A Flexible Selection Scheme for Minimum-Effort Transfer Learning

Amélie Royer, Christoph H. Lampert

Institute of Science and Technology Austria, Klosterneuburg, Austria

Motivation

Transfer learning

- Fine-tuning **all layers**, or from scratch
  - → Efficient, but requires enough data, or else, overfitting risk
- Fine-tuning **last layer**
  - → Assume similar visual inputs, otherwise, limited benefit

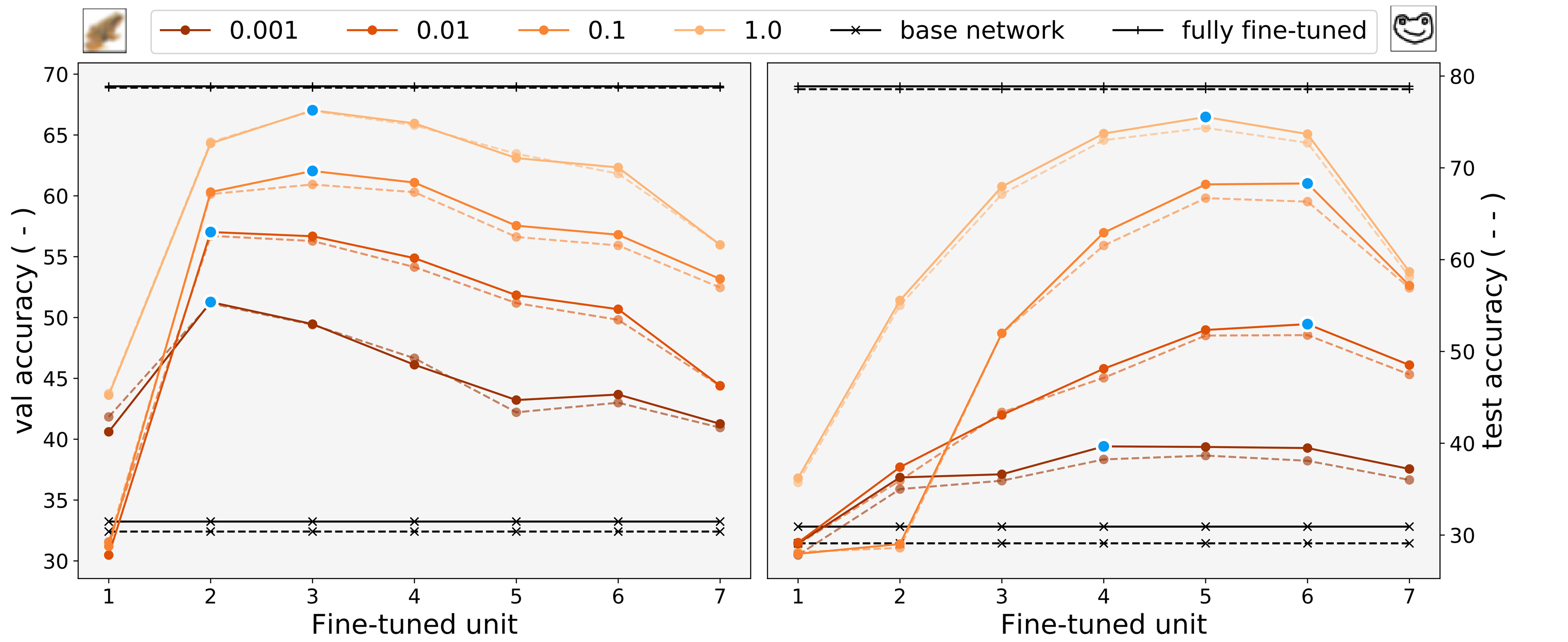
Orthogonal scenario ?

- Similar semantics (e.g. classes), different visual domains
  - → Common for applications with noisy data sources, e.g. noisy, blurry, rotated images...
  - → Data augmentation requires knowledge *before* training

Objective: Extend the common fine-tuning scheme to adapt intermediate *individual* units for visually dissimilar domains

Flex-tuning

- Model selection criterion based on validation accuracy.
- Early stopping to avoid overfitting bias
- Complex behavior depending on *(i)* data amount and *(ii)* the severity of the domain shift. **But**, beneficial in several scenarios.



Fast and Faster Flex-tuning

Fine-tuning all models to perform model selection is **costly**. Instead, we want to estimate how important each unit is with a lightweight and fast selection criterion.

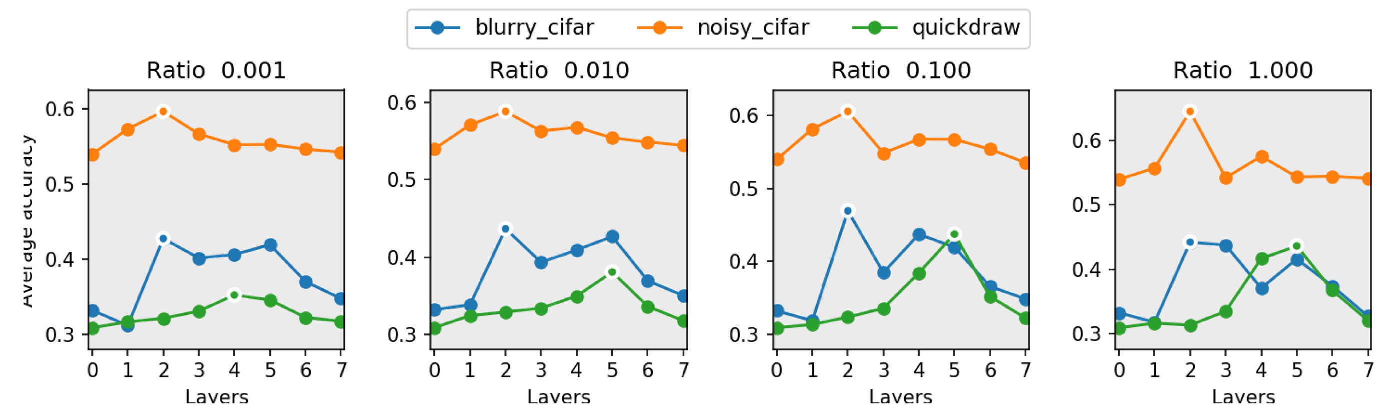
Fast Flex-tuning:

- Fine-tune all the layers of the model on the given training target dataset
- Estimate the influence of each unit by *network surgery* on the validation set

$$\phi_i = f^1 \circ \cdots \circ f^{i-1} \circ f^i_{\text{tgt}} \circ f^{i+1} \circ \cdots \circ f^N$$

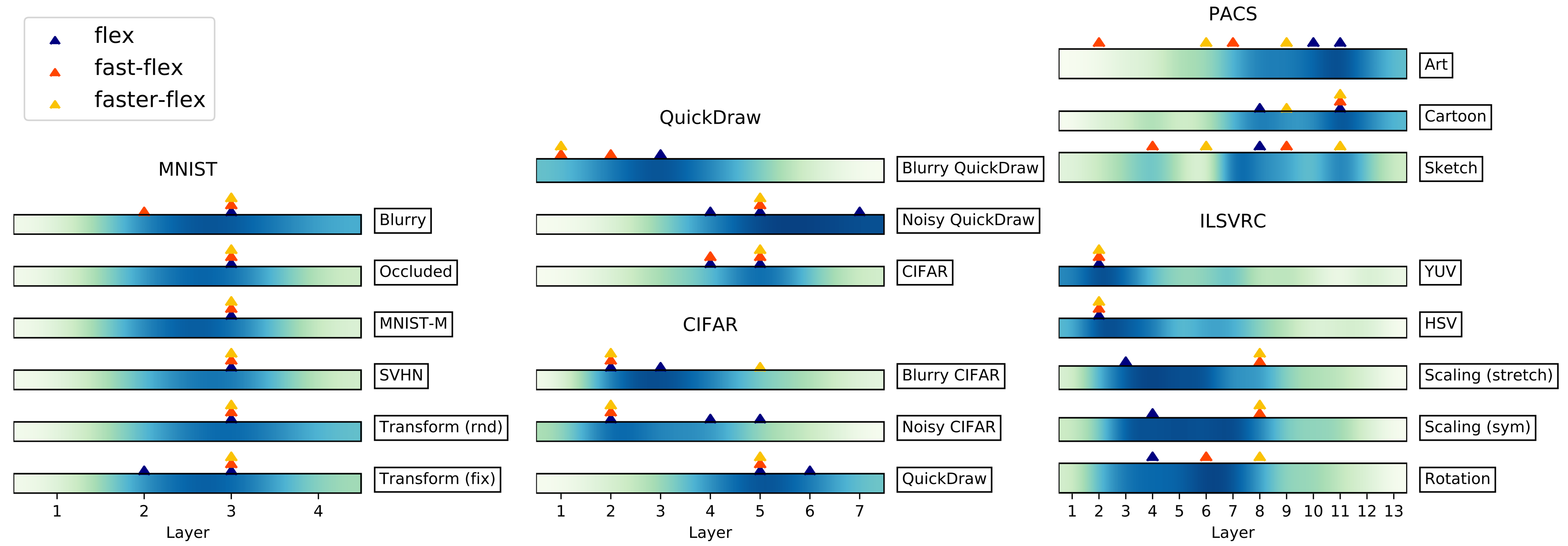
Faster Flex-tuning:

- Fine-tuning all layers is impossible or detrimental (overfitting) in data-scarce settings, hence, unreliable estimates.
- Instead, train the networks for *only a few epochs*: Not fully trained, but representative of the general gradient direction.



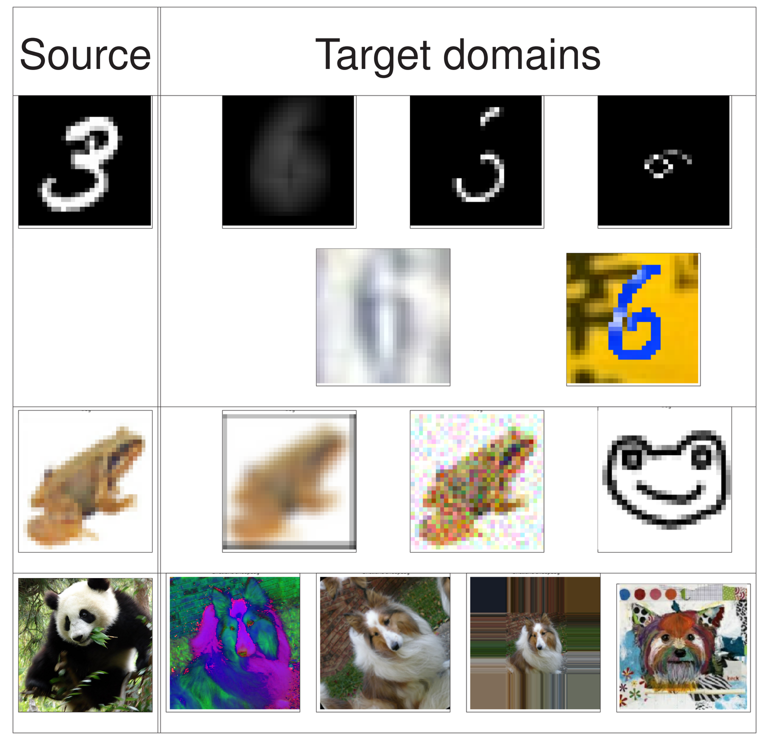
method	computational cost
`flex`	$LE_{\text{one}}c_{\text{one}} + E_{\text{all}}c_{\text{all}}$
`fast-flex`	$E_{\text{one}}c_{\text{one}} + E_{\text{all}}c_{\text{all}}$
`faster-flex`	$E_{\text{one}}c_{\text{one}} + c_{\text{all}}$
`ft-fc`	$E_{\text{one}}c_{\text{one}}$
`ft-all`	$E_{\text{all}}c_{\text{all}}$

Qualitative Results



Experiments

- Classification accuracy in different data scarcity scenarios
- Different domain shifts and associated network architectures
- Three **baselines**: fine-tuning last layer (`ft-fc`), all layers (`ft-all`) and extra scaling-shifting operations (`ft-ss`) [1].



ILSVRC	flex			ft-		
	`flex`	`fast`	`faster`	`fc`	`ss`	`all`
ratio: 2 images per class						
Art (0.53)	**0.669**	**0.703**	**0.655**	0.626	0.630	0.628
Cartoon(0.32)	0.639	**0.683**	0.593	0.618	0.647	0.507
Sketch (0.14)	**0.625**	**0.606**	0.414	0.554	0.581	0.337
ratio: 20 images per class						
Art (0.53)	**0.870**	**0.851**	**0.861**	0.729	0.849	0.724
Cartoon(0.32)	**0.912**	**0.893**	0.841	0.820	0.887	0.709
Sketch (0.14)	**0.852**	0.638	0.638	0.766	0.801	0.542
ratio: 200 images per class						
Art (0.53)	**0.906**	**0.906**	0.823	0.791	0.887	0.746
Cartoon(0.32)	**0.958**	0.956	0.952	0.868	0.956	0.925
Sketch (0.14)	**0.924**	**0.924**	0.890	0.767	0.916	0.875

- **Small architectures**: Fine-tuning all (4, 6) layers is often more beneficial. `ft-flex` recovers `ft-all`
- **Local pixel-level transformations**: Fine-tuning an early layer of the architecture yields the best results
- **Art style transformations**: More complex patterns, often captured by more than one unique unit

Conclusions

- **[+]** Fine-tuning an intermediate unit is beneficial in domain shifts with different visual inputs and similar semantics
- **[+]** Fast and faster selection criteria to compensate the fact that we have to estimate which unit is best to tune
- **[-]** So far, the study is limited to similar output domains, e.g. target classes form a subset of the source classes

[1] Q. Sun et al., *"Meta-transfer learning for few-shot learning"*. CVPR, 2019.