Dense Classification and Implanting for Few-Shot Learning

Yann Lifchitz[1,2], Yannis Avrithis[1], Sylvaine Picard[2], Andrei Bursuc[3]

[1]Univ Rennes, Inria, CNRS, IRISA [2]Safran [3]valeo.ai

Few-shot learning

- Recognize previously unseen classes with very few annotated examples
- Take advantage of a large dataset of images from a variety of base classes to learn a class independent embedding
- Use the few-shot data to adapt to novel classes disjoint from the base classes

Contributions

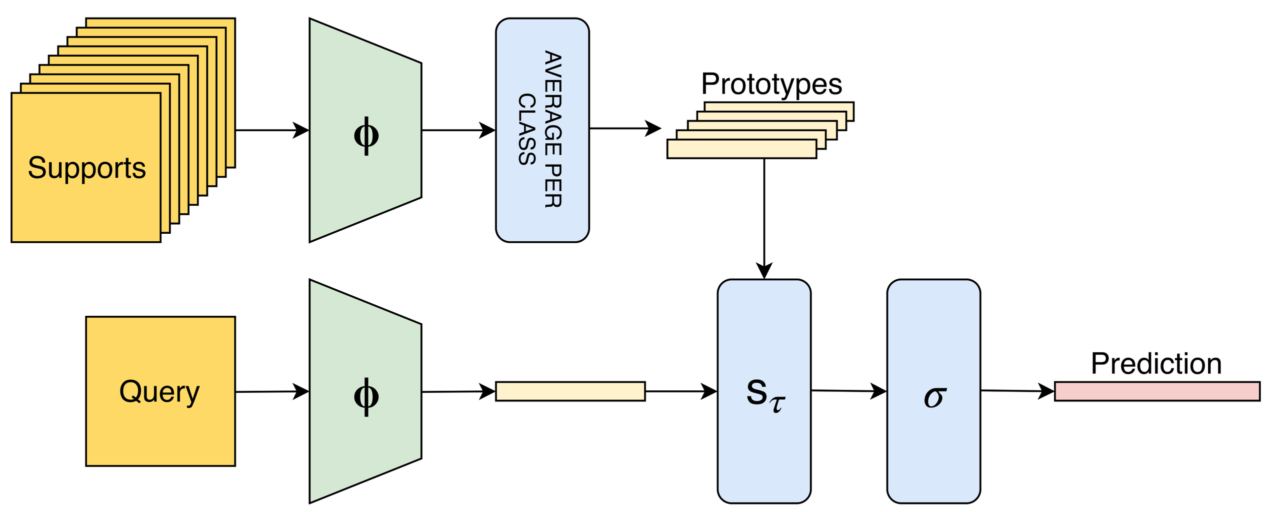
- **Dense classification**: Use local activations to better learn an embedding function
- **Implanting**: Attach new neurons to a previously trained network to learn new, task-specific features
- Improve the prior state-of-the-art on few-shot classification of *mini*ImageNet and FC100

First learning stage (base classes)

- Goal: Learn a task-agnostic embedding function
- Data: Large dataset in base class set C
- Embedding function $\phi_\theta(\mathbf{x}) \in \mathbb{R}^d$: CNN+pooling
- Classifier $f_{\theta,W}(\mathbf{x}) := \left([s_\tau(\phi_\theta(\mathbf{x}), \mathbf{w}_j)]_{j=1}^{c}\right)$, where $\mathbf{w}_j \in \mathbb{R}^d$ the weights of class j
- Similarity function s_τ: scaled cosine similarity [2, 4] $s_\tau(\mathbf{x}, \mathbf{y}) := \tau \langle \hat{\mathbf{x}}, \hat{\mathbf{y}} \rangle$
- Loss: softmax + cross-entropy, with τ as a free parameter

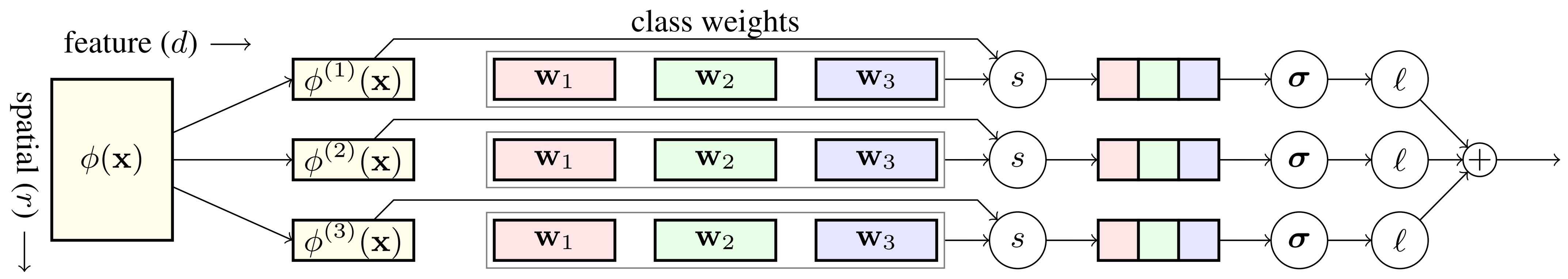
Second learning stage (novel classes)

- Goal: Classify queries of novel classes C' disjoint from C
- Data: k labeled images per novel class: k-shot c'-way classification
- Novel class weights: prototypes [5] from labeled examples



Dense classification (DC)

- Embedding tensor in $\mathbb{R}^{r \times d}$ is usually flattened or pooled
- Here, seen as collection of vectors $[\phi^{(k)}(\mathbf{x})]_{k=1}^{r}$, where $\phi^{(k)}(\mathbf{x})$ represents spatial location k
- Average pooling of losses at training, average pooling of predictions at testing



Dense classification study

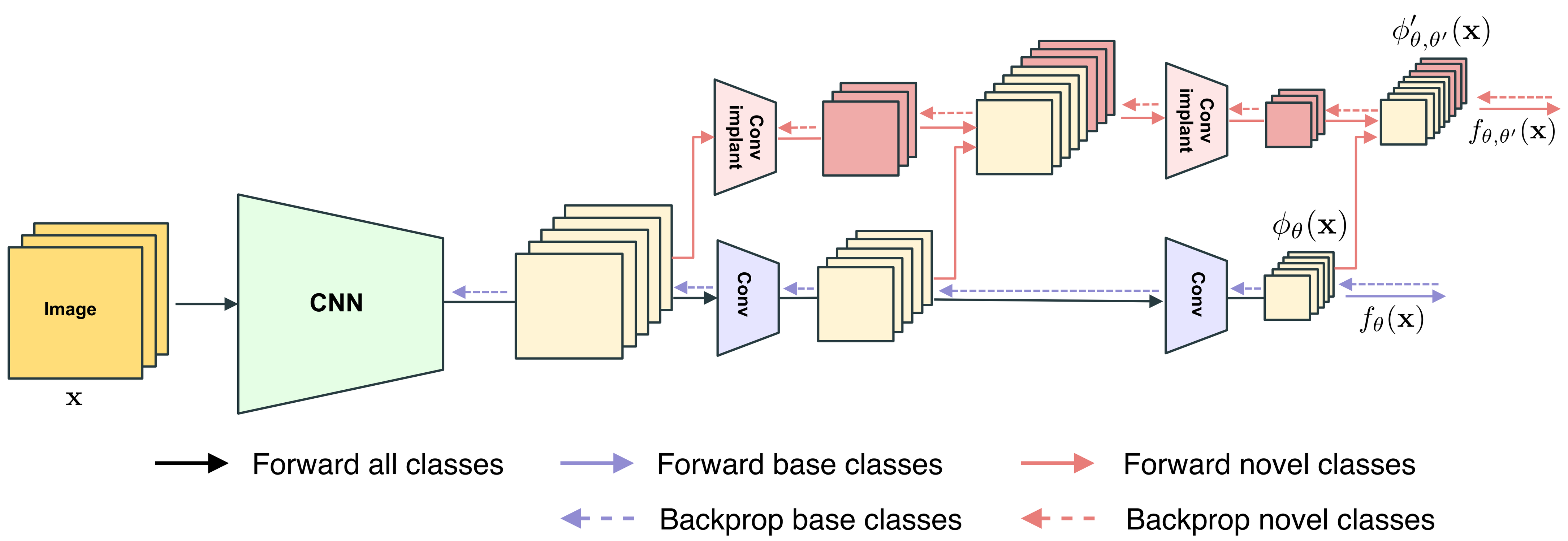
Stage 1 training		Support/query pooling at testing			
	Support →	GMP		GAP	
	Queries →	GMP	DC	GAP	DC
Global average pooling	Base classes	63.55 ±0.20	77.17 ±0.11	79.37 ±0.09	77.15 ±0.11
	Novel classes	72.25 ±0.13	70.71 ±0.14	76.40 ±0.13	73.28 ±0.14
	Both classes	37.74 ±0.07	38.65 ±0.05	56.25 ±0.10	54.80 ±0.09
Dense classification	Base classes	79.28 ±0.10	**80.67 ±0.10**	**80.61 ±0.10**	**80.70 ±0.10**
	Novel classes	**79.01 ±0.13**	77.93 ±0.13	78.55 ±0.13	**78.95 ±0.13**
	Both classes	42.45 ±0.07	57.98 ±0.10	67.53 ±0.10	**67.78 ±0.10**

Average 5-way 5-shot accuracy on base, novel and both classes of miniImageNet with ResNet-12
GMP: global max-pooling; GAP: global average pooling.

- Best: DC at training, global average pooling (GAP) on support, DC on queries
- DC allows integrating novel classes without forgetting the base ones

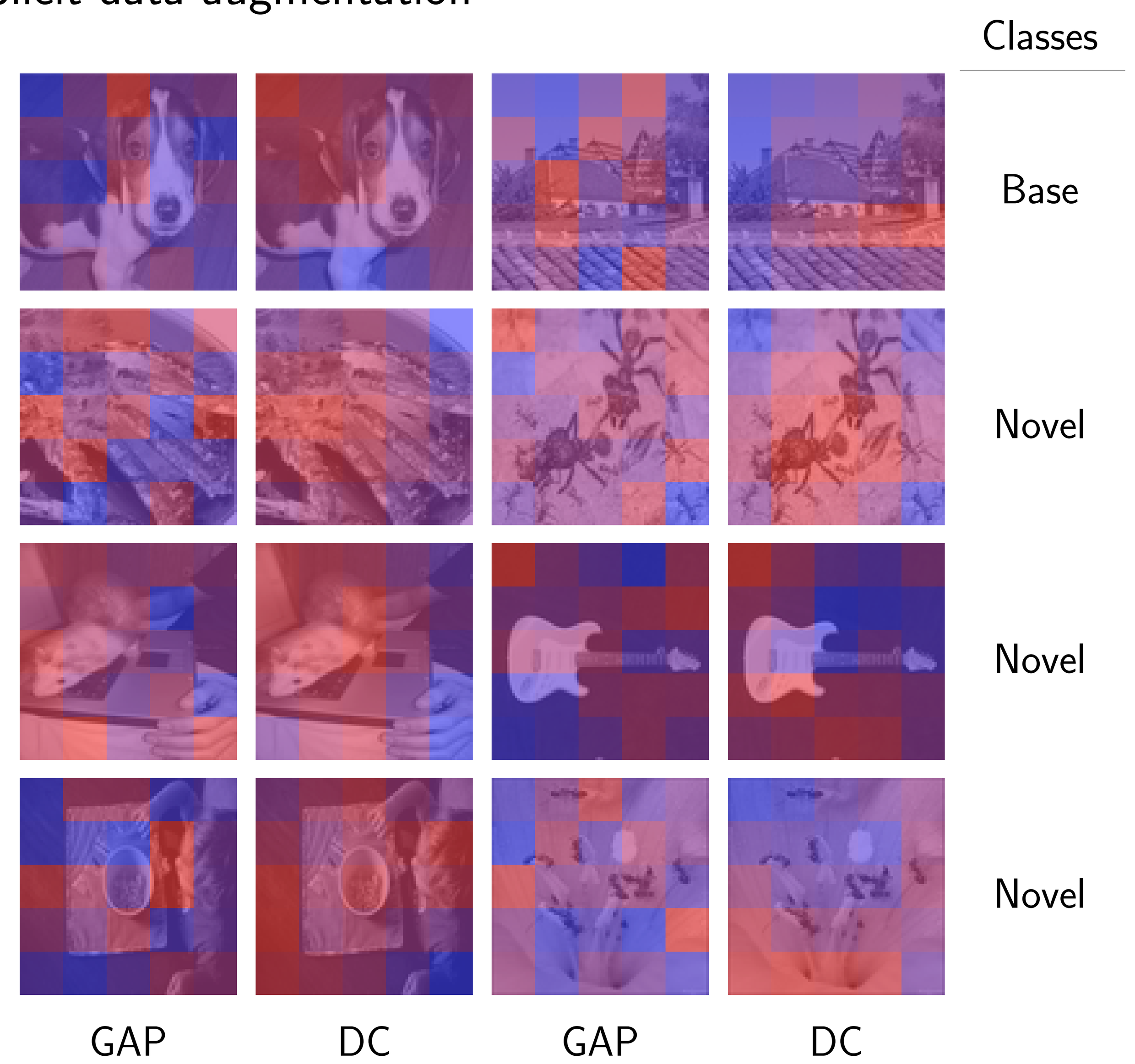
Implanting

- Goal: Learn new features specific to the few-shot task using only few-shot data
- Implants: Additional convolution kernels trained on few-shot examples
- Base network frozen during this stage
- Implant features and base network features concatenated



Effect of dense classification

- Improves spatial distribution of class activation
- Encourages correct classification at all spatial locations
- Implicit data augmentation



Images overlaid with correct class activation maps [6] with models learned with global average pooling (GAP) or dense classification (DC)

Results

Method	1-shot	5-shot	10-shot
GAP	58.61 ± 0.17	76.40 ± 0.13	80.76 ± 0.11
DC (ours)	**62.53 ± 0.19**	78.95 ± 0.13	82.66 ± 0.11
DC + WIDE	61.73 ± 0.19	78.25 ± 0.14	82.03 ± 0.12
DC + IMP (ours)	-	**79.77 ± 0.19**	**83.83 ± 0.16**
MAML [1]	48.70 ± 1.8	63.10 ± 0.9	-
PN [5]	49.42 ± 0.78	68.20 ± 0.66	-
LWF [2]	55.45 ± 0.7	73.00 ± 0.6	-
PN [3]	56.50 ± 0.4	74.20 ± 0.2	78.60 ± 0.4
TADAM [3]	58.50	76.70	80.80

Average 5-way novel-class accuracy on miniImageNet with ResNet-12. GAP: global average pooling; WIDE: last residual block widened by 16 channels; IMP: last residual block implanted by 16 channels.

References

[1] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.

[2] S. Gidaris and N. Komodakis. Dynamic few-shot visual learning without forgetting. In *CVPR*, 2018.

[3] B. N. Oreshkin, A. Lacoste, and P. Rodriguez. Tadam: Task dependent adaptive metric for improved few-shot learning. *NIPS*, 2018.

[4] H. Qi, M. Brown, and D. G. Lowe. Low-shot learning with imprinted weights. In *CVPR*, 2018.

[5] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. In *NIPS*, 2017.

[6] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *CVPR*, 2016.

Contact: yann.lifchitz@safrangroup.com