Few-shot Learning from Meta-Learning, Statistical Understanding to Applications

<u>Part I: Introduction</u>

Da Li

Samsung AI Centre, Cambridge, UK

CVPR 2023

- Introduction
  - Background
  - Conventional FSL
  - FSL lately …

- Introduction
  - Background
  - Conventional FSL
  - FSL lately …

Can you guess the category?

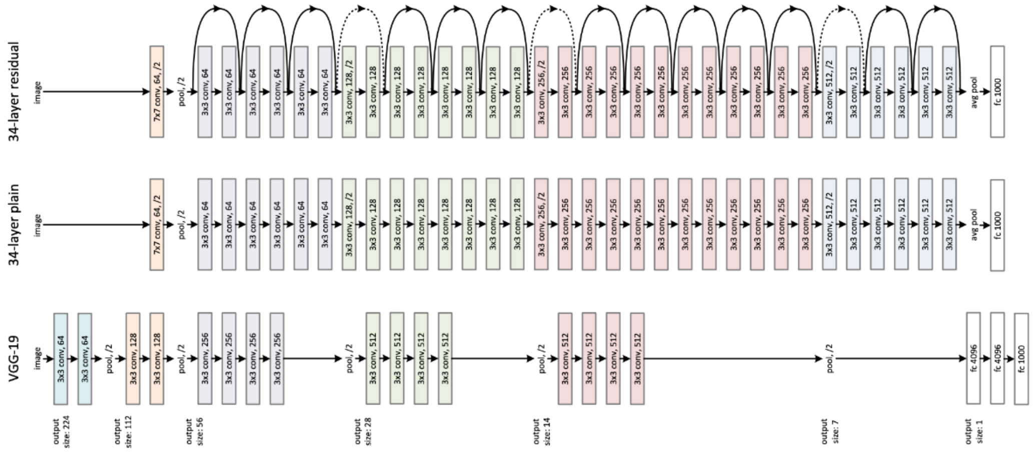
- Human is good at visual FSL.

echidna **porcupine**

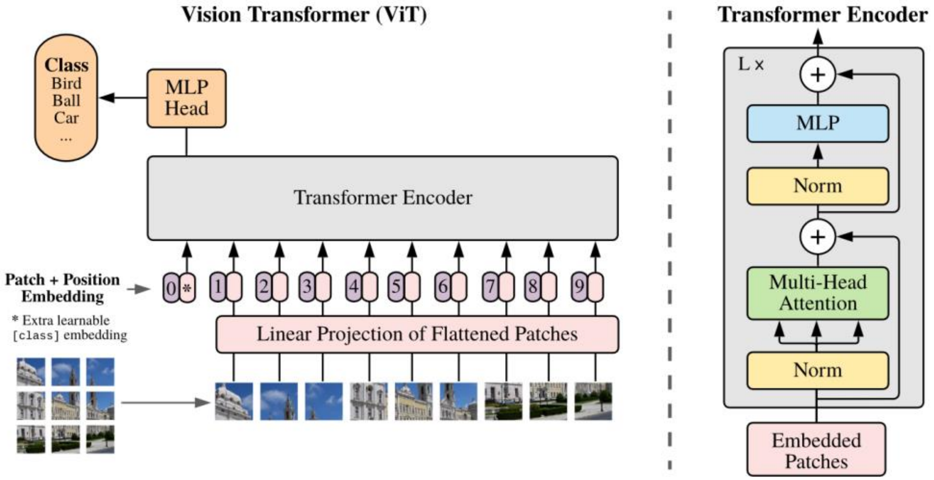
??

Motivation

- Why do we need few shot learning?



VGG, ResNet

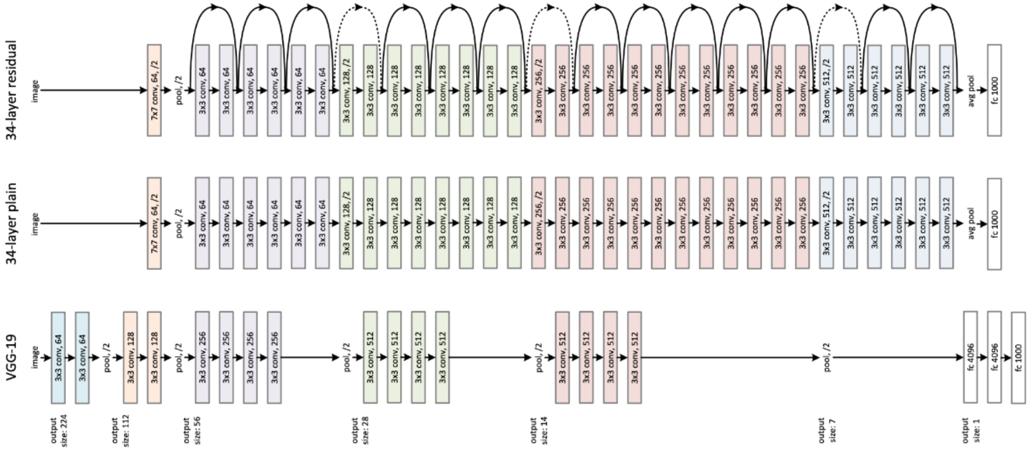


ViT

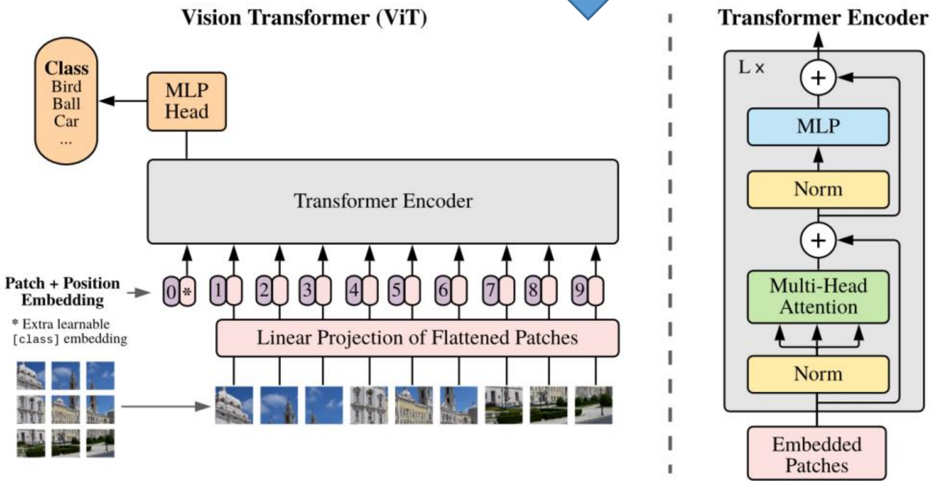
Dosovitskiy et al, AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE, in ICLR 2021.
He et al, Deep Residual Learning for Image Recognition, in CVPR 2016.
Simonyan and Zisserman, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGERECOGNITION." in ICLR 2015.

Motivation

- Why do we need few shot learning?



VGG, ResNet

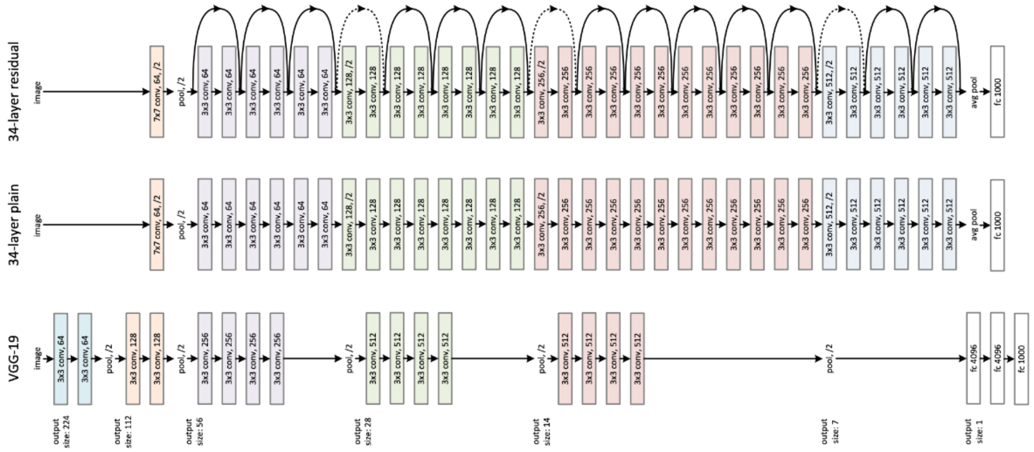


ViT

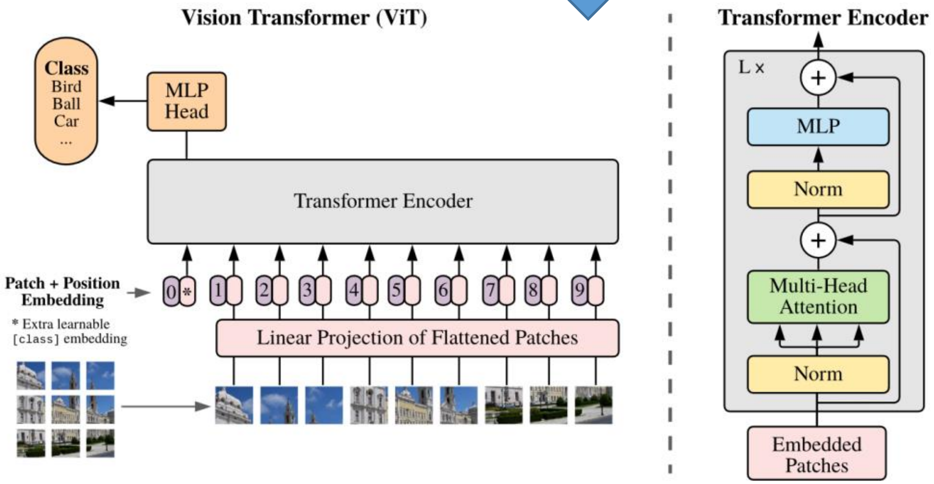
Dosovitskiy et al, AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE, in ICLR 2021.
He et al, Deep Residual Learning for Image Recognition, in CVPR 2016.
Simonyan and Zisserman, VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGERECOGNITION." in ICLR 2015.

Motivation

- Why do we need few shot learning?



VGG, ResNet



ViT

$$\forall \mathcal{D},\ \mathbf{P}_{S\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}_0(S)} \epsilon(h) \leq C\frac{\mathrm{VC}(\mathcal{H})+\ln\frac{1}{\delta}}{m}\right] \geq 1-\delta$$

Valle-Pérez et al, Generalization bounds for deep learning, in arXiv 2020.

Distinction to related topics

- Domain adaptation
- Long-tailed recognition
- Zero-shot learning
- … and open set recognition

Adapt source domain model to unlabelled target domain.

Distinction to related topics

- Domain adaptation
- Long-tailed recognition
- Zero-shot learning
- … and open set recognition

Train a model with imbalanced data of different classes.

Distinction to related topics

- Domain adaptation
- Long-tailed recognition
- Zero-shot learning
- … and open set recognition

Enable a pre-trained model to recognize a novel class without using any training image of that class.

Distinction to related topics

- Domain adaptation
- Long-tailed recognition
- Zero-shot learning
- … and open set recognition

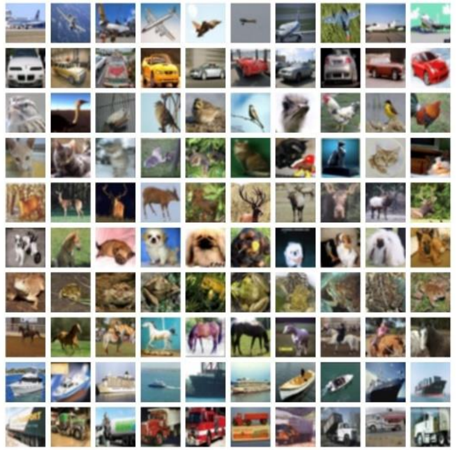
Distinguish between known and unknown samples during inference.

Benchmarks

Omniglot (Lake et al)

1,623 characters from 50 alphabets, while each character has 20 images of size 28 x 28.
1,200 characters for training and the rest 423 for testing

FC100 (Oreshkin et al)



CIFAR-100 for few-shot learning.
60, 20, 20 classes for traing, val and test.
600 images of size 32 × 32 per class

mini-ImageNet (Vinyals et al)

100 classes randomly selected from ImageNet and each class contains 600 images with the size of 84 × 84.
64 classes for training, 16 for validation and 20 for testing.

Vinyals et al, Matching Networks for One Shot Learning, in NeurIPS 2016.
Oreshkin et al, TADAM: Task dependent adaptive metric for improved few-shot learning, in NeurIPS 2018.
Lake et al"The Omniglot challenge: a 3-year progress report." in Current Opinion in Behavioral Sciences, 2019.

Benchmarks

Omniglot (Lake et al)

Meta-train

1,623 characters from 50 alphabets, while each character has 20 images of size 28 x 28.

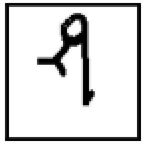
Base classes

1,200 characters for training and the rest 423 for testing

Novel classes

Episode sampling

Meta-test



…

K way N shot:
E.g. 20 way 1 shot

Lake et al"The Omniglot challenge: a 3-year progress report." in Current Opinion in Behavioral Sciences, 2019.

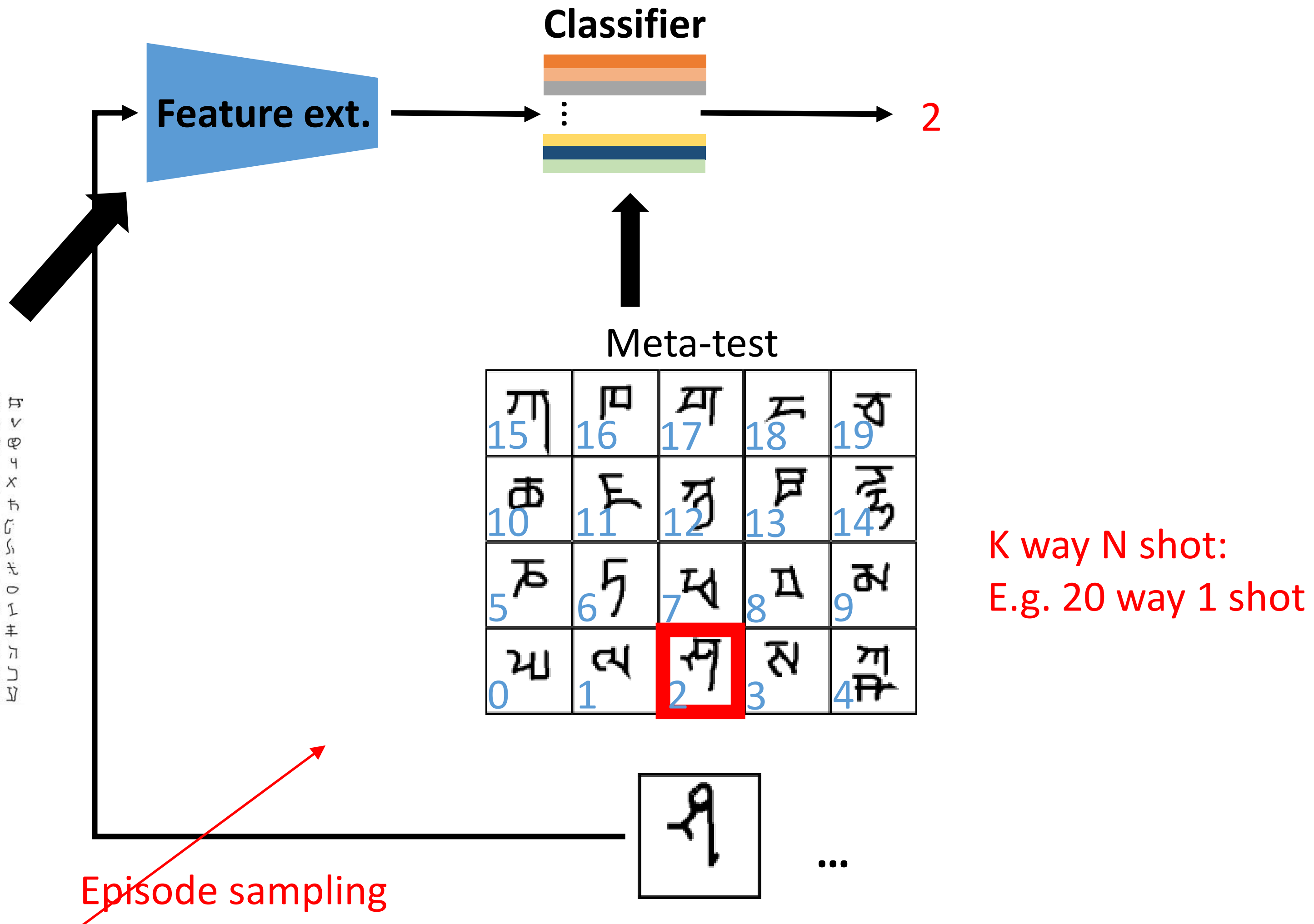
Benchmarks

Omniglot (Lake et al)

Meta-train

1,623 characters from 50 alphabets, while each character has 20 images of size 28 x 28.

Base classes 1,200 characters for training and the rest 423 for testing Novel classes



Lake et al"The Omniglot challenge: a 3-year progress report." in Current Opinion in Behavioral Sciences, 2019.

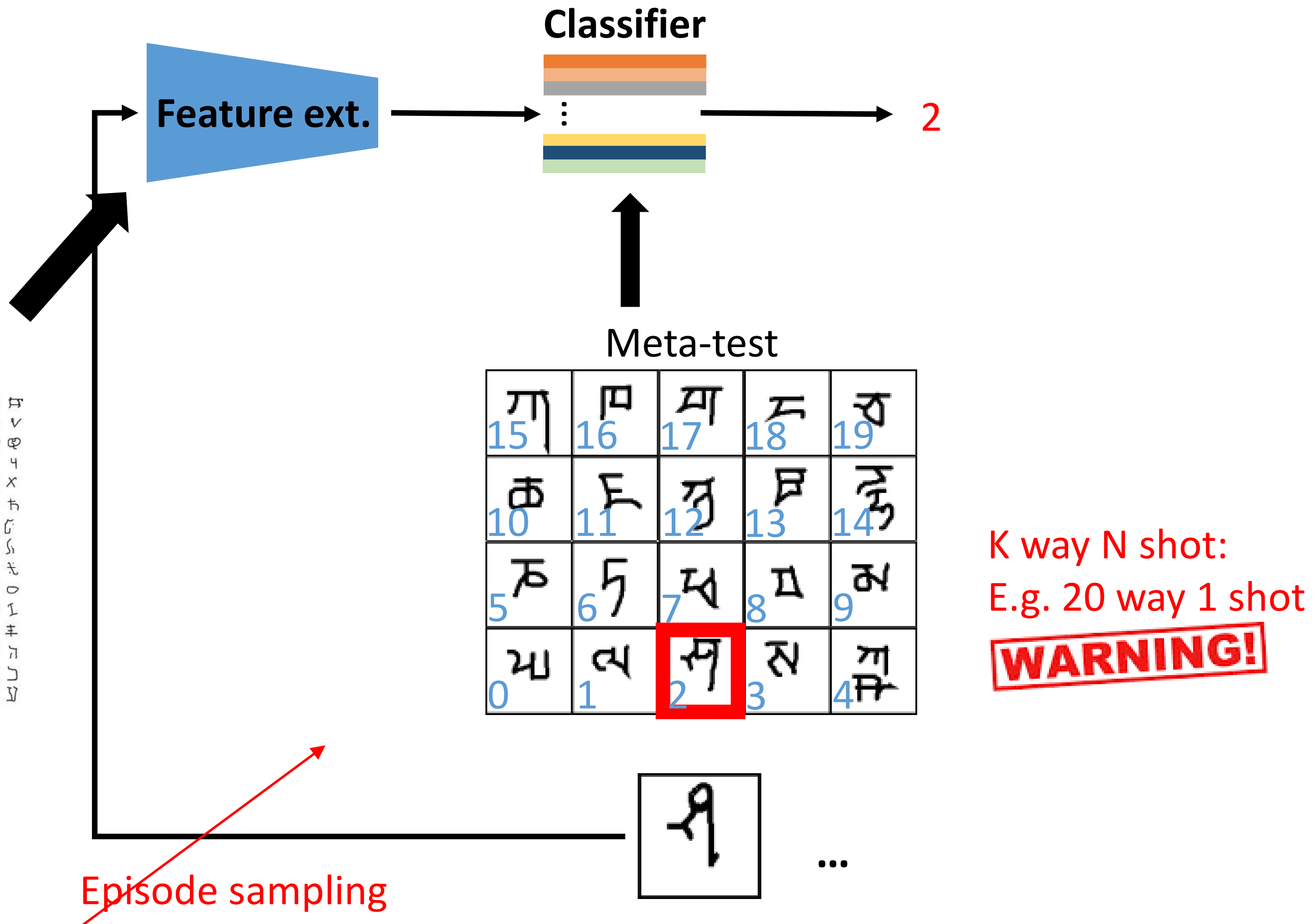
Benchmarks

Omniglot (Lake et al)

Meta-train

1,623 characters from 50 alphabets, while each character has 20 images of size 28 x 28.

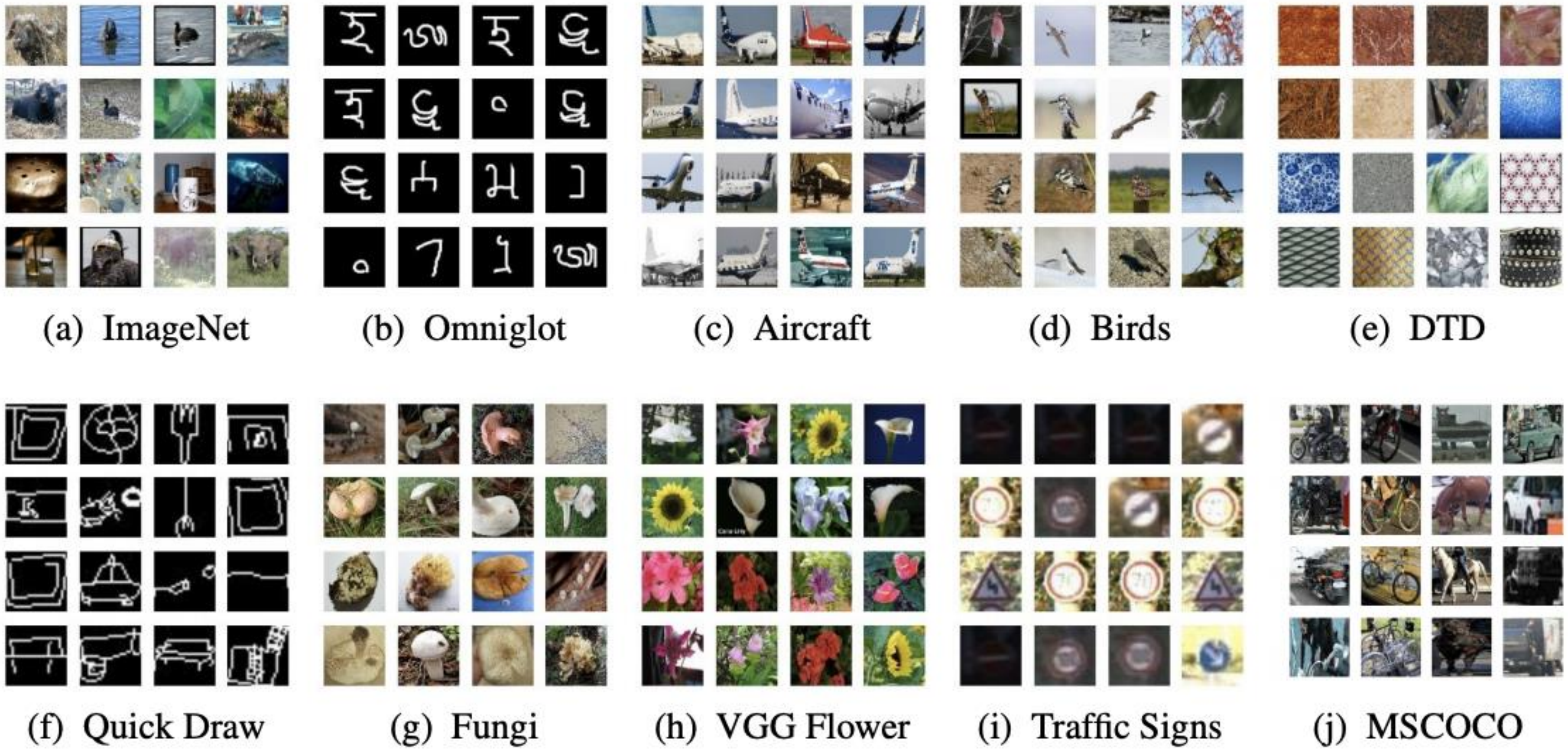
Base classes 1,200 characters for training and the rest 423 for testing Novel classes



Lake et al"The Omniglot challenge: a 3-year progress report." in Current Opinion in Behavioral Sciences, 2019.

Benchmarks

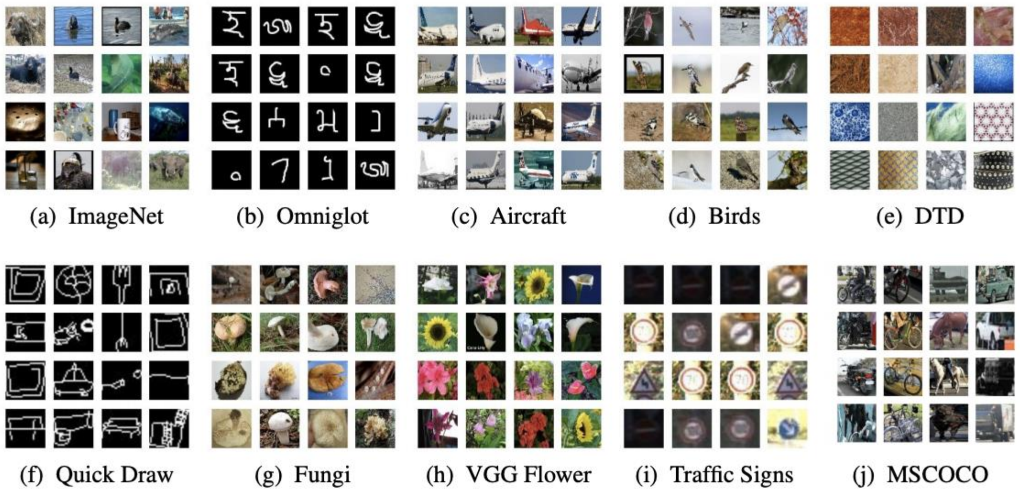
Meta-Dataset



Triantafillou et al, Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, in ICLR 2020.

Benchmarks

Meta-Dataset

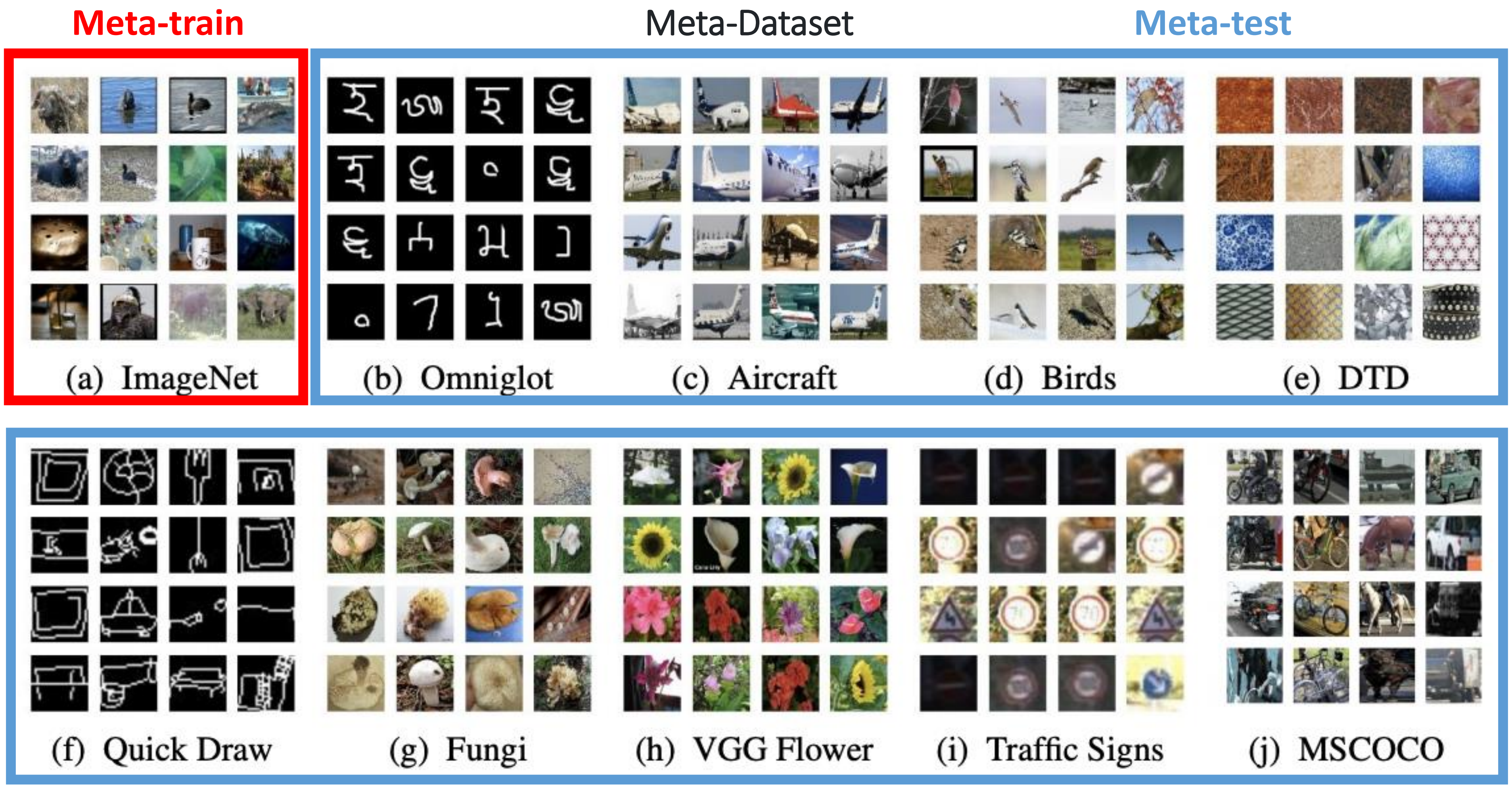


- ILSVRC-2012 (the ImageNet dataset, consisting of natural images with 1000 categories)
- Omniglot (hand-written characters, 1623 classes)
- Aircraft (dataset of aircraft images, 100 classes)
- CUB-200-2011 (dataset of Birds, 200 classes)
- Describable Textures (different kinds of texture images with 43 categories)
- Quick Draw (black and white sketches of 345 different categories)
- Fungi (a large dataset of mushrooms with 1500 categories)
- VGG Flower (dataset of flower images with 102 categories),
- Traffic Signs (German traffic sign images with 43 classes)
- MSCOCO (images collected from Flickr, 80 classes).

Triantafillou et al, Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, in ICLR 2020.

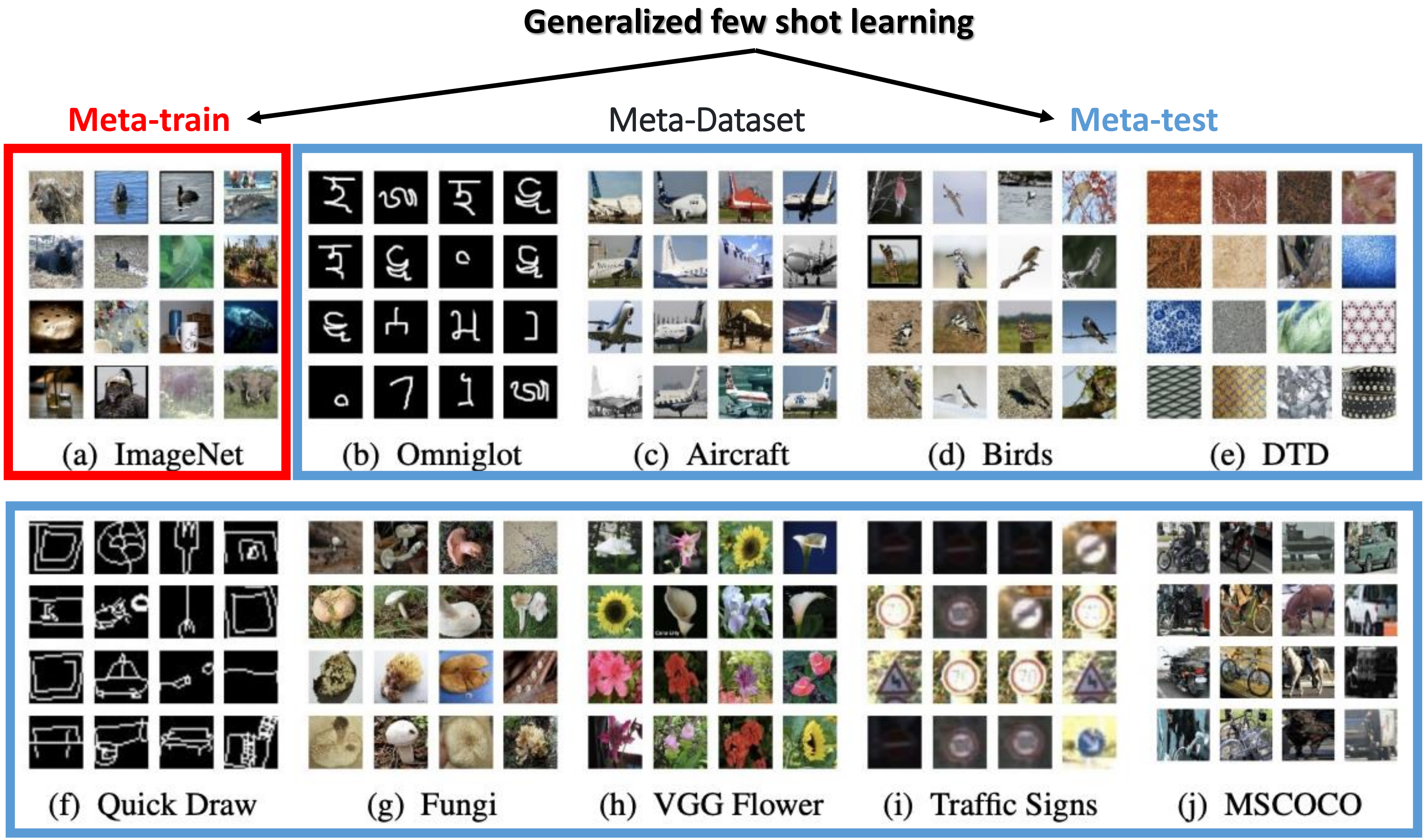
Benchmarks

Cross-domain few shot learning



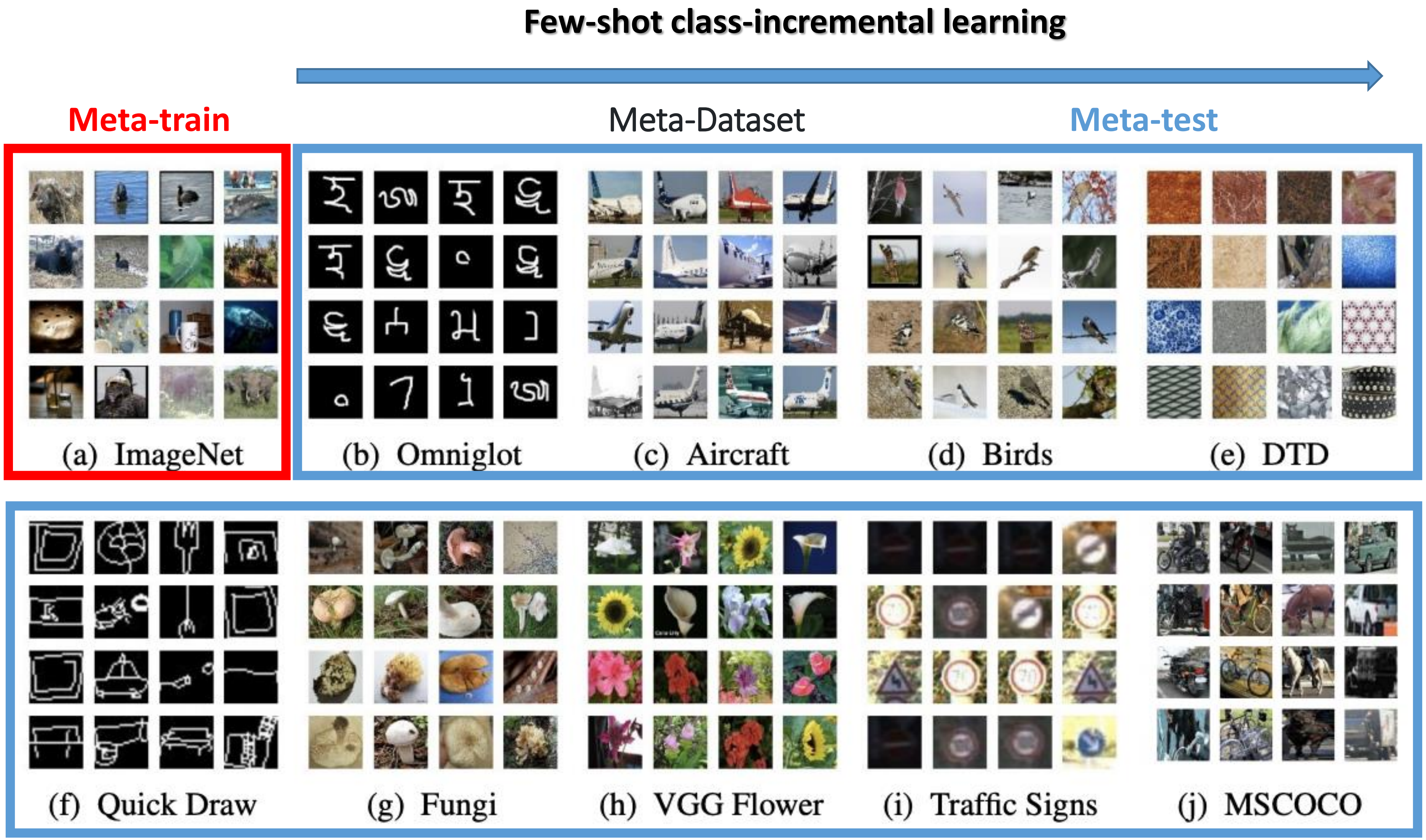
Triantafillou et al, Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, in ICLR 2020.

Benchmarks



Triantafillou et al, Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, in ICLR 2020.

Benchmarks



Triantafillou et al, Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples, in ICLR 2020.

Benchmarks

Benchmarks

Add a Result

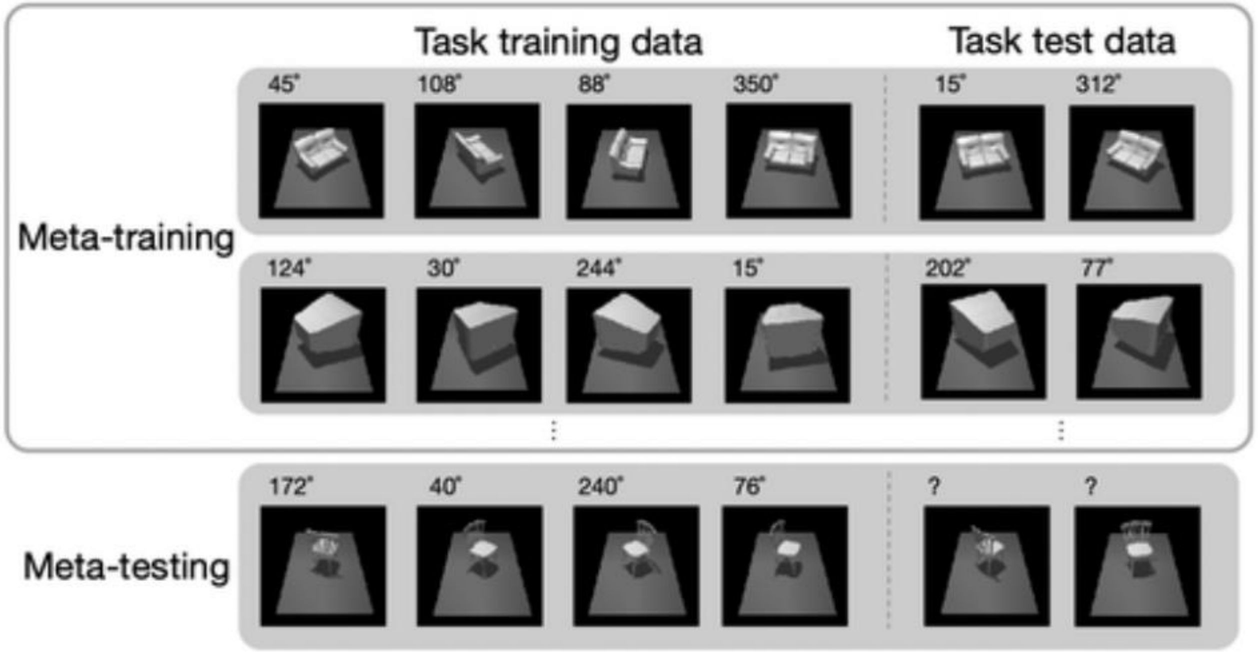
These leaderboards are used to track progress in Few-Shot Image Classification

Trend	Dataset	Best Model	Paper	Code	Compare
	Mini-Imagenet 5-way (1-shot)	P>M>F (P=DINO-ViT-base, M=ProtoNet)			See all
	Mini-Imagenet 5-way (5-shot)	P>M>F (P=DINO-ViT-base, M=ProtoNet)			See all
	Tiered ImageNet 5-way (5-shot)	TRIDENT			See all
	Tiered ImageNet 5-way (1-shot)	TRIDENT			See all
	CIFAR-FS 5-way (5-shot)	PT+MAP+SF+SOT (transductive)			See all
	CIFAR-FS 5-way (1-shot)	PT+MAP+SF+SOT (transductive)			See all
	CUB 200 5-way 1-shot	PT+MAP+SF+SOT (transductive)			See all
	CUB 200 5-way 5-shot	PT+MAP+SF+SOT (transductive)			See all

https://paperswithcode.com/task/few-shot-image-classification#benchmarks

Further tasks

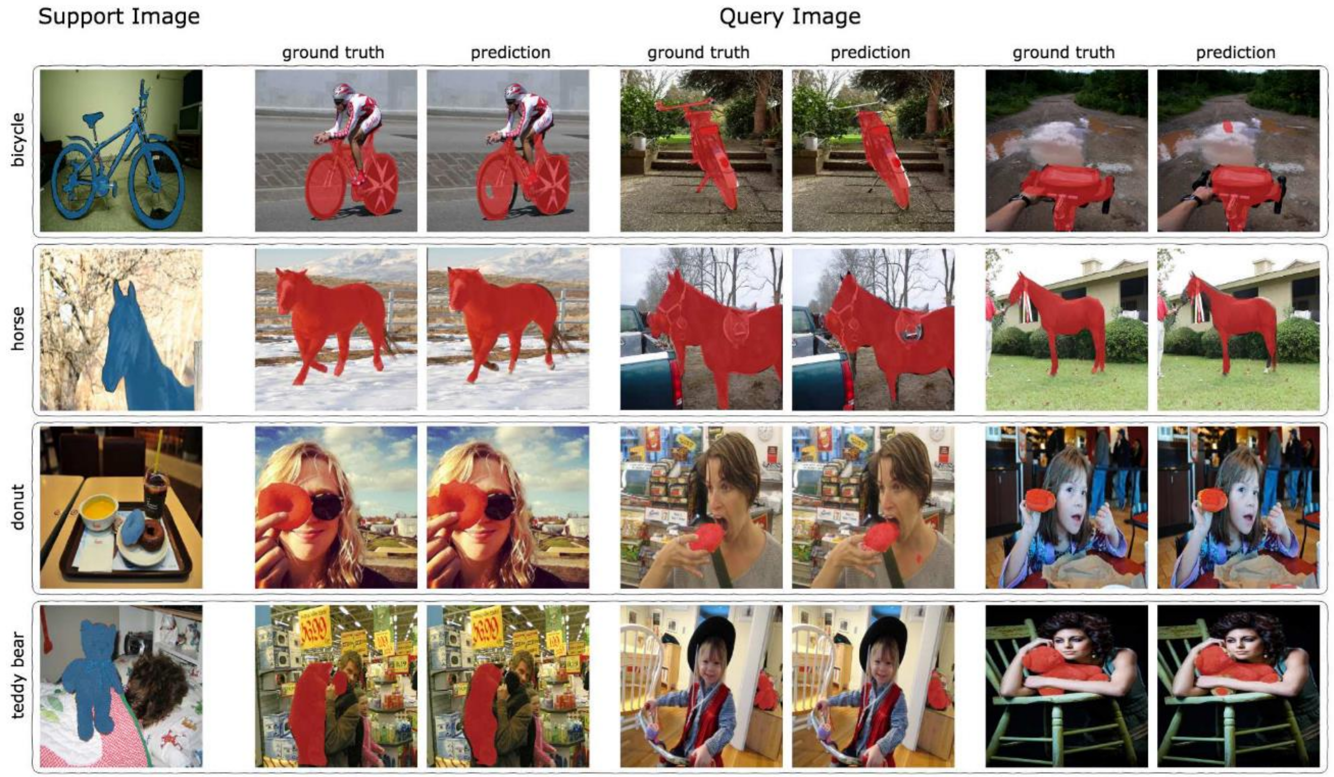
- Regression/Segmentation/Object detection/2D-to-3D/Reinforcement learning…



Yin et al, META-LEARNING WITHOUT MEMORIZATION, in ICLR 2020.

Further tasks

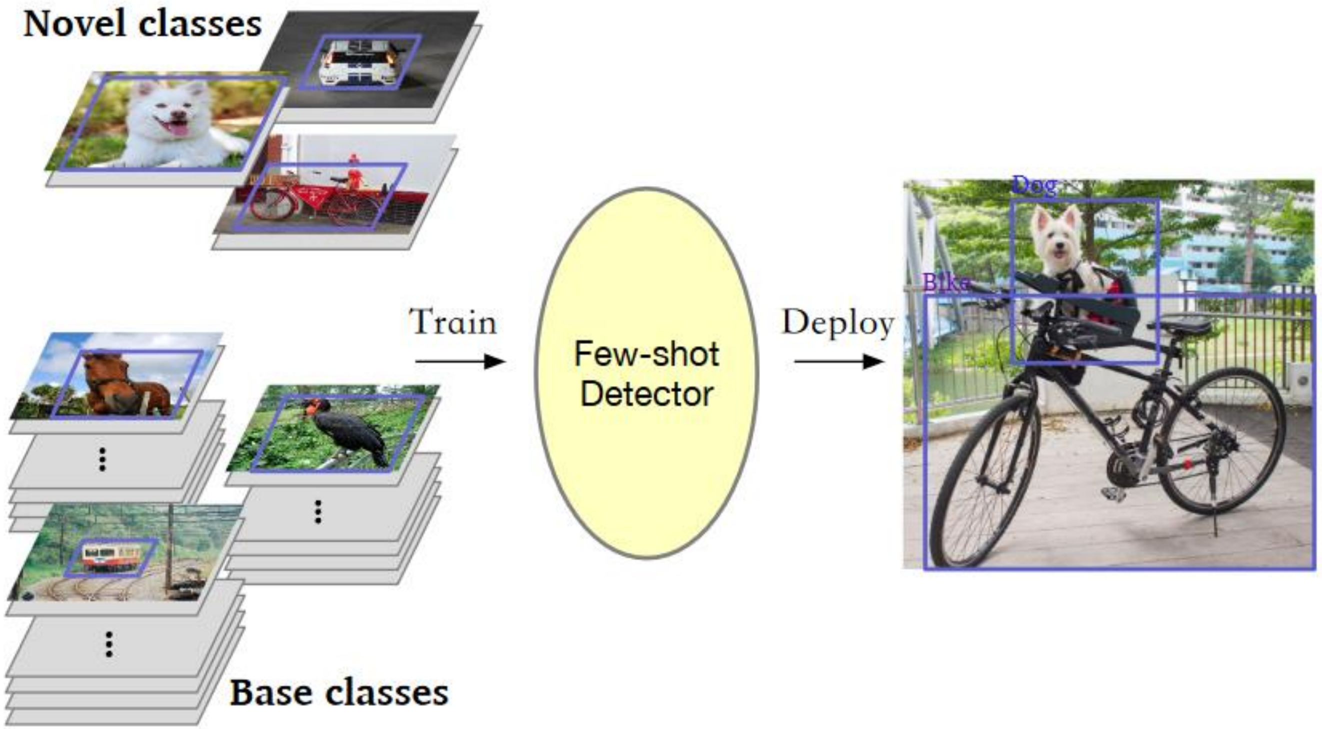
- Regression/**Segmentation**/Object detection/2D-to-3D/Reinforcement learning…



Wang et al, PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment, in ICCV 2019.

Further tasks

- Regression/Segmentation/Object detection/2D-to-3D/Reinforcement learning…



Kang et al, Few-shot Object Detection via Feature Reweighting, in ICCV 2019.

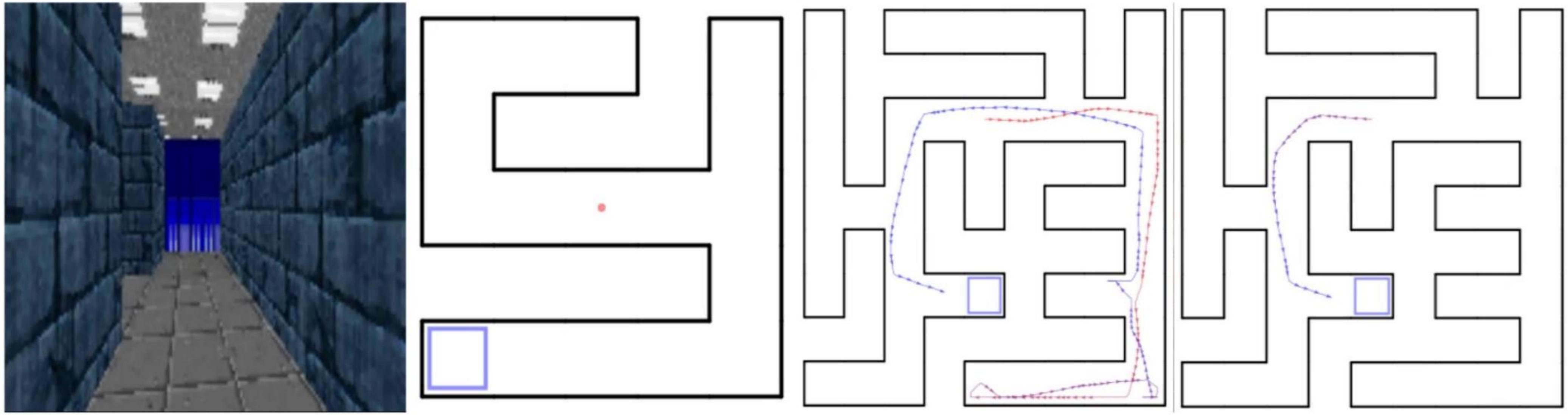
Further tasks

- Regression/Segmentation/Object detection/2D-to-3D/Reinforcement learning…

Mildenhall et al, Representing Scenes as Neural Radiance Fields for View Synthesis, in ECCV 2020.

Further tasks

- Regression/Segmentation/Object detection/2D-to-3D/Reinforcement learning…

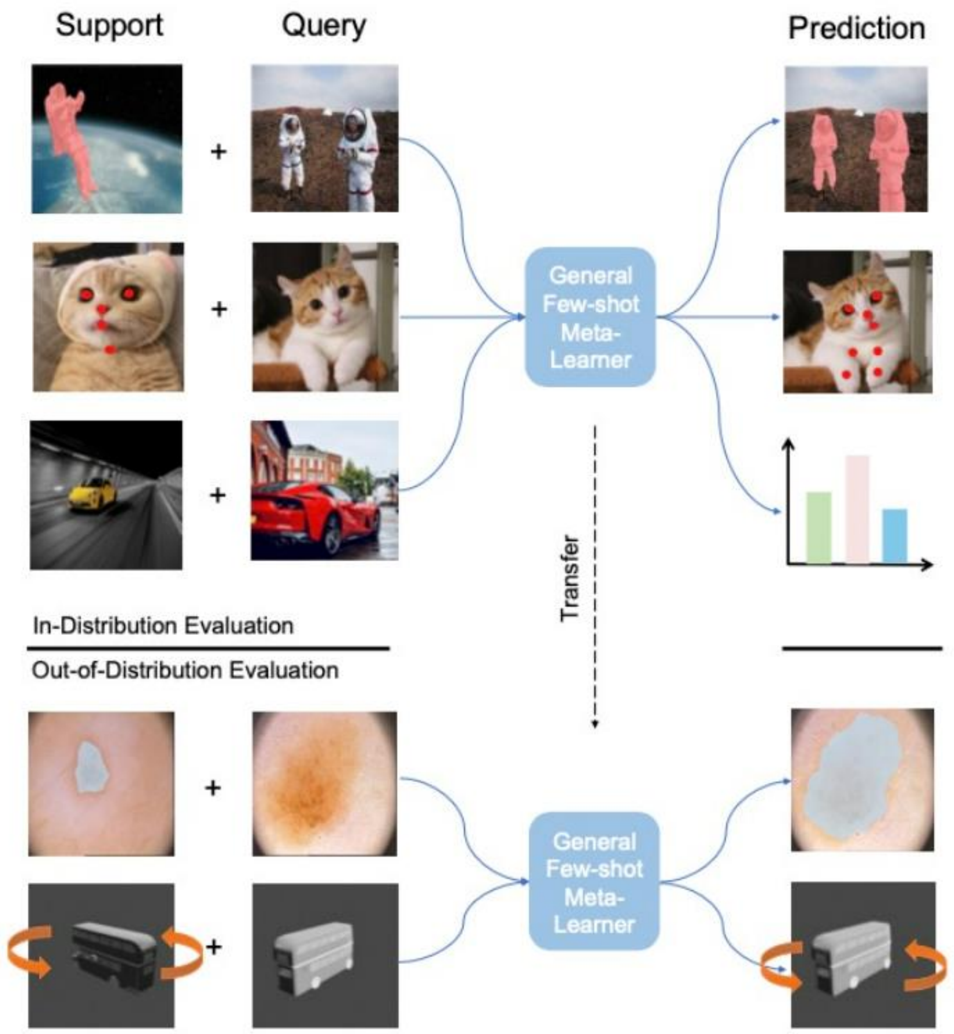


Mishra et al, A SIMPLE NEURAL ATTENTIVE META-LEARNER, in ICLR 2018.

General-purpose l2l

Meta Omnium: A Benchmark for General-Purpose Learning-to-Learn

https://github.com/edi-meta-learning/meta-omnium



Bohdal et al. in Meta Omnium: A Benchmark for General-Purpose Learning-to-Learn, in CVPR 2023.

- Introduction
  - Background
  - Conventional FSL
  - FSL lately …

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …

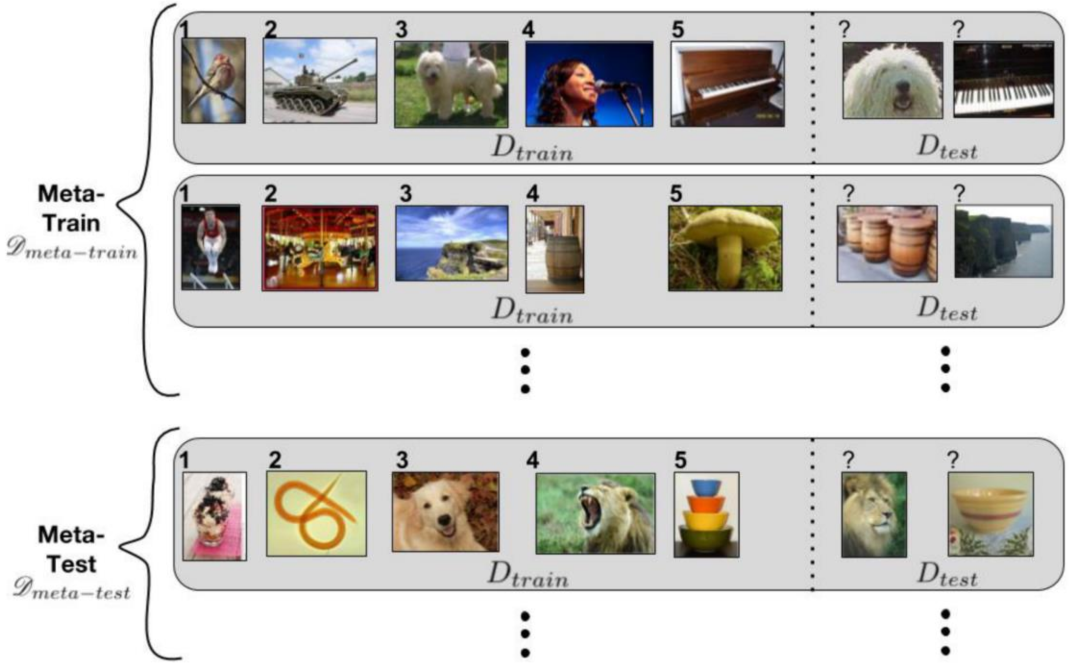
Hospedales et al, Meta-Learning in Neural Networks: A Survey, in PAMI 2020.

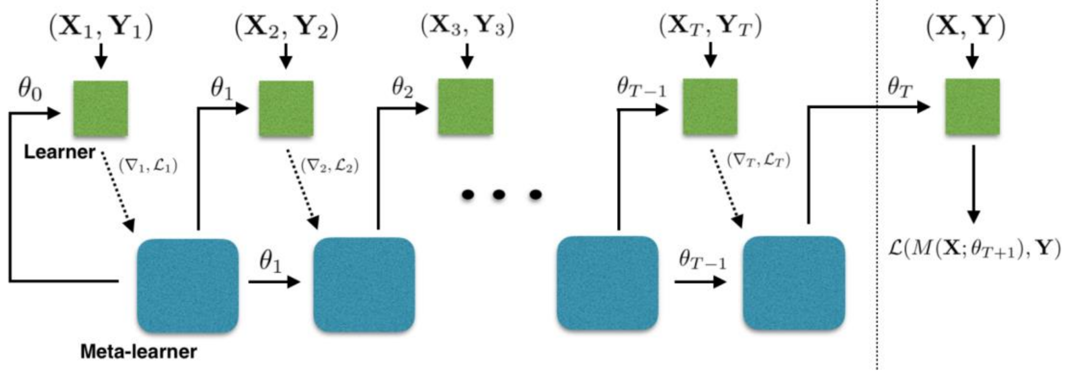
Conventional FSL

- Meta-learning methods
    - Optimizer (Ravi et al 2017)
    - Model initialization (Finn et al 2017)
    - Metric learning (Vinyals et al 2017, Snell et al 2017)
    - …
- Non meta-learning methods
    - Fine-tuning
    - Nearest neighbour (Wang et al 2019)
    - …

…

- Improve FSL by statistical tools
    - Sparsity (Wang et al)
    - Causality (Xu et al)
    - …





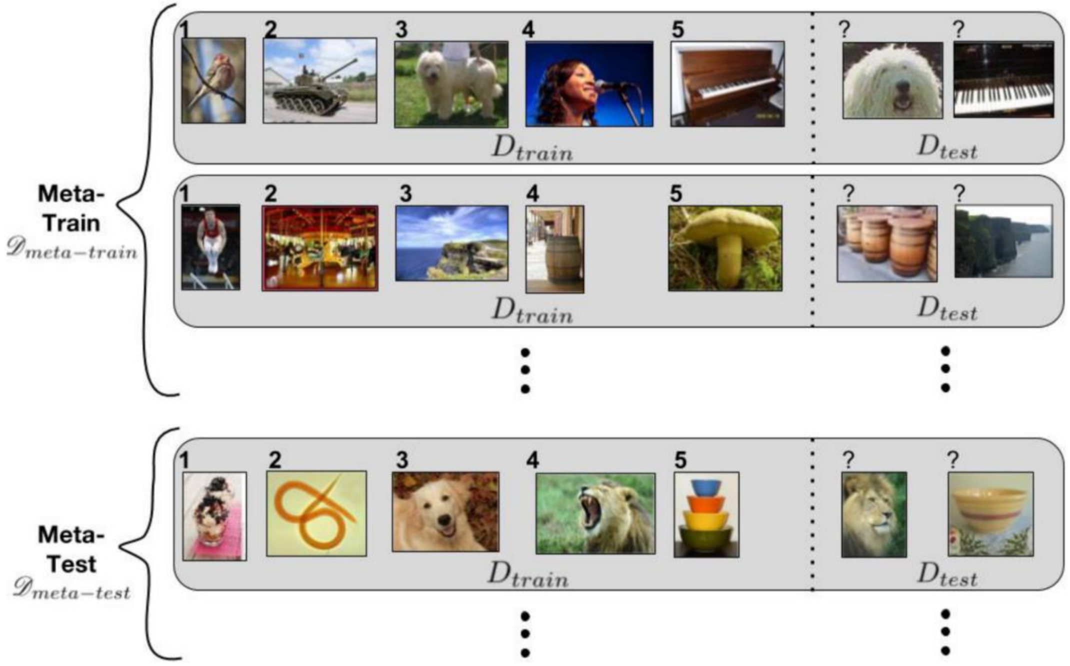
Ravi et al, OPTIMIZATION AS A MODEL FOR FEW-SHOT LEARNING, in ICLR 2017.

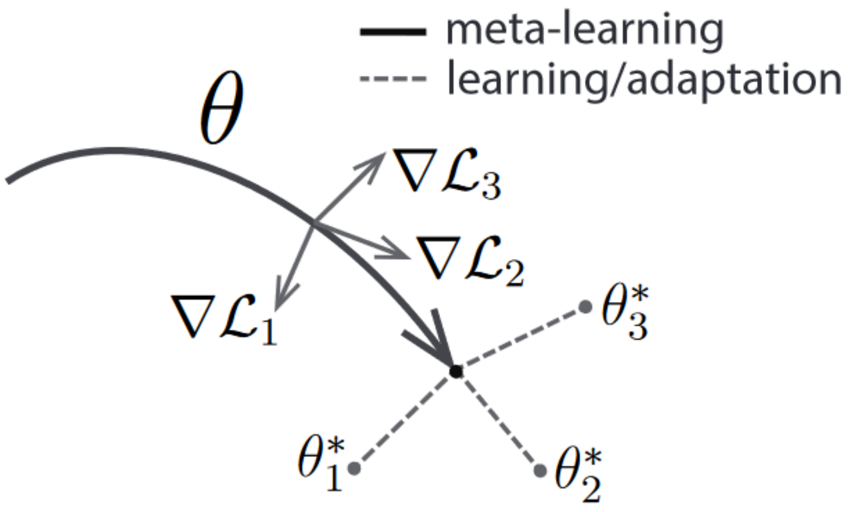
Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …





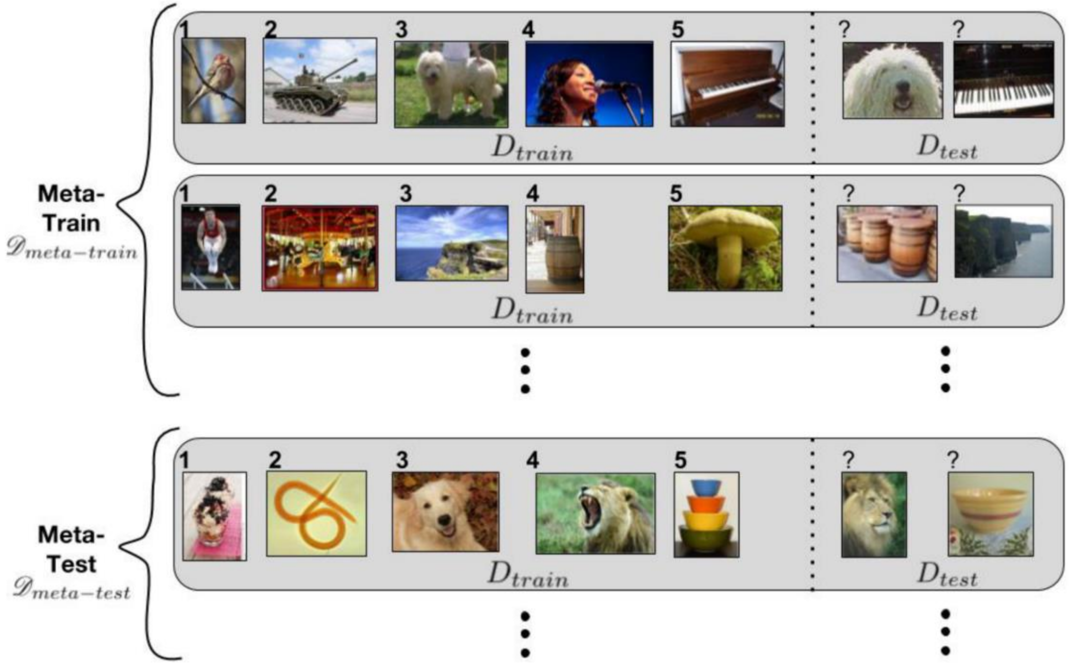
Finn et al, Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, in ICML 2017.

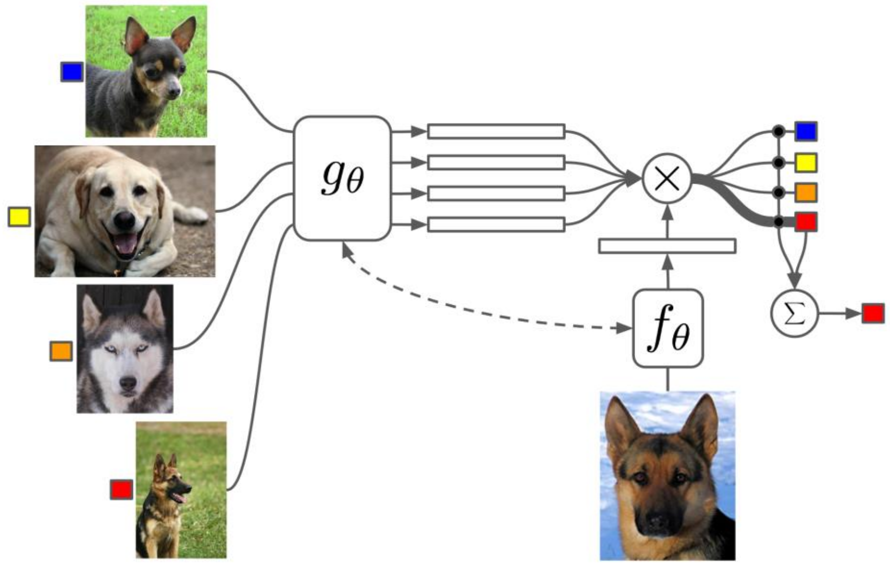
Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2016, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …





Snell et al, Prototypical Networks for Few-shot Learning, in NeurIPS 2017.
Vinyals et al, Matching Networks for One Shot Learning, in NeurIPS 2016.

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

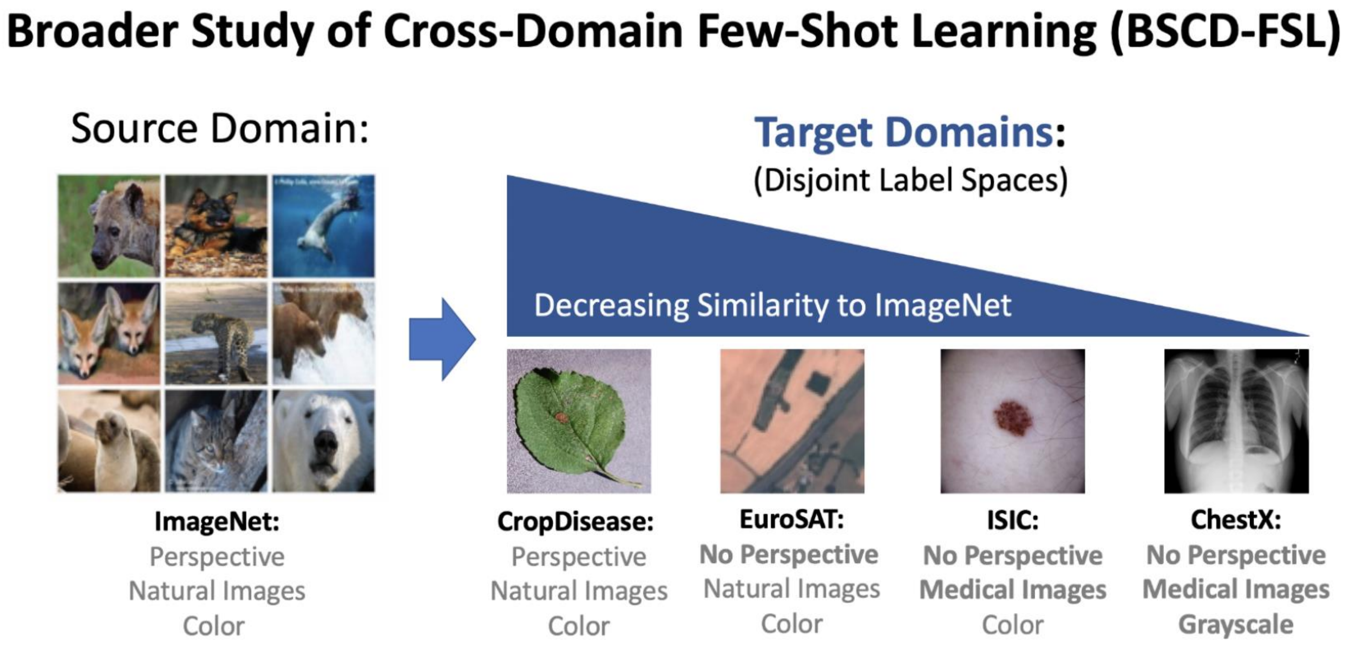
- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Sı
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

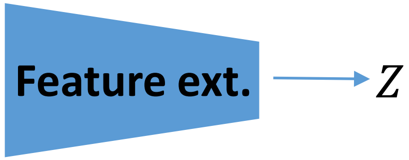
…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …



Guo et al. A Broader Study of Cross-Domain Few-Shot Learning, in ECCV 2020.

Conventional FSL



$$Z = \frac{Z - \bar{Z}}{\|Z - \bar{Z}\|_2}$$

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …

Approach	**Network**	**One shot**	**Five shots**
Reptile [23]$^\sharp$	Conv-4	48.97 ± 0.21	66.47 ± 0.21
ProtoNet [30]$^\sharp$	Conv-4	**53.31 ± 0.89**	**72.69 ± 0.74**
SimpleShot (UN)	Conv-4	33.12 ± 0.18	65.23 ± 0.18
SimpleShot (L2N)	Conv-4	50.21 ± 0.20	69.02 ± 0.18
SimpleShot (CL2N)	Conv-4	51.02 ± 0.20	68.98 ± 0.18
SimpleShot (UN)	ResNet-10	58.60 ± 0.22	79.99 ± 0.16
SimpleShot (L2N)	ResNet-10	64.58 ± 0.23	82.31 ± 0.16
SimpleShot (CL2N)	ResNet-10	65.37 ± 0.22	81.84 ± 0.16
SimpleShot (UN)	ResNet-18	62.69 ± 0.22	83.27 ± 0.16
SimpleShot (L2N)	ResNet-18	68.64 ± 0.22	84.47 ± 0.16
SimpleShot (CL2N)	ResNet-18	**69.09 ± 0.22**	**84.58 ± 0.16**
Meta SGD [18]†	WRN	62.95 ± 0.03	79.34 ± 0.06
LEO [29]	WRN	66.33 ± 0.05	81.44 ± 0.09
SimpleShot (UN)	WRN	63.85 ± 0.21	84.17 ± 0.15
SimpleShot (L2N)	WRN	66.86 ± 0.21	**85.50 ± 0.14**
SimpleShot (CL2N)	WRN	**69.75 ± 0.20**	85.31 ± 0.15
SimpleShot (UN)	MobileNet	63.65 ± 0.22	84.01 ± 0.16
SimpleShot (L2N)	MobileNet	68.66 ± 0.23	**85.43** ± 0.15
SimpleShot (CL2N)	MobileNet	**69.47 ± 0.22**	85.17 ± 0.15
SimpleShot (UN)	DenseNet	64.35 ± 0.23	85.69 ± 0.15
SimpleShot (L2N)	DenseNet	69.91 ± 0.22	86.42 ± 0.15
SimpleShot (CL2N)	DenseNet	**71.32 ± 0.22**	**86.66 ± 0.15**

Wang et al, SimpleShot: Revisiting Nearest-Neighbor Classification for Few-Shot Learning, in arXiv 2019.

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …

- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

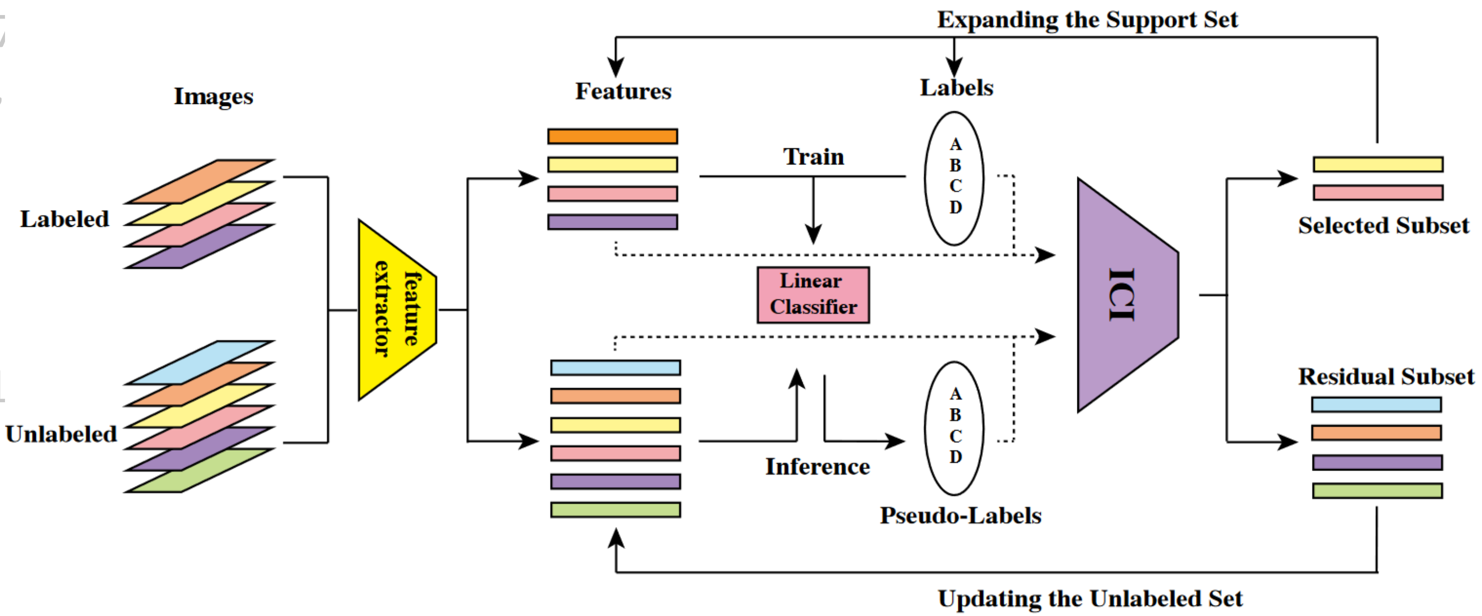
- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017
  - Metric learning (Vinyals et al 2017,
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 201
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …



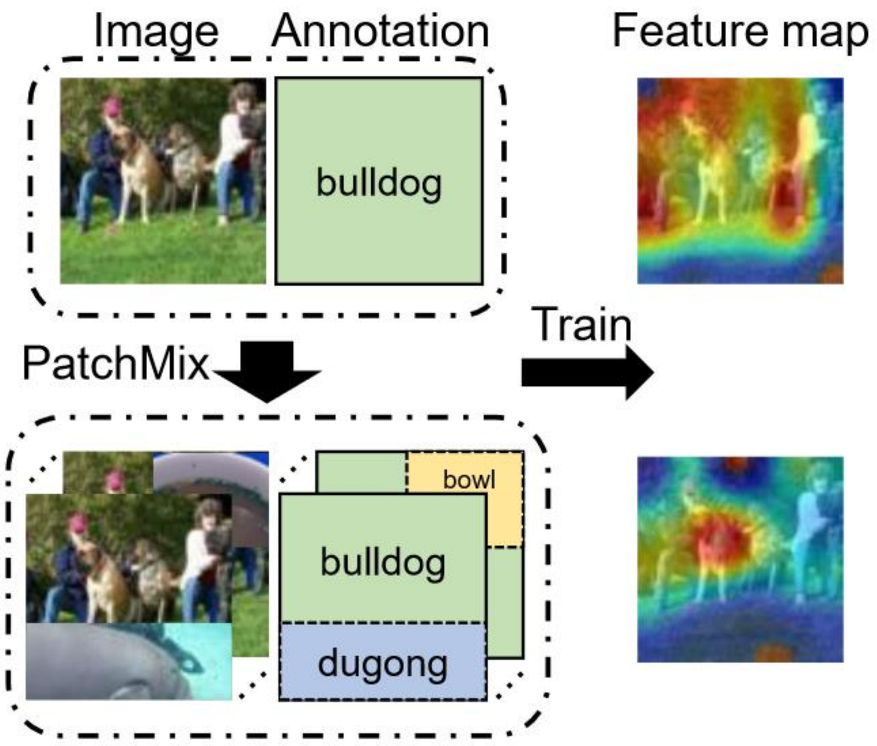
Wang et al, How to Trust Unlabeled Data? Instance Credibility Inference for Few-Shot Learning, in PAMI 2021.

Conventional FSL

- Meta-learning methods
  - Optimizer (Ravi et al 2017)
  - Model initialization (Finn et al 2017)
  - Metric learning (Vinyals et al 2017, Snell et al 2017)
  - …
- Non meta-learning methods
  - Fine-tuning
  - Nearest neighbour (Wang et al 2019)
  - …

…

- Improve FSL by statistical tools
  - Sparsity (Wang et al)
  - Causality (Xu et al)
  - …



Xu et al, PatchMix Augmentation to Identify Causal Features in Few-shot Learning, in PAMI 2022.

- Introduction
  - Background
  - Conventional FSL
  - FSL lately …

FSL lately …

- FSL with foundation models
- Visual (/language) in-context learning

FSL lately …

- FSL with foundation models
- Visual (/language) in-context learning

P>M>F

E.g. DINO > ProtoNet (PN) > Fine-tuning (FT)

	Training Configuration			Benchmark Results		
ID	Arch	Pre Train	MetaTr	MD	miniIN	CIFAR
0	ViT-small	DINO (IN1K)	-	67.4	97.0	79.8
1	ViT-small	DeiT (IN1K)	-	67.5	98.8	84.6
2	ResNet50	DINO (IN1K)	-	63.8	91.5	76.1
3	ResNet50	Sup. (IN1K)	-	62.4	96.4	82.3
4	ViT-small	DINO (IN1K)	PN	78.4	98.0	92.5
5	ViT-small	DEIT (IN1K)	PN	79.3	99.4	93.6
6	ViT-small	-	PN	52.8	49.1	59.8
7	ResNet50	DINO (IN1K)	PN	72.4	92.0	84.0
8	ResNet50	Sup. (IN1K)	PN	70.2	97.4	87.6
9	ResNet50	-	PN	62.9	72.2	68.4
10	ResNet18	-	PN	63.3	73.7	70.2
11	ViT-base	DINO (IN1K)	PN	79.2	98.4	92.2
12	ViT-base	CLIP (YFCC)	PN	80.0	98.1	93.2
13	ViT-base	Sup (IN21K)	PN	81.4	99.2	96.7
14	ViT-base	BEIT (IN21K)	PN	82.8	99.0	97.5
15	ResNet50	CLIP (YFCC)	PN	75.0	92.2	82.6

Influence of pre-training.

M	Arch	PreTr	MetaTr	MetaTe	Avg	Out-D
1	ViT-small	DINO	PN (IN)	PN	68.38	67.68
2	ViT-small	DINO	PN (IN)	PN+FT(lr=0.01)	76.05	76.54
3	ViT-small	DINO	PN (IN)	PN+FT(lr=0.001)	74.47	74.51
4	ViT-small	DINO	PN (IN)	PN+FT(Tuned)	77.53	77.85
5	ViT-small	DINO	PN (MD)	PN	78.43	55.71
6	ViT-small	DINO	PN (MD)	PN+FT(lr=0.01)	76.09	73.26
7	ViT-small	DINO	PN (MD)	PN+FT(lr=0.001)	74.64	69.97
8	ViT-small	DINO	PN (MD)	PN+FT(Tuned)	83.13	75.72

Influence of fine-tuning.

Hu et al, Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference, in CVPR 2022.

FSL lately …

- FSL with foundation models
- Visual (/language) in-context learning

P>M>F

E.g. DINO > ProtoNet (PN) > Fine-tuning (FT)

Algorithm 1 PyTorch pseudo code for fine-tuning

```
# Inputs: a task including supp_x, supp_y, query_x
# backbone_state: meta-trained backbone weights
# optimizer: Adam optimizer
# Outputs: logits

backbone = create_model_from_checkpoint(backbone_state)

def single_step(z):
    supp_f = backbone(supp_x)
    proto = compute_prototypes(supp_f, supp_y)
    f = backbone(z)
    logits = f.norm() @ proto.norm().T # cos similarity
    loss = cross_entropy_loss(logits, supp_y)
    return logits, loss

# fine-tuning loop
for i in range(num_steps):
    aug_supp_x = rand_data_augment(supp_x)
    _, loss = single_step(aug_supp_x)
    loss.backward() # back-prop
    optimizer.step() # gradient descent

logits, _ = single_step(query_x) # classification
```

8 in-domain datasets	In-domain								Out-of-domain		
	INet	Omglot	Acraft	CUB	DTD	QDraw	Fungi	Flower	Sign	COCO	Avg
ProtoNet [65] (RN18)	67.01	44.5	79.56	71.14	67.01	65.18	64.88	40.26	86.85	46.48	63.29
CNAPs [56] (RN18+Adapter)	50.8	91.7	83.7	73.6	59.5	74.7	50.2	88.9	56.5	39.4	66.90
SUR [26] (RN18+Adapter)	57.2	93.2	**90.1**	82.3	73.5	81.9	67.9	88.4	67.4	51.3	75.32
T-SCNAPs [7] (RN18+Adapter)	58.8	93.9	84.1	76.8	69.0	78.6	48.8	91.6	76.1	48.7	72.64
URT [48] (RN18+Adapter)	55.7	94.4	85.8	76.3	71.8	82.5	63.5	88.2	69.4	52.2	73.98
FLUTE [64] (RN18)	51.8	93.2	87.2	79.2	68.8	79.5	58.1	91.6	58.4	50.0	71.78
URL [44] (RN18+Adapter)	57.51	94.51	88.59	80.54	76.17	81.94	68.75	92.11	63.34	54.03	75.75
ITA [43] (RN18+Adapter)	57.35	**94.96**	89.33	81.42	76.74	**82.01**	67.4	92.18	83.55	55.75	78.07
P>M>F (DINO/IN1K, RN50)	67.51	85.91	80.3	81.67	**87.08**	72.84	60.03	94.69	87.17	58.92	77.61
P>M>F (DINO/IN1K, ViT-small)	74.59	91.79	88.33	91.02	86.61	79.23	74.2	94.12	88.85	62.59	83.13
P>M>F (DINO/IN1K, ViT-base)	**77.02**	91.76	89.73	**92.94**	86.94	80.2	**78.28**	**95.79**	**89.86**	**64.97**	**84.75**

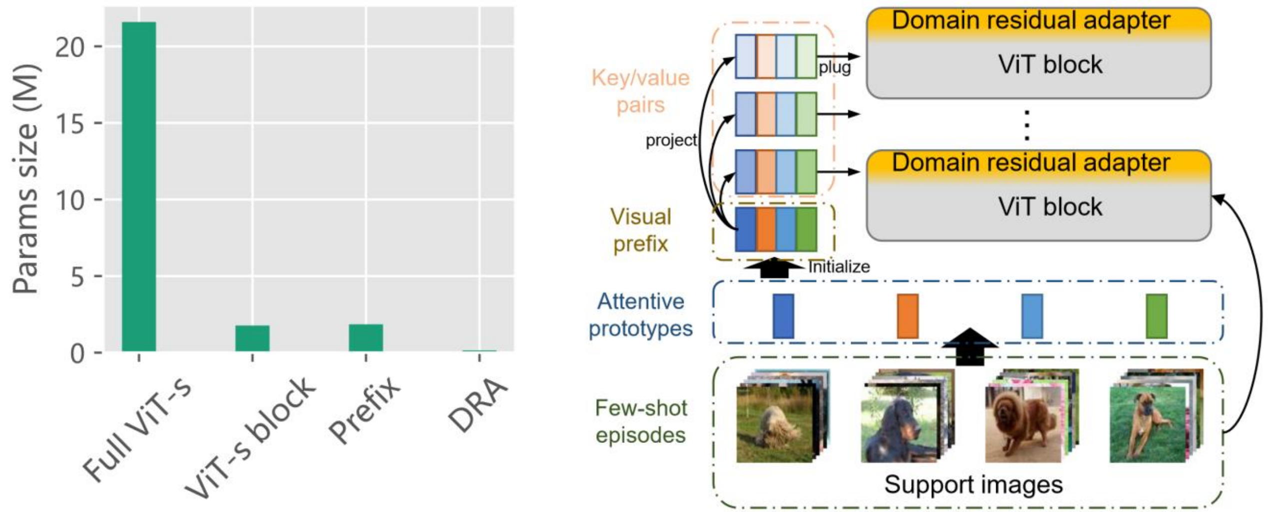
In-domain = ImageNet	In-domain	Out-of-domain									
	INet	Omglot	Acraft	CUB	DTD	QDraw	Fungi	Flower	Sign	COCO	Avg
ProtoNet [65] (RN18)	50.5	59.98	53.1	68.79	66.56	48.96	39.71	85.27	47.12	41	56.10
ALFA+FP-MAML [5] (RN12)	52.8	61.87	63.43	69.75	70.78	59.17	41.49	85.96	60.78	48.11	61.41
BOHB [58] (RN18)	51.92	67.57	54.12	70.69	68.34	50.33	41.38	87.34	51.8	48.03	59.15
CTX [24] (RN34)	62.76	82.21	79.49	80.63	75.57	72.68	51.58	**95.34**	82.65	59.9	74.28
P>M>F (DINO/IN1K, RN50)	67.08	75.33	75.39	72.08	86.42	66.79	50.53	94.14	86.54	58.2	73.25
P>M>F (DINO/IN1K, ViT-small)	74.69	80.68	76.78	85.04	86.63	71.25	54.78	94.57	88.33	62.57	77.53
P>M>F (DINO/IN1K, ViT-base)	**76.69**	**81.42**	**80.33**	**84.38**	**86.87**	**75.43**	**55.93**	95.14	**89.68**	**65.01**	**79.09**

Meta-Dataset – Comparison with SOTA FSL algorithms.

Hu et al, Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference, in CVPR 2022.

FSL lately …

- FSL with foundation models
- Visual (/language) in-context le



(a) Tunable parameters in Backbone Finetuning (b) Attentive Prefix Tuning in Task Tuning

P>~~M~~>F

E.g. DINO > ~~ProtoNet (PN)~~ > Fine-tuning (FT)

Model	Backbone	ILSVRC	Omni	Acraft	CUB	DTD	QDraw	Fungi	Flower	Sign	COCO	Avg	Rank
Finetune	Res18	45.78	60.85	68.69	57.31	69.05	42.60	38.20	85.51	66.79	34.86	56.96	10.2
Proto		50.50	59.98	53.10	68.79	66.56	48.96	39.71	85.27	47.12	41.00	56.10	10.5
Relation		34.69	45.35	40.73	49.51	52.97	43.30	30.55	68.76	33.67	29.15	42.87	14.6
P-MAML		49.53	63.37	55.95	68.66	66.49	51.52	39.96	87.15	48.83	43.74	57.52	9.2
BOHB		51.92	67.57	54.12	70.69	68.34	50.33	41.38	87.34	51.80	48.03	59.15	8.2
TSA		**59.50**	**78.20**	72.20	**74.90**	77.30	67.60	44.70	90.90	82.50	**59.00**	**70.68**	4.3
Ours	ViT-t	56.40	72.52	**72.84**	73.79	**77.57**	**67.97**	**51.23**	**93.30**	**84.09**	55.68	70.54	4.1
Proto	Res34	53.70	68.50	58.00	74.10	68.80	53.30	40.70	87.00	58.10	41.70	60.39	7.4
CTX		62.76	82.21	79.49	80.63	75.57	**72.68**	51.58	95.34	82.65	59.90	74.28	2.8
TSA		63.73	**82.58**	**80.13**	83.39	79.61	71.03	51.38	94.05	81.71	61.67	74.93	2.5
P>M>F*	ViT-s	74.69	80.68	76.78	85.04	86.63	71.25	54.78	94.57	88.33	62.57	77.53	—
Ours		**67.37**	78.11	79.94	**85.93**	**87.62**	71.34	**61.80**	**96.57**	85.09	**62.33**	**77.61**	**1.6**

Meta-Dataset – Comparison with SOTA FSL algorithms.

Xu et al, Exploring Efficient Few-shot Adaptation for Vision Transformers, in TMLR 2022.

FSL lately …

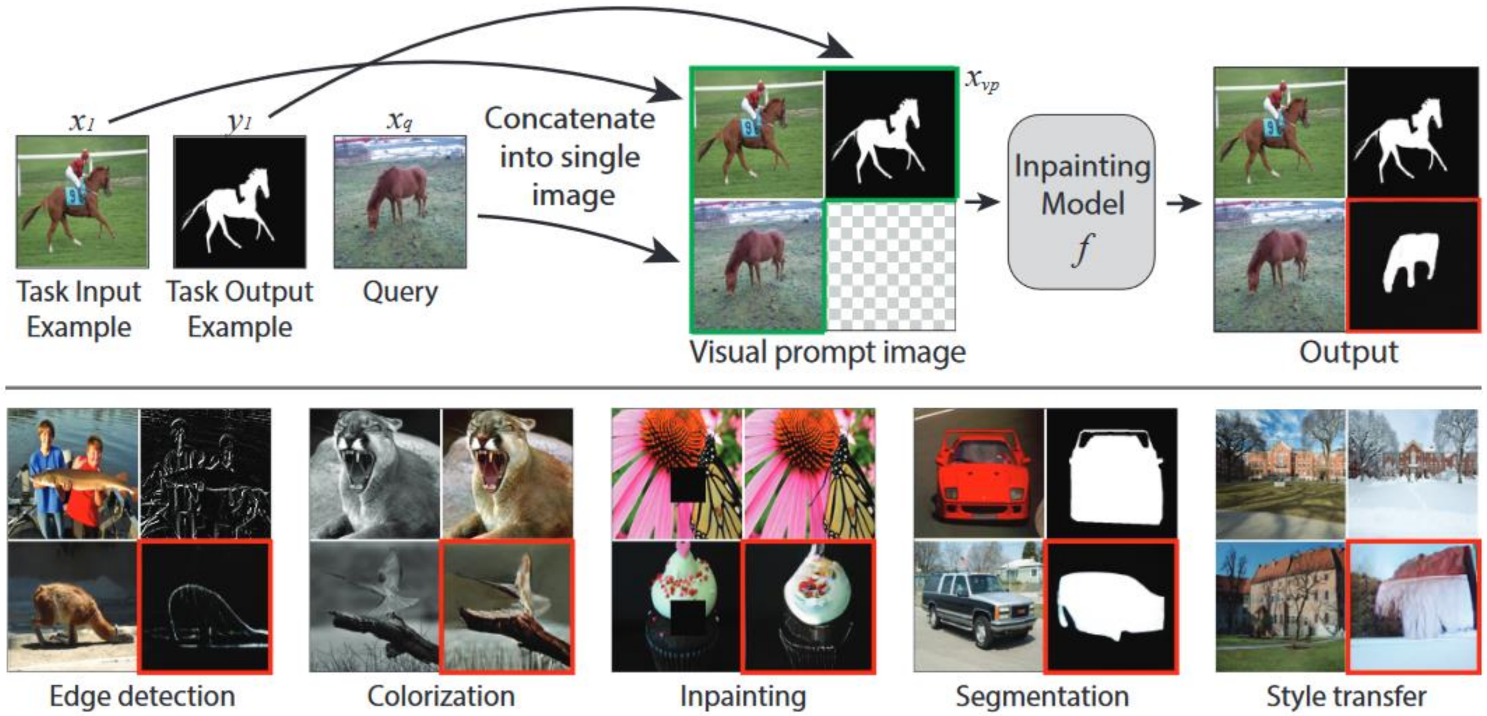
- FSL with foundation models
- Visual (/language) in-context learning

FSL lately …

Je suis désolé — I'm sorry
J'adore la glace — ??

I love ice cream

- FSL with foundation models
- Visual (/language) in-context learning



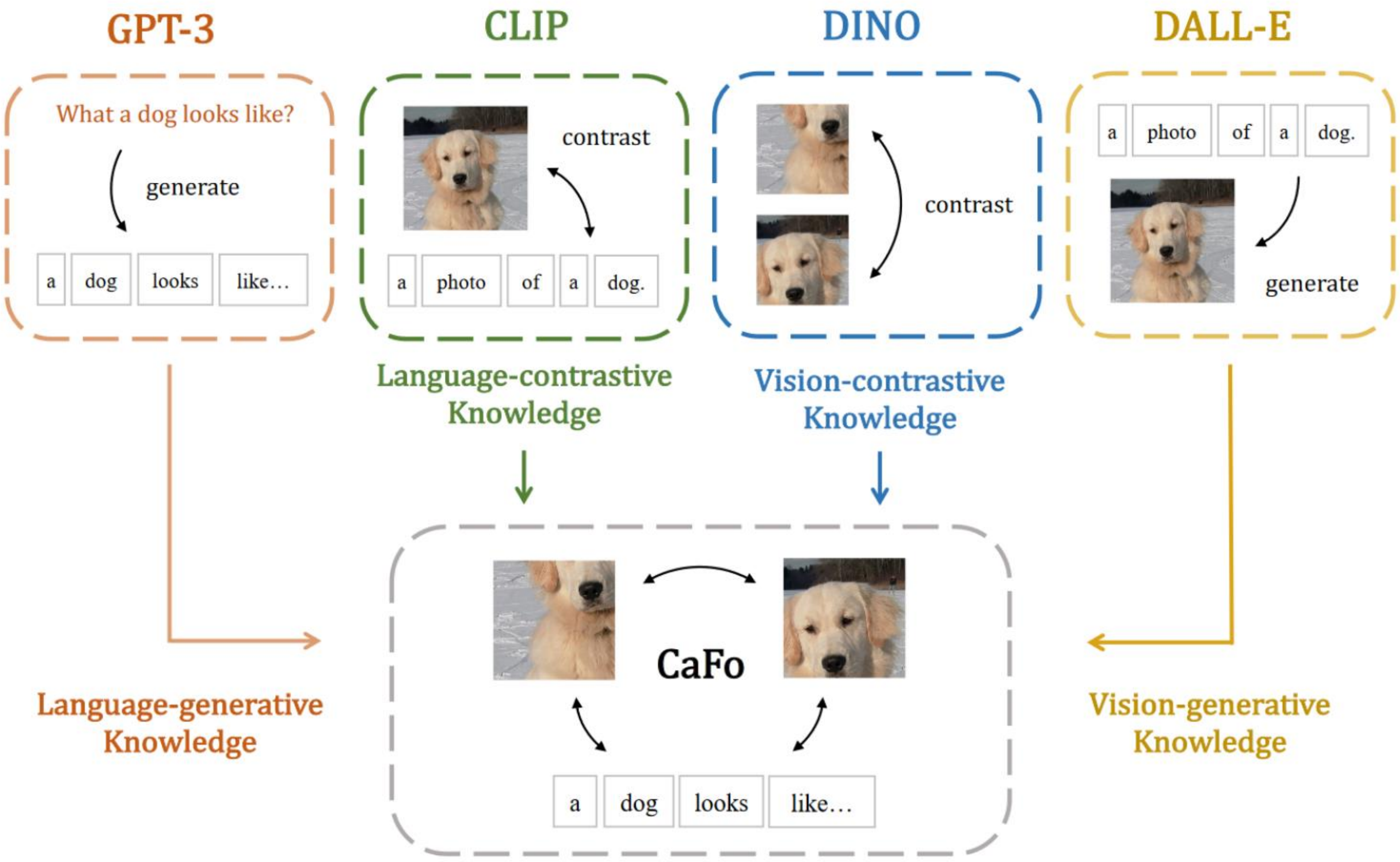
Bar et al, Visual Prompting via Image Inpainting, in NeurIPS 2022.

FSL lately …

- FSL with foundation models
- Visual (/language) in-context learning

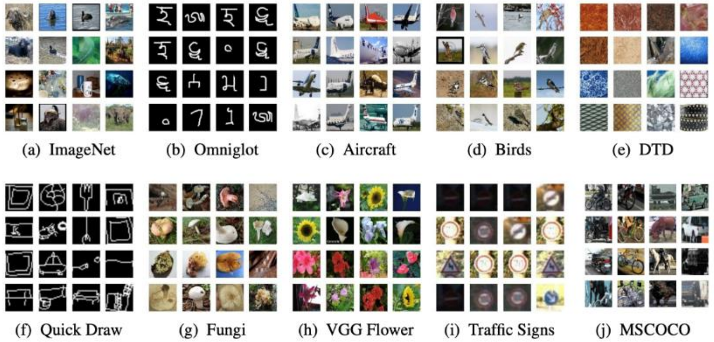
Tsimpoukelli et al, Multimodal Few-Shot Learning with Frozen Language Models, in NeurIPS 2021.

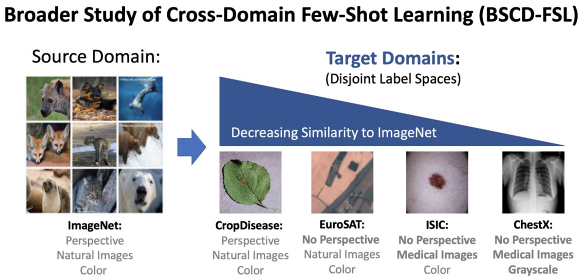
Take home

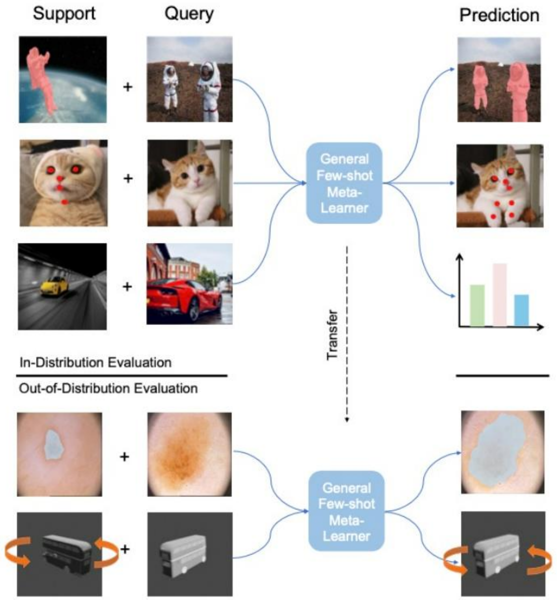


A Cascade Paradigm of Foundation Models

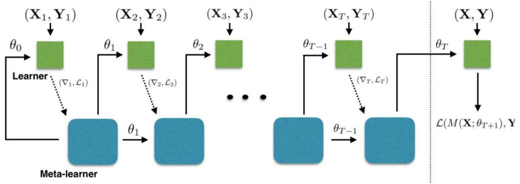
Zhang et al, Prompt, Generate, then Cache: Cascade of Foundation Models makes Strong Few-shot Learners, in CVPR 2023.

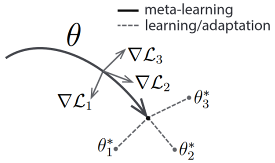


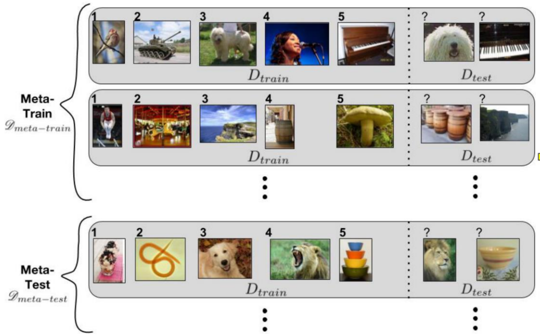


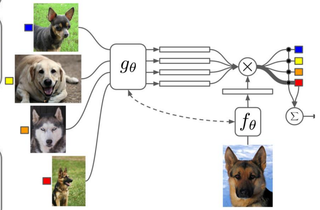


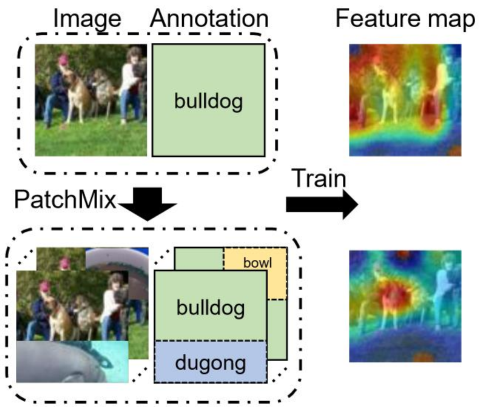
Thanks!











Algorithm 1 PyTorch pseudo code for fine-tuning

```
# Inputs: a task including supp_x, supp_y, query_x
# backbone_state: meta-trained backbone weights
# optimizer: Adam optimizer
# Outputs: logits

backbone = create_model_from_checkpoint(backbone_state)

def single_step(z):
    supp_f = backbone(supp_x)
    proto = compute_prototypes(supp_f, supp_y)
    f = backbone(z)
    logits = f.norm() @ proto.norm().T # cos similarity
    loss = cross_entropy_loss(logits, supp_y)
    return logits, loss

# fine-tuning loop
for i in range(num_steps):
    aug_supp_x = rand_data_augment(supp_x)
    _, loss = single_step(aug_supp_x)
    loss.backward() # back-prop
    optimizer.step() # gradient descent

logits, _ = single_step(query_x) # classification
```

P>M>F