Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference

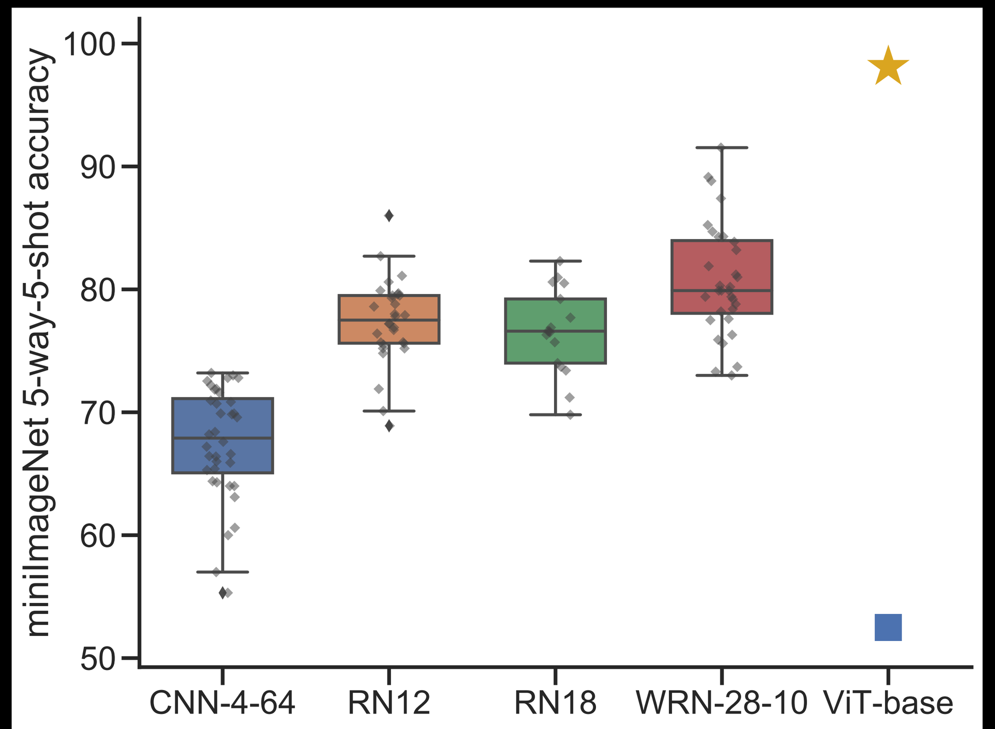
Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim and Timothy M. Hospedales

SAMSUNG Samsung AI Center-Cambridge

† Introduction

Few-shot learning (FSL) is commonly characterized as a meta-learning problem. We however find empirically that meta-learning is the least important part in a three-stage pipeline for FSL: Pre-training → Meta-training → Fine-tuning (P>M>F). In this work, we investigate three previously under-studied design choices: external source data, network architecture, and meta-test time fine-tuning. We show that a simple transformer-based pipeline yields surprisingly good performance on standard benchmarks such as Mini-ImageNet, CIFAR-FS, CDFSL and Meta-Dataset.

† Motivation



We aggregated dozens of studies from the past 5 years of FSL research on Mini-ImageNet and compared them with the results of our simple pipeline: ProtoNet + ViT (Yellow star uses CLIP. Blue square has no pre-training). Pre-training made a huge difference, even compared with previous SOTA.

† P>M>F pipeline: pre-training backbone on external data → meta-training → fine-tuning backbone in meta-test

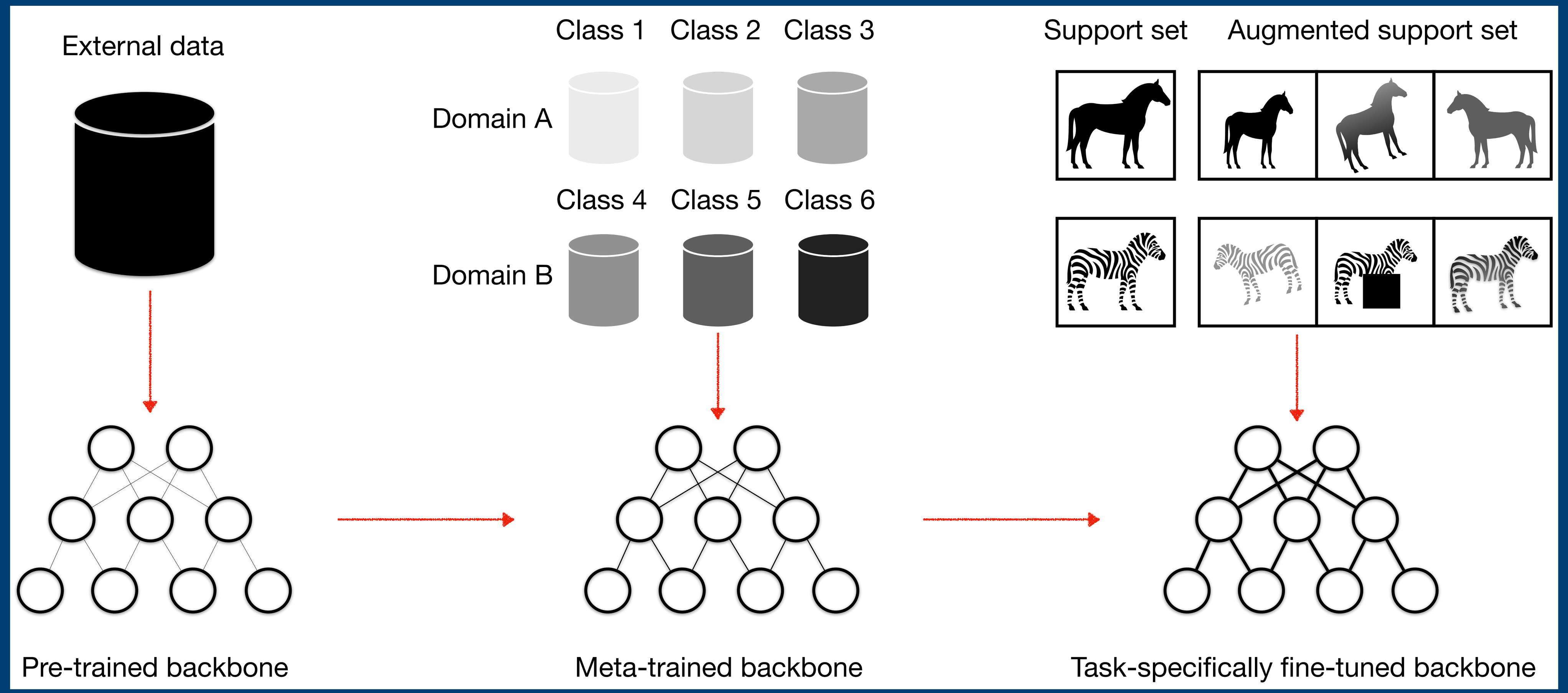
- **Pre-training**: we compared DINO [15], BEiT [6] and CLIP [53]. They got equally good results, which can be further boosted if supervision is available.
- **Meta-training**: we compared ProtoNet [59], MetaOptNet [42] and MetaQDA [72]. However, the simplest ProtoNet outperformed the more sophisticated counterparts.
- **Fine-tuning**: we fine-tune the whole backbone on the support set of a novel task with mild data augmentation, where the learning rate is selected for each domain using 5 validation tasks.

What are the simple ingredients for achieving SOTA few-shot learning?

a) Pre-trained foundation model

b) ~~SOTA meta-learning~~ ProtoNet

c) Fine-tuning on meta-test tasks



Links to our paper, code and demo:

arXiv GitHub

† How do pre-training and architecture choices affect FSL?

	Training Configuration			Benchmark Results		
ID	Arch	Pre-Train	MetaTr	MD	Mini-IN	CIFAR
0	ViT-small	DINO (IN)	-	67.4	97.0	79.8
2	ResNet50	DINO (IN)	-	63.8	91.5	76.1
3	ResNet50	Sup. (IN)	-	62.4	96.4	82.3
4	ViT-small	DINO (IN)	ProtoNet	78.4	98.0	92.5
7	ResNet50	DINO (IN)	ProtoNet	72.4	92.0	84.0
9	ResNet50	-	ProtoNet	62.9	72.2	68.4

- 7 vs. 9: pre-training on ImageNet (IN) offers a strong feature to boost classical ProtoNet baseline.
- 2 vs. 9: pre-training alone is already good.
- 2 vs. 3: self-supervised pre-training is as good as supervised.
- 0 vs. 2 & 4 vs. 7: ViT-small > ResNet50.

† How to best exploit fine-tuning for meta-testing?

ID	Arch	PreTrain	MetaTrain	FineTune	Avg	Out-D
1	ViT-small	DINO	ProtoNet (IN only)	✗	68.38	67.68
4	ViT-small	DINO	ProtoNet (IN only)	✓	77.53	77.85
5	ViT-small	DINO	ProtoNet (MD)	✗	78.43	55.71
8	ViT-small	DINO	ProtoNet (MD)	✓	83.13	75.72

- 1 v 4, 5 v 8: fine-tuning in meta-test improves substantially.
- 1 v 5, 4 v 8: meta-training is recommended if possible.

† Comparison with SOTA: Meta-Dataset (MD)

Standard MD	In-domain								Out-of-domain		
	INet	Omglot	Acraft	CUB	DTD	QDraw	Fungi	Flower	Sign	COCO	Avg
ProtoNet	67.01	44.5	79.56	71.14	67.01	65.18	64.88	40.26	86.85	46.48	63.29
CNAPs [56]	50.8	91.7	83.7	73.6	59.5	74.7	50.2	88.9	56.5	39.4	66.90
ITA [43]	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	**77.02**	91.76	**89.73**	**92.94**	**86.94**	80.2	**78.28**	**95.79**	**89.86**	**64.97**	**84.75**

† Comparison with SOTA: Cross-domain FSL (CDFSL)

	ChestX			ISIC			EuroSAT			CropDisease		
	5w5s	5w20s	5w50s	5w5s	5w20s	5w50s	5w5s	5w20s	5w50s	5w5s	5w20s	5w50s
ProtoNet	24.05	28.21	29.32	39.57	49.50	51.99	73.29	82.27	80.48	79.72	88.15	90.81
Finetune [33]	25.97	31.32	35.49	48.11	59.31	66.48	79.08	87.64	90.89	89.25	95.51	97.68
STARTUP [52]	26.94	33.19	36.91	47.22	58.63	64.16	82.29	89.26	91.99	**93.02**	97.51	98.45
P>M>F	**27.27**	**35.33**	**41.39**	**50.12**	**65.78**	**73.50**	**85.98**	**91.32**	**95.40**	92.96	**98.12**	**99.24**