Neural Fine-Tuning Search for Few-Shot Learning

Panagiotis Eustratiadis[1], Łukasz Dudziak[2], Da Li[2], Timothy Hospedales[1,2]

[1] University of Edinburgh, UK [2] Samsung AI Centre Cambridge, UK

{ p.eustratiadis, t.hospedales }@ed.ac.uk { l.dudziak, da.li1 }@samsung.com

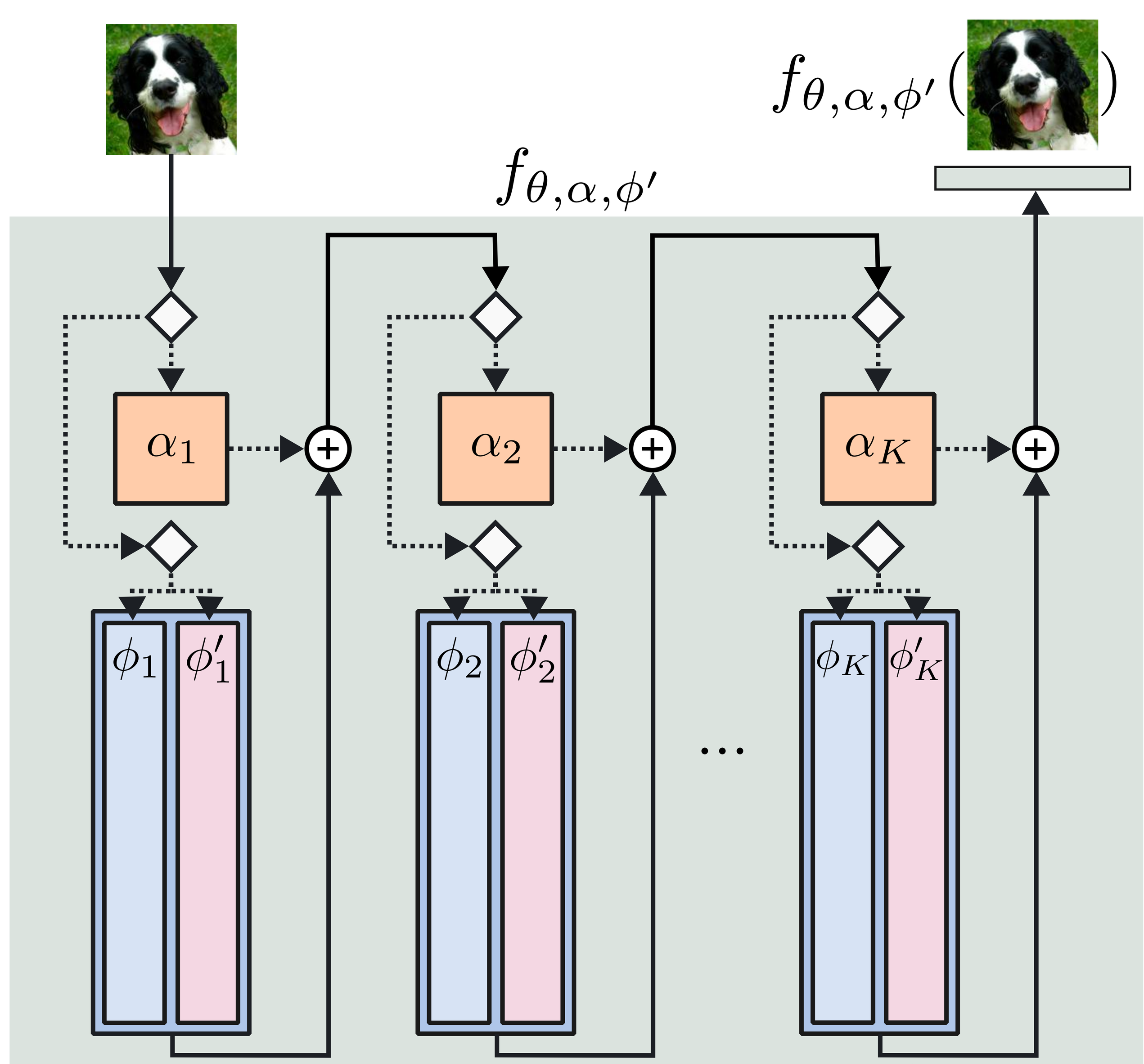
Main question

In multi-domain few-shot adaptation of a pretrained model, *what* layers and *how* should be fine-tuned for optimal results?

Main challenge

Solution depends on a downstream task! Needs to find a sweet spot between under- and over-fitting at testing time.

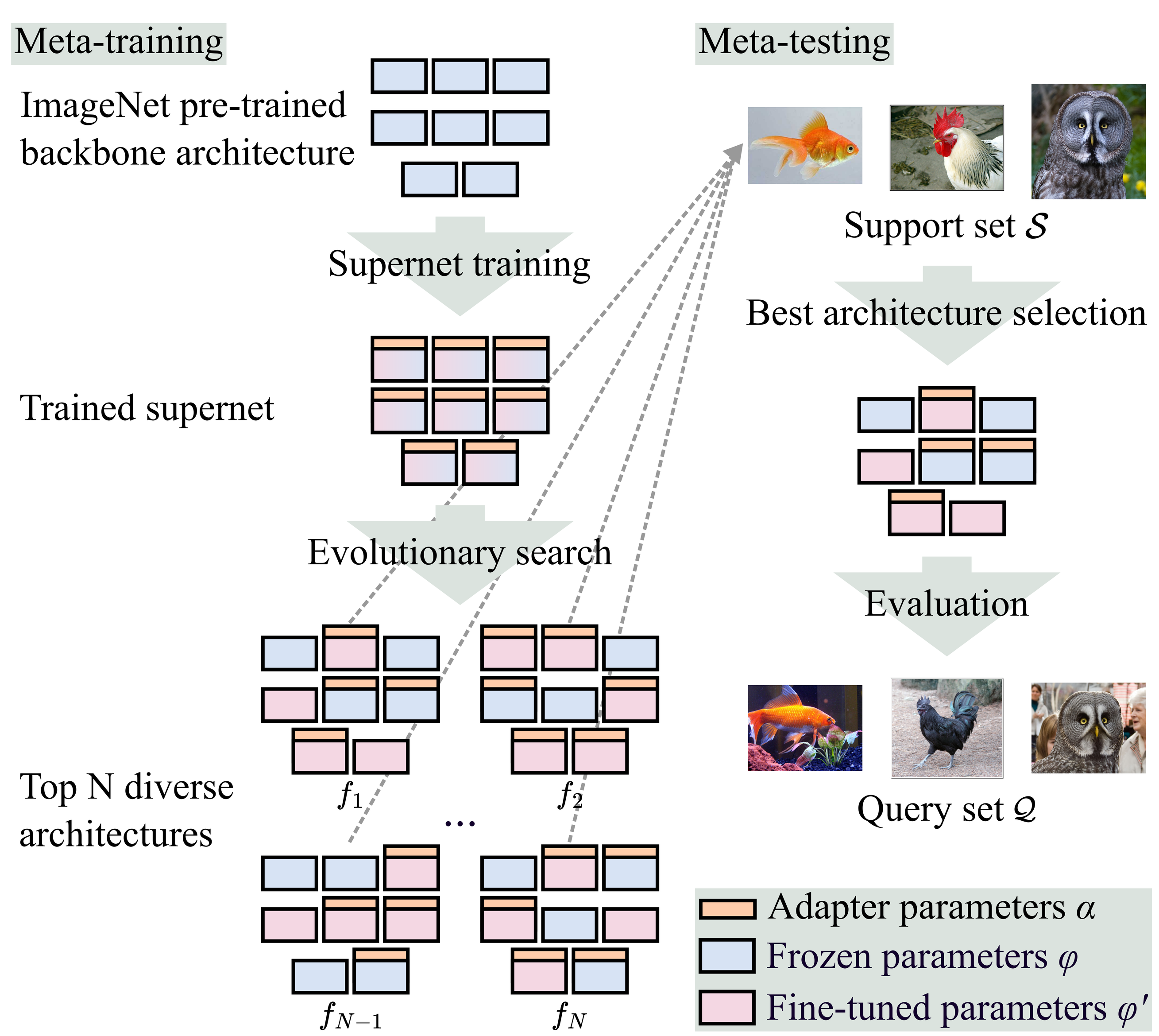
❶ The *what* and the *how*



What? ⇒ any standard block in a contemporary architecture is considered (ϕ_i)
How? ⇒ either by directly fine-tuning (ϕ'_i) or by using state-of-the-art adapters (α_i)

In total, for K blocks, 4^K possible designs!

❷ Searching for a solution



Training $$\arg\min_{\alpha,\phi'} \mathbb{E}_{p\sim P}\, \mathbb{E}_{\mathcal{S},\mathcal{Q}}\, \mathcal{L}(f^p_{\theta,\alpha,\phi'},\mathcal{S},\mathcal{Q})$$

$$\arg\max_{\{p_1,\dots,p_N\}\in P} \mathbb{E}_{\mathcal{S},\mathcal{Q}}\, A(f^p_{\theta,\alpha^*,\phi'^*},\mathcal{S},\mathcal{Q})$$

Testing $$\arg\min_{p\in\{p_1,\dots,p_N\}} \mathcal{L}(f^p_{\theta,\alpha^*,\phi'^*},\mathcal{S},\mathcal{S})$$

❸ Understanding and characterising the search

How N affects the risk of under- and over-fitting at testing time. First: 30 episodes, second: 600.

N →	1	3	10	100
Test acc. (S)	96.5	97.1	99.9	99.8
Test acc. (Q)	72.8	**72.9**	71.5	71.4
S/Q correl.	0.31	**0.35**	0.28	0.18
Test acc. (Q)	73.6	**75.2**	-	-

Robustness of the final selection and how diverse arch. help.

NTFS-3	Arch. 1	Arch. 2	Arch. 3
CIFAR10	82.0	81.2	**83.3**
CIFAR100	**75.9**	75.0	75.1
MNIST	**95.5**	94.4	95.1
MSCOCO	**58.1**	57.8	56.4
Tr. Signs	81.7	**82.2**	81.8

Search perf. vs. naive approaches.

	RN-18	ViT-S
ϕ , –	67.8	71.8
ϕ , α	70.4	73.8
ϕ' , –	70.2	74.0
ϕ' , α	**70.8**	**74.4**
NTFS	**75.2**	**79.2**

❹ Results

	Method	Aircrafts	Birds	DTD	Fungi	ImageNet	Omniglot	QuickDraw	Flowers	MSCOCO	Tr. Sign	Average
ResNet-18	FLUTE	48.5	47.9	63.8	31.8	46.9	61.6	57.5	80.1	41.4	46.5	52.6
	ProtoNet	53.1	68.8	66.6	39.7	50.5	60.0	49.0	85.3	41.0	47.1	56.1
	BOHB	54.1	70.7	68.3	41.4	51.9	67.6	50.3	87.3	48.0	51.8	59.2
	FO-MAML	63.4	69.8	70.8	41.5	52.8	61.9	59.2	86.0	48.1	60.8	61.4
	TSA	72.2	74.9	77.3	44.7	59.5	78.2	**67.6**	90.9	59.0	**82.5**	73.3
	NTFS	**74.9**	**76.5**	**81.6**	**50.5**	**62.7**	**80.2**	67.2	**94.5**	**59.7**	81.9	**75.2**
ViT-S	*PMF	76.8	85.0	86.6	54.8	**74.7**	80.7	71.3	94.6	**62.6**	**88.3**	77.5
	ETT	79.9	**85.9**	**87.6**	61.8	67.4	78.1	71.3	**96.6**	62.3	85.1	77.6
	NTFS	**83.0**	85.5	**87.6**	**62.2**	71.0	**81.9**	**74.5**	96.0	**62.6**	87.9	**79.2**

Partial single domain Meta-Dataset results. More in the paper! (* additional data)