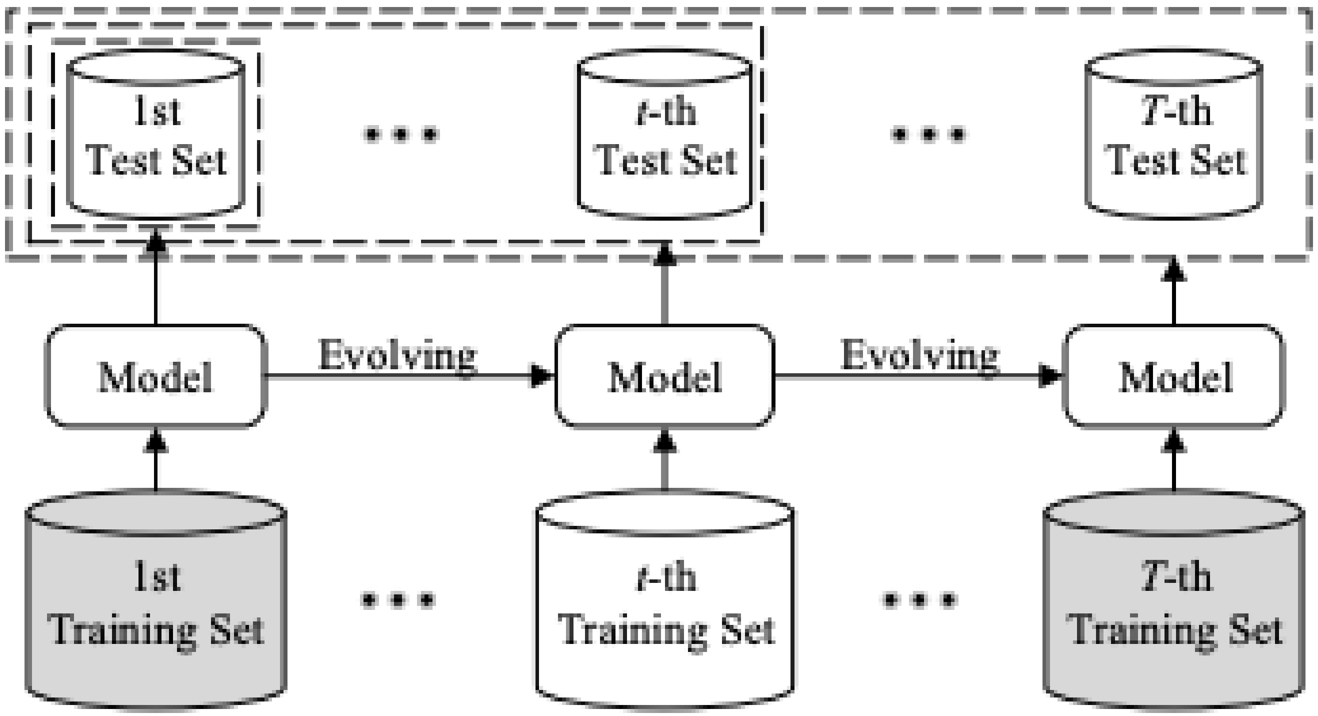


Fig. 1. Illustration of continual learning for BIQA. The grey cylinders denote the inaccessibility of previous and future training data. During testing, we use all previous and the current test sets to evaluate the stability and plasticity of the learned BIQA model, respectively.

Online learning

- Learning to adapt
- Application
  - *Concept drift in data stream*
  - *Nonstationary scenario in generation*

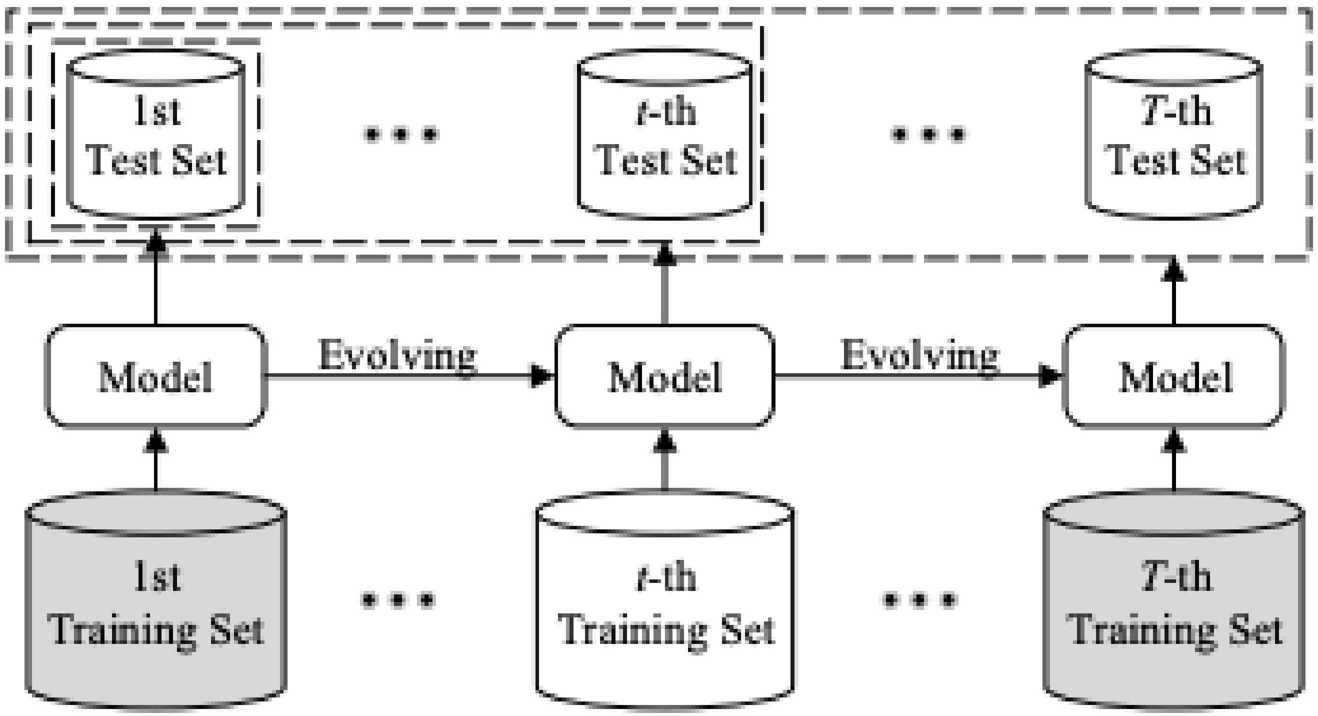


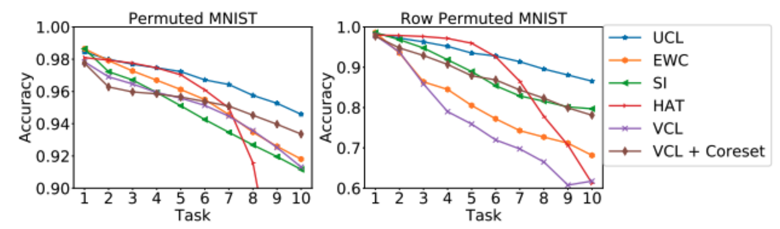
Fig. 1. Illustration of continual learning for BIQA. The grey cylinders denote the inaccessibility of previous and future training data. During testing, we use all previous and the current test sets to evaluate the stability and plasticity of the learned BIQA model, respectively.

Continual Learning on ASC

- Aspect Sentiment Classification[ASC,19 tasks]:
  - *classify the review sentences of 19 products into 4 sentiment(positive negative neutral). Each dataset represents a task from 4 sources.[1]*
- Permuted MNIST[3]:
  - *a different permutation of the pixels for the old task and the new task.*

Data source	Liu3domain			HL5domain					Ding9domain									SemEval14	
Task/ domain	Speaker	Router	Computer	Nokia6610	Nikon4300	Creative	CanonG3	ApexAD	CanonD500	Canon100	Diaper	Hitachi	Ipod	Linksys	MicroMP3	Nokia6600	Norton	Restaurant	Laptop
Train	352	245	283	271	162	677	228	343	118	175	191	212	153	176	484	362	194	3452	2163
Val.	44	31	35	34	20	85	29	43	15	22	24	26	19	22	61	45	24	150	150
Test	44	31	36	34	21	85	29	43	15	22	24	27	20	23	61	46	25	1120	638

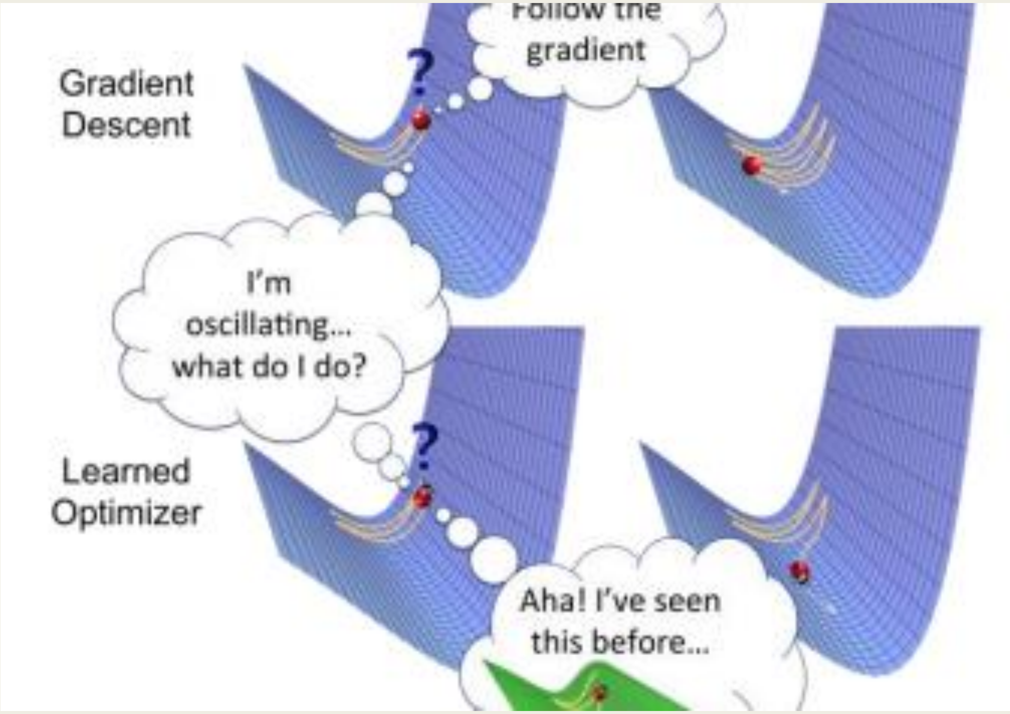
Table 1: Statistics of datasets for ASC. The datasets statistics for DSC and 20News have been described in the text. More detailed data statistics are given in *Supplementary*.

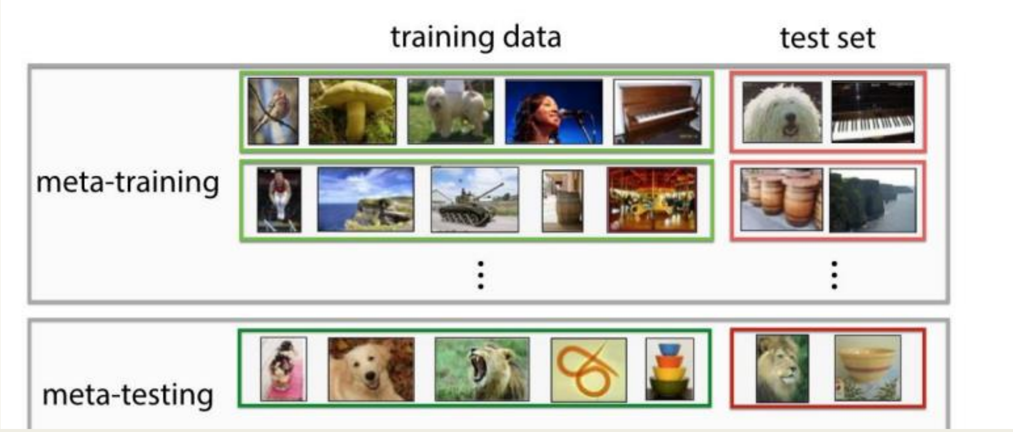


[1] Achieving Forgetting Prevention and Knowledge Transfer in Continual Learning NeurIPS 2021
[2] Uncertainty-based Continual Learning with Adaptive Regularization NeurIPS 2019
[3] An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks 2015
[4] Task-Specific Normalization for Continual Learning of Blind Image Quality Models

Meta-Learning

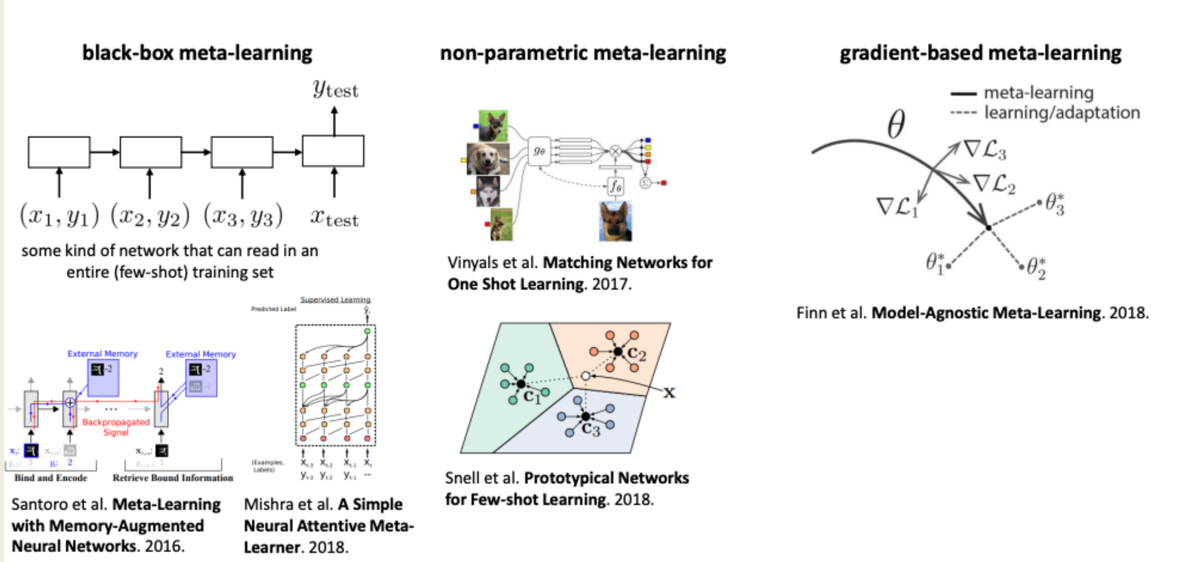
- **Learning to learn**
- If you've learned 100 tasks already, can you figure out how to learn more efficiently?
- Now having multiple tasks is a huge advantage!
- In practice, very closely related to multi-task learning





[1] Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks ICML 2017

META-LEARNING METHODS



Meta Learning in few-shot Classification

- **MiniImagenet:** In total, 100 classes are divided into 64, 16, and 20 classes respectively for sampling tasks for meta-training, meta-validation, and meta-test.[1]
- Provide a paradigm to learn new concepts rapidly from little data.[1]

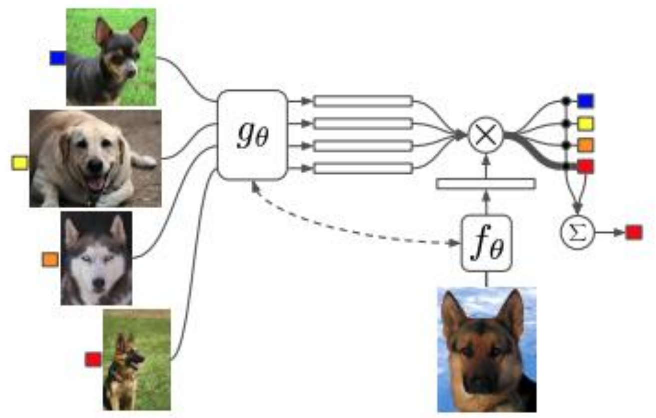
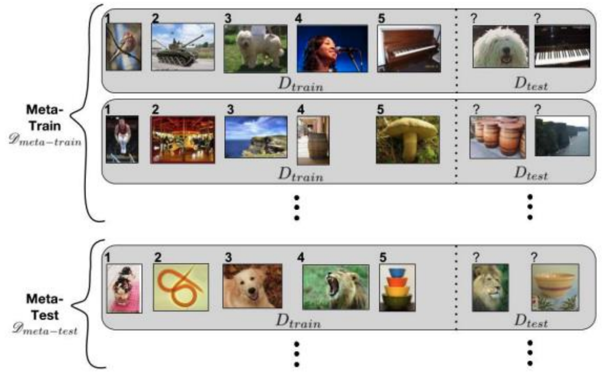
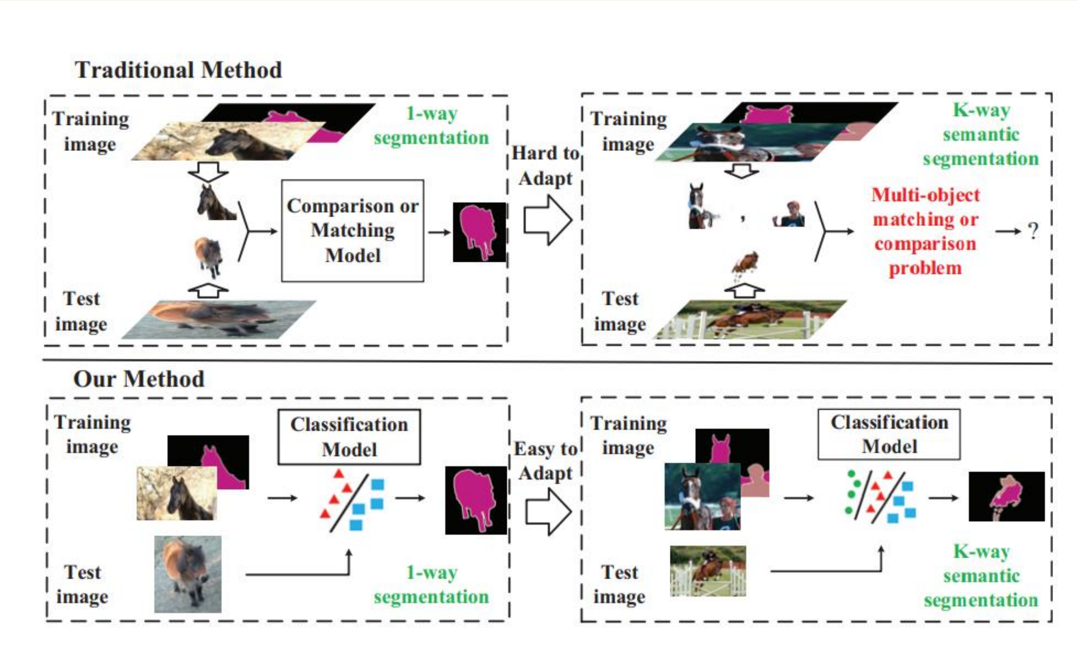
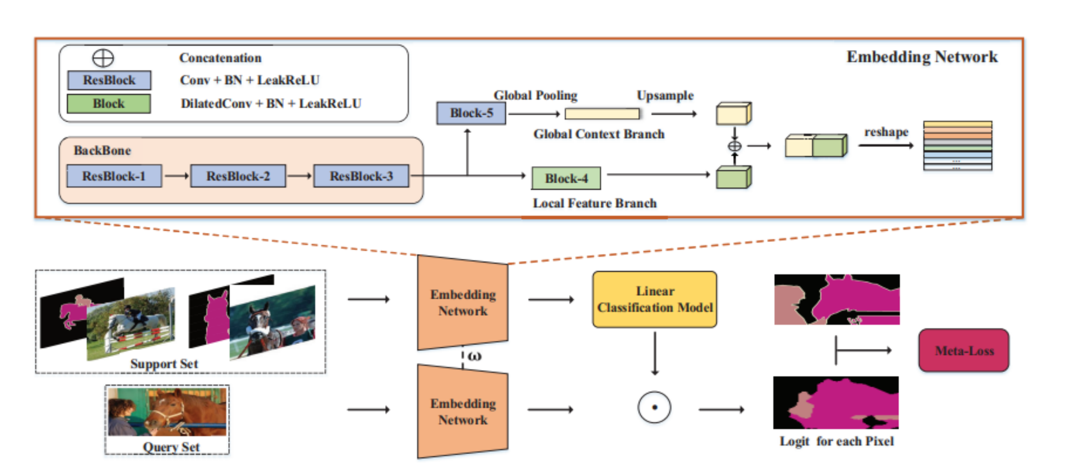


Figure 1: Matching Networks architecture



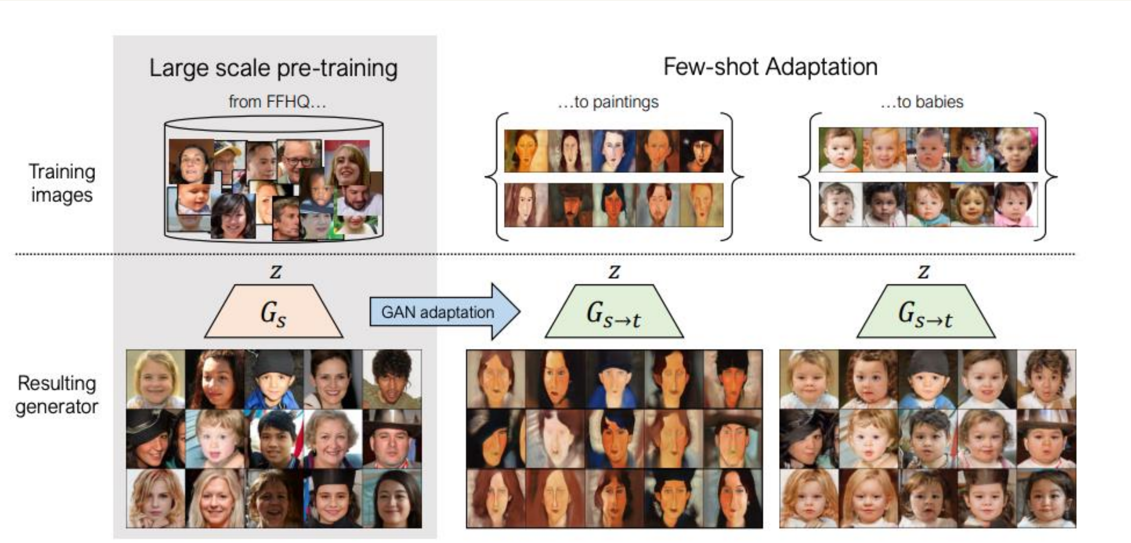
[1] Matching Networks for One Shot Learning, NIPS 2016





META LEARNING IN FEW-SHOT SEGMENTATION

[1] Differentiable Meta-Learning Model for Few-Shot Semantic Segmentation, AAAI2020

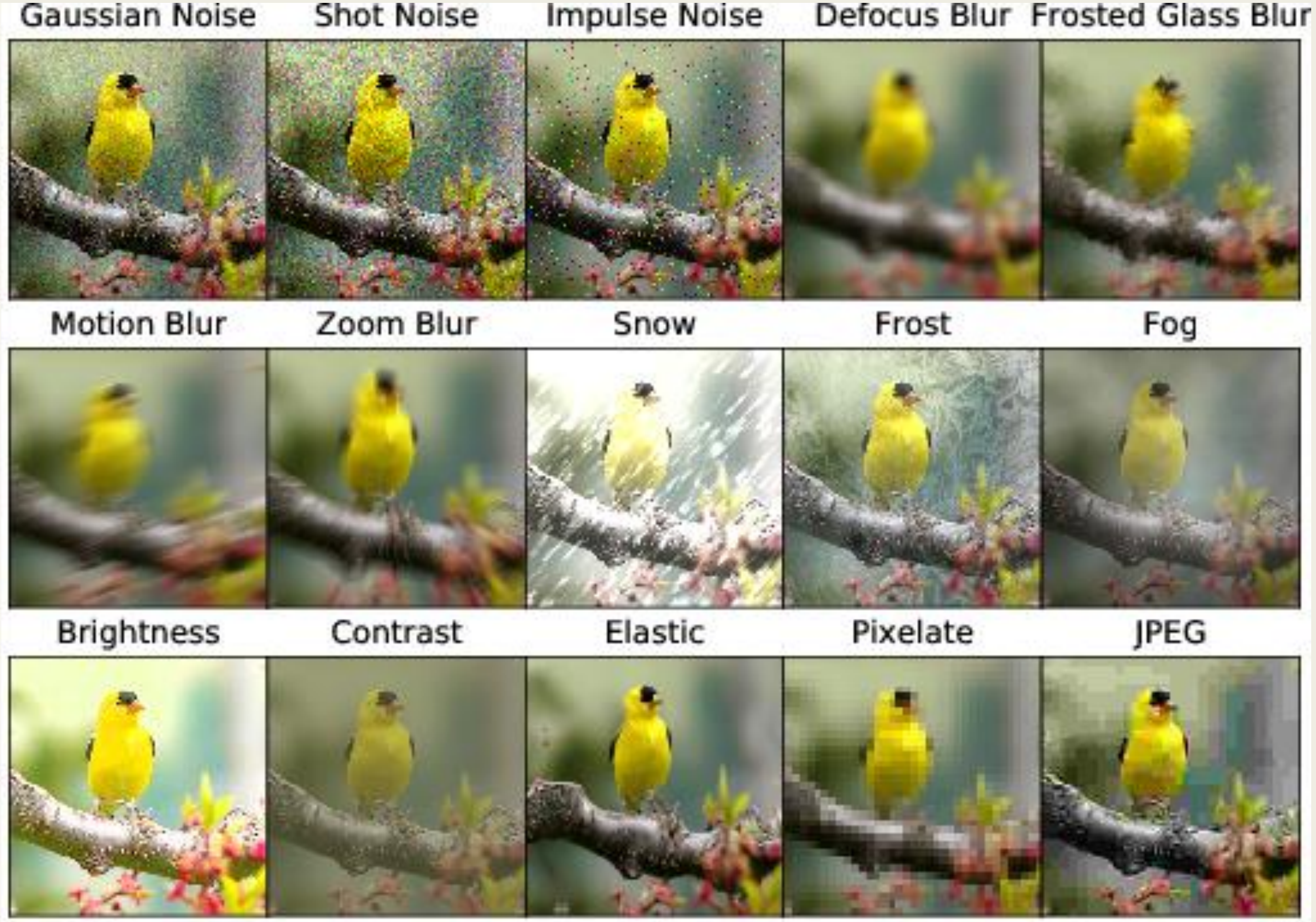


META LEARNING IN FEW-SHOT GENERATION

[1] Few-shot Image Generation via Cross-domain Correspondence CVPR2022

Test Time Adaptation

- Transmission
- Privacy



[1] Improving robustness against common corruptions by covariate shift adaptation (Neurips2020)
[2] Test-Time Training with Self-Supervision for Generalization under Distribution Shifts (ICML2020)

Test Time Adaptation

- Corruption
  - Noise： Gaussian, Shot, Impulse
  - Blur: Defocus, Glass, Motion, Zoom
  - Weather: Snow, Frost, Fog, Bright
  - Digital: Contrast, Elastic, Pixel, JPEG

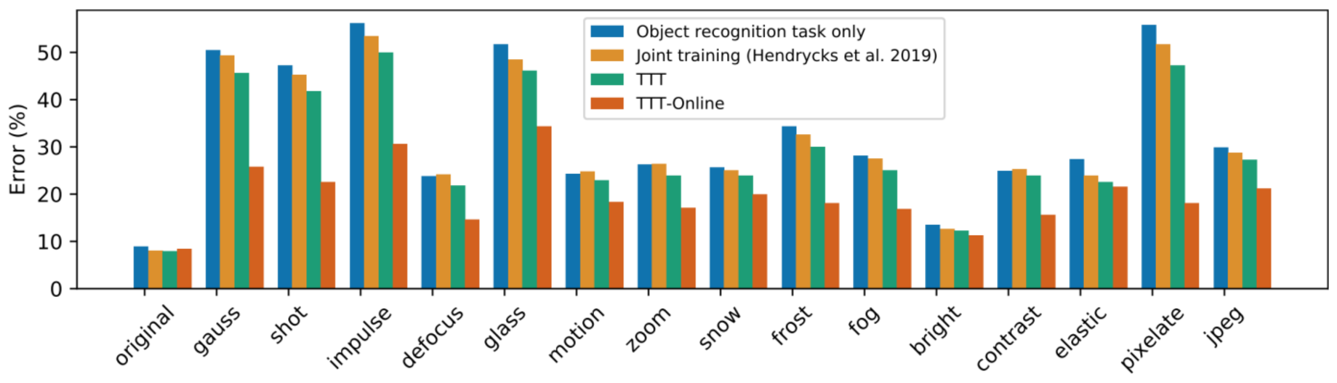


Figure 1. **Test error (%) on CIFAR-10-C with level 5 corruptions.** We compare our approaches, Test-Time Training (TTT) and its online version (TTT-Online), with two baselines: object recognition without self-supervision, and joint training with self-supervision but keeping the model fixed at test time. TTT improves over the baselines and TTT-Online improves even further.

[1] Improving robustness against common corruptions by covariate shift adaptation (Neurips2020)
[2] Test-Time Training with Self-Supervision for Generalization under Distribution Shifts (ICML2020)

THANKS