ECML PKDD 2021

Spatial Contrastive Learning for Few-Shot Classification

Yassine Ouali, Céline Hudelot, Myriam Tami
Paris-Saclay University, CentraleSupélec, MICS

université PARIS-SACLAY

CentraleSupélec

MICS

Few-shot Learning: an introduction

Few-shot learning: equip the learner with ability to rapidly learn new concept with few training samples.

?????

Few-shot Learning: an introduction

Few-shot learning: equip the learner with ability to rapidly learn new concept with few training samples.

Elephant Shrew

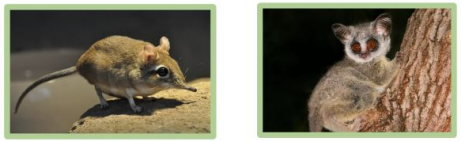
Few training samples

A few-shot classification task: N-way K-shot classification task

- N: number of classes
- K: number of examples per class (small)

5–way (classes) **1**–shot (example per class) Task

Train set (support set)

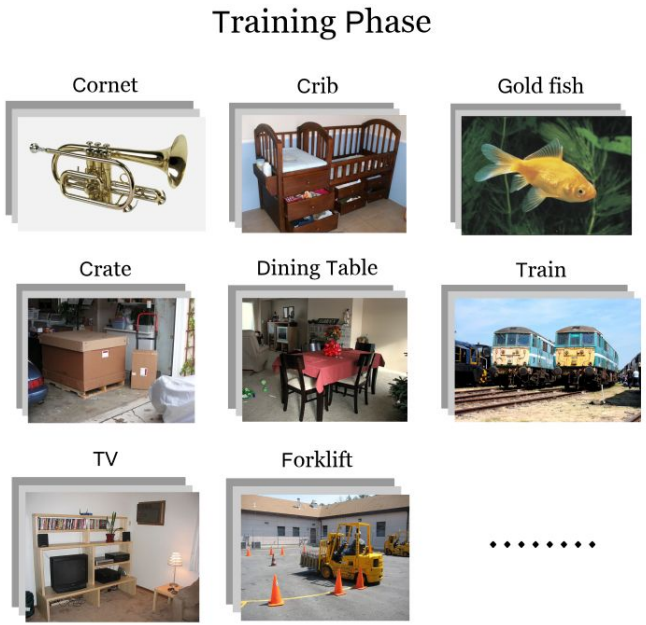


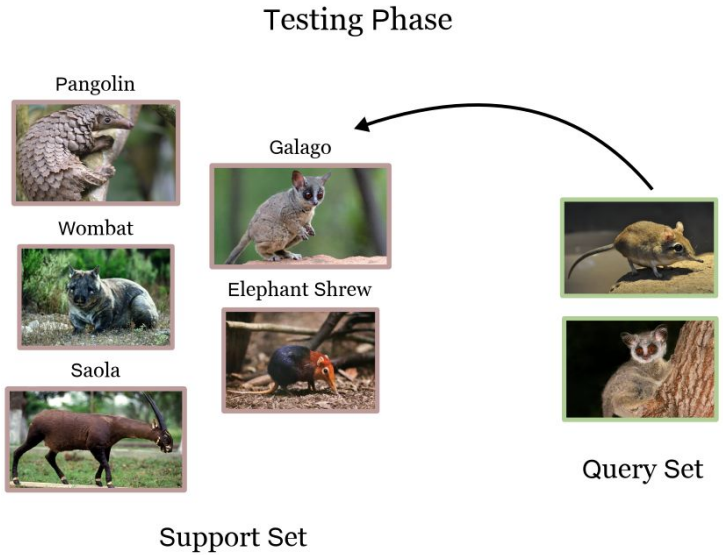
Test set (query set)

Few-shot Learning: an introduction

A few-shot classification task: N-way K-shot classification task

- N: number of classes
- K: number of examples per class (small)

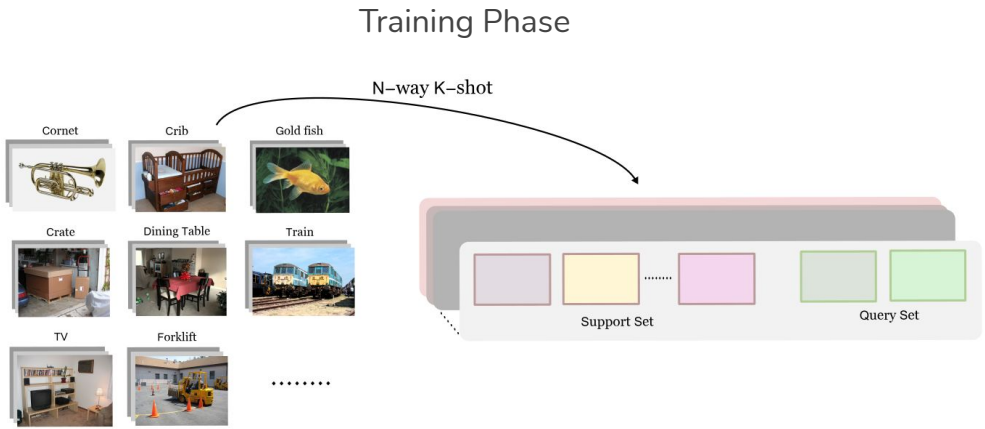




Meta-Learning: initialize then adapt

Training phase has matching conditions compared to the testing phase.

- The training set is transformed into many training tasks.
- The learner is then trained on a distribution of such tasks (episodes).



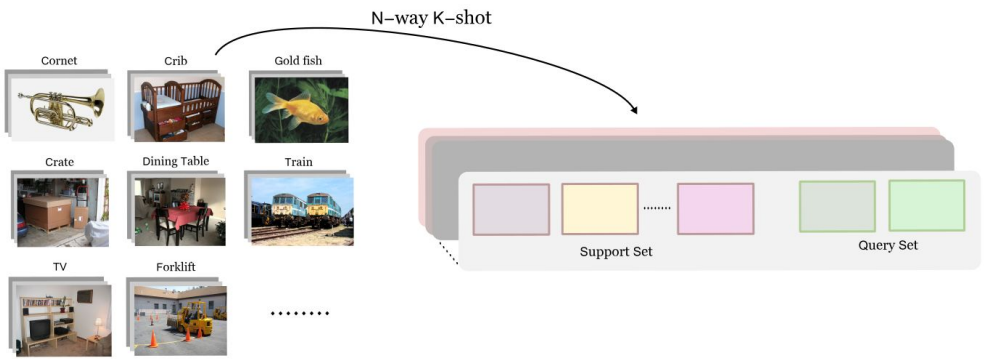
Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, Finn et al. 2017.

Meta-Learning: initialize then adapt

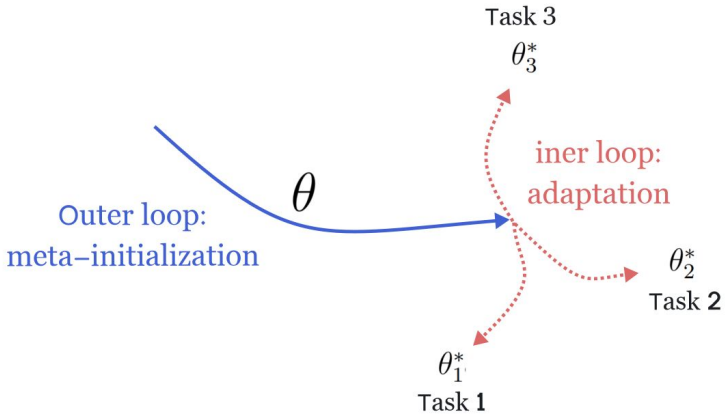
Training phase has matching conditions compared to the testing phase.

- The training set is transformed into many training tasks.
- The learner is then trained on a distribution of such tasks (episodes).

Training Phase



Model-Agnostic Meta-Learning

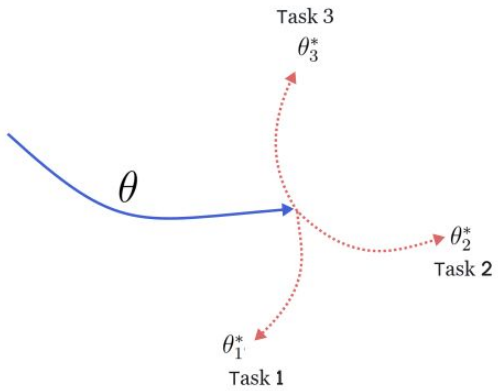


Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, Finn et al. 2017.

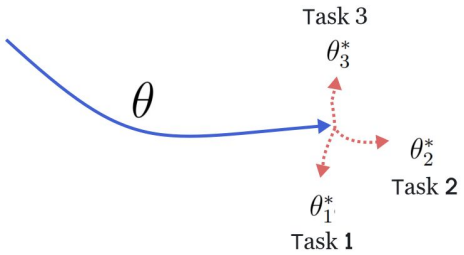
Meta-Learning: rapid learning or feature reuse

Is this inner/outer loop necessary?
How much of the effectiveness of such methods is contingent on the inner/outer loop structure?

Rapid Learning



Feature Reuse

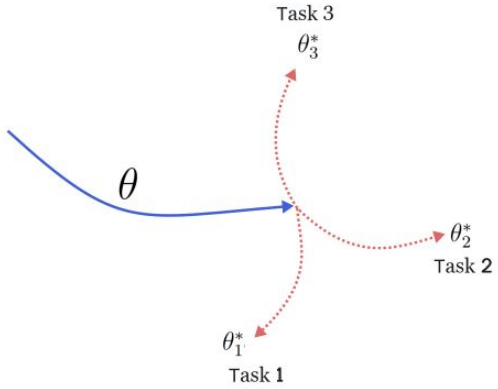


Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. Raghu et al. 2020.

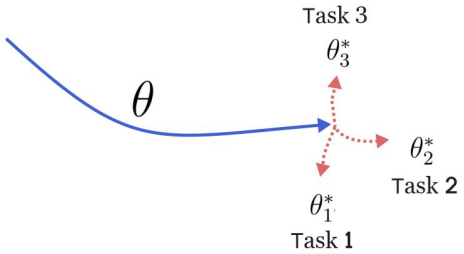
Meta-Learning: rapid learning or feature reuse

Is this inner/outer loop necessary?
How much of the effectiveness of such methods is contingent on the inner/outer loop structure?

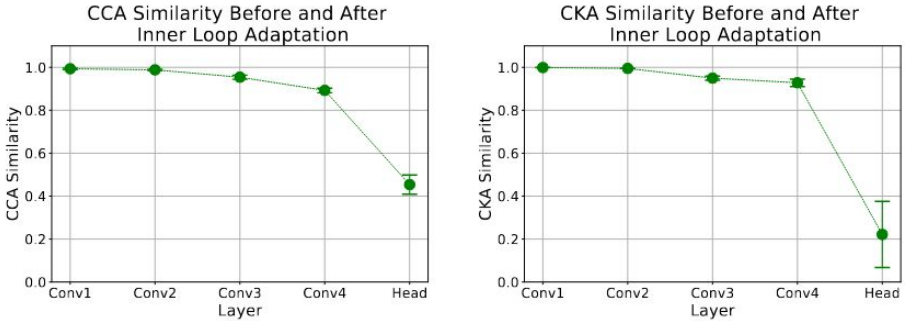
Rapid Learning



Feature Reuse



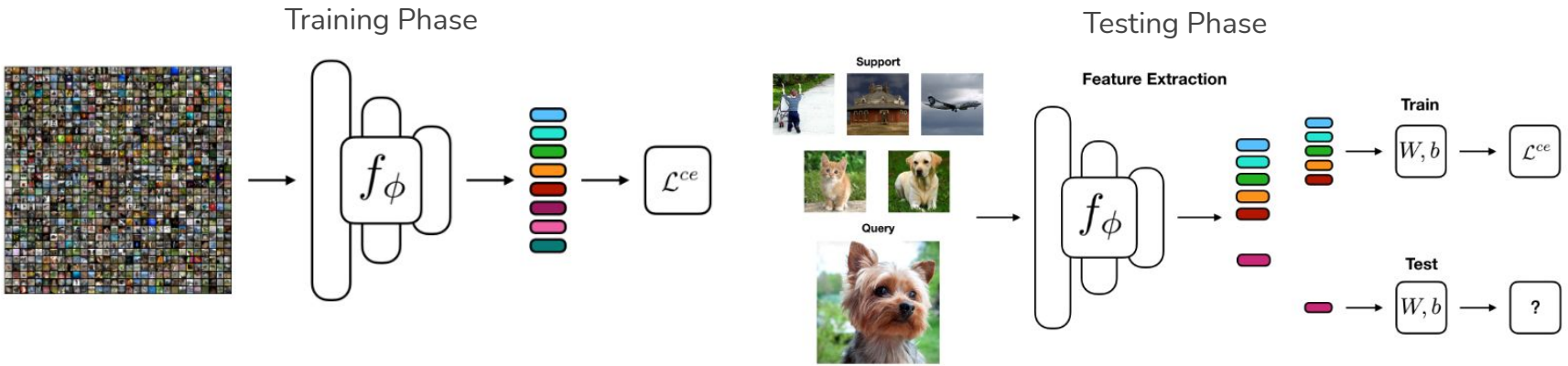
Similarity of the representation before and after adaptation



-> Feature reuse: Having general purpose representations is important for few-shot recognition

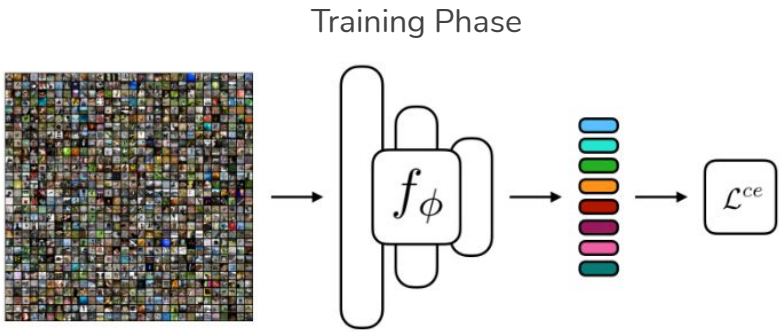
Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. Raghu et al. 2020.

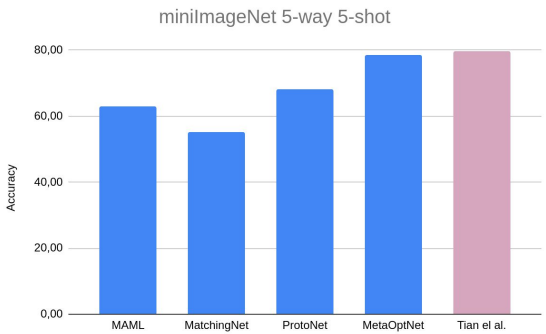
Transfer Learning: a simple and effective baseline

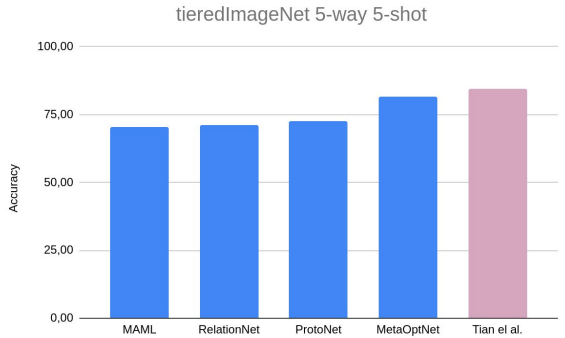


Rethinking Few-Shot Image Classification: a Good Embedding Is All You Need? Tian et al. 2020.

Transfer Learning: a simple and effective baseline

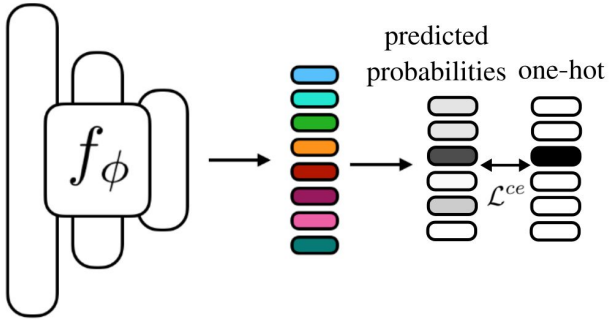






Rethinking Few-Shot Image Classification: a Good Embedding Is All You Need? Tian et al. 2020.

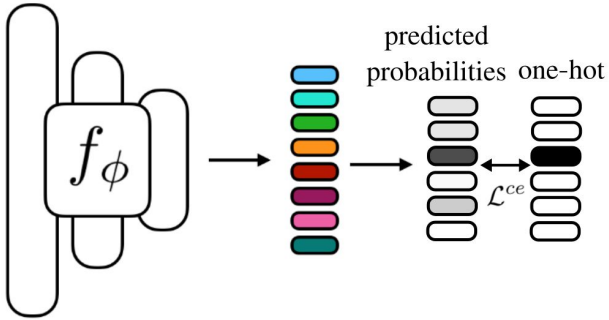
Analysis: the limitations of the CE loss



k-Nearest Neighbors Analysis:

- Train on the whole training set with CE loss.
- Find the nearest neighbors of test instances.

Analysis: the limitations of the CE loss



k-Nearest Neighbors Analysis:

- Train on the whole training set with CE loss.
- Find the nearest neighbors of test instances.

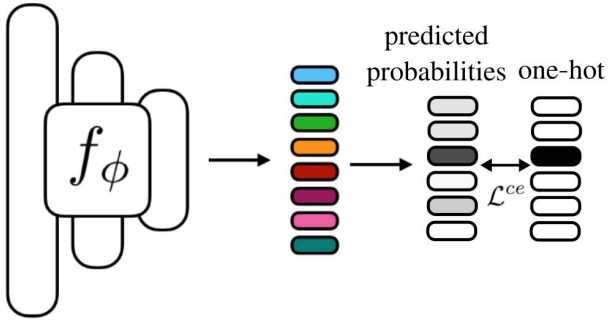
Test Image

cuirass

golden retriever

hourglass

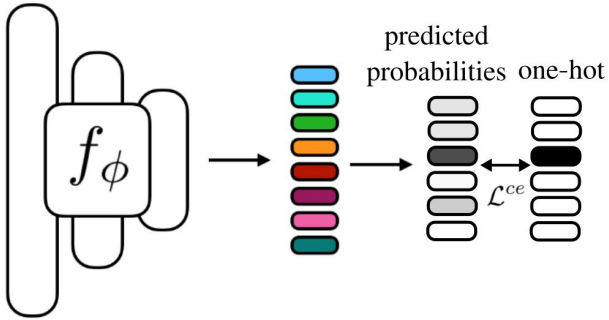
Analysis: the limitations of the CE loss



k-Nearest Neighbors Analysis:

- Train on the whole training set with CE loss.
- Find the nearest neighbors of test instances.

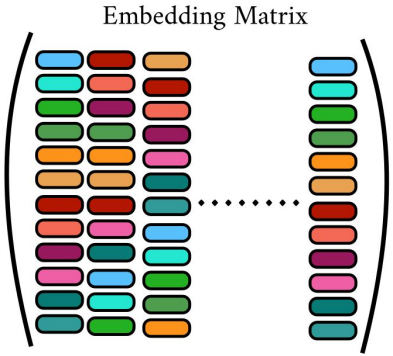
Analysis: the limitations of the CE loss



k-Nearest Neighbors Analysis:

- Train on the whole training set with CE loss.
- Find the nearest neighbors of test instances.

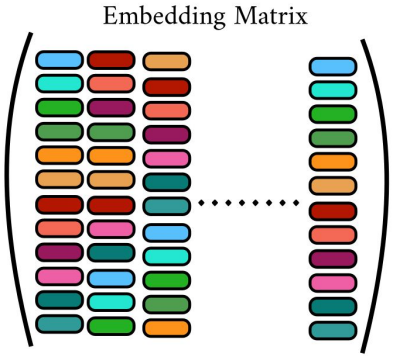
Analysis: the limitations of the CE loss



Spectral Analysis:

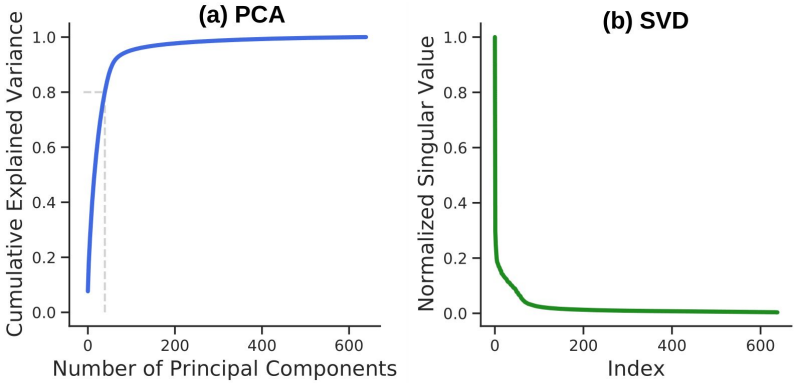
- The explained cumulative variance.
- Normalized singular values.

Analysis: the limitations of the CE loss

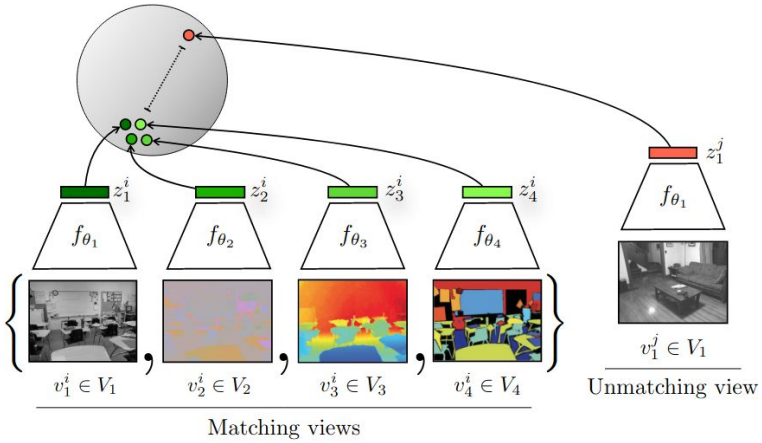


Spectral Analysis:

- The explained cumulative variance.
- Normalized singular values.

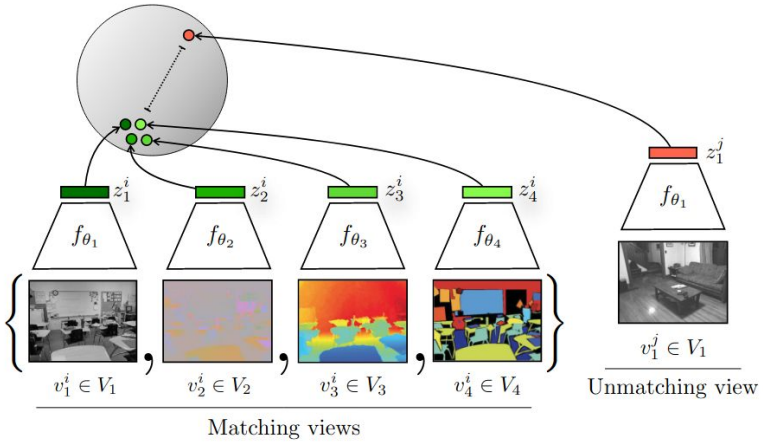


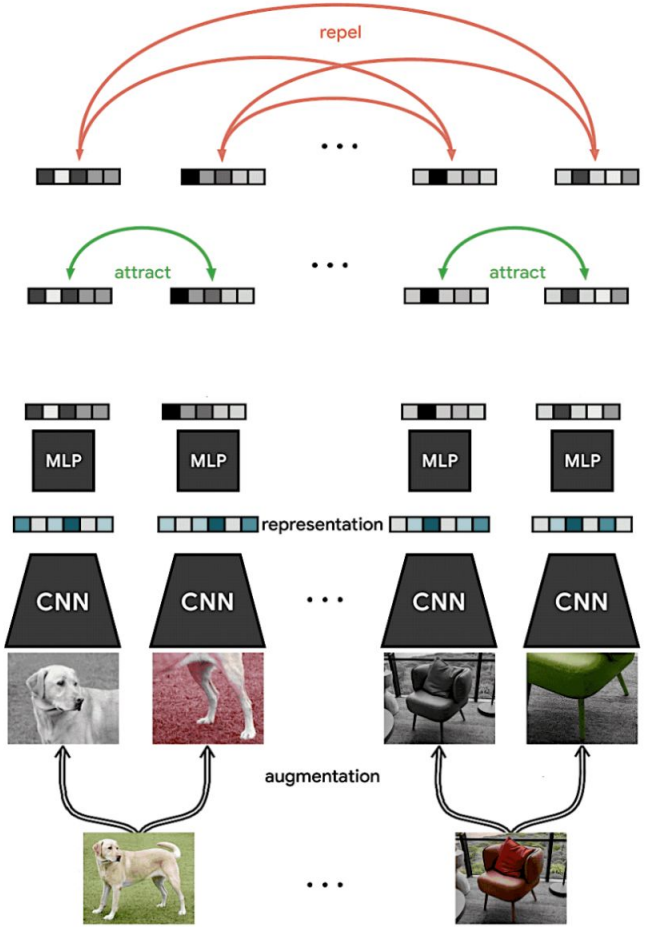
Contrastive Learning: a better pre-training loss



Contrastive Multiview Coding. Tial el al. 2020.
A Simple Framework for Contrastive Learning of Visual Representations. Chen et al. 2020.

Contrastive Learning: a better pre-training loss





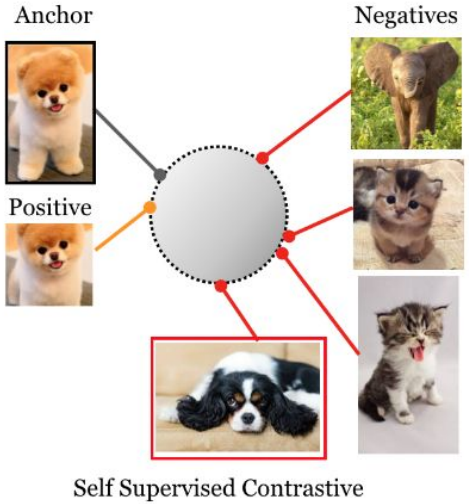
Self-Supervised contrastive loss:

- For each anchor i, attract its positive j against the rest of the examples (negatives).
- The positive examples is a transformed version of the anchor i.

$$-\log \frac{\exp(\text{sim}_g(\mathbf{f}_i, \mathbf{f}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(\text{sim}_g(\mathbf{f}_i, \mathbf{f}_k)/\tau)}$$

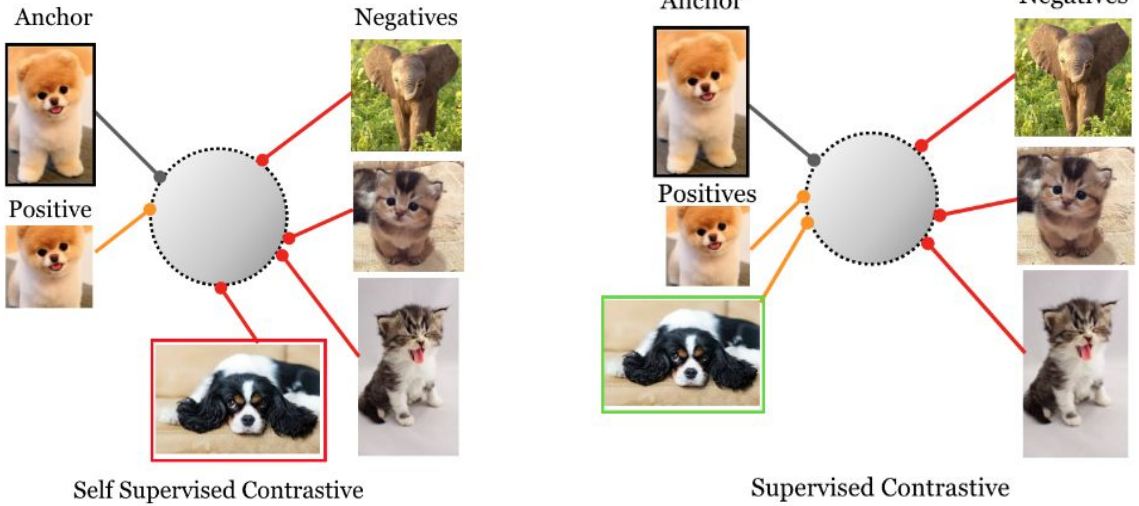
Contrastive Multiview Coding. Tial el al. 2020.
A Simple Framework for Contrastive Learning of Visual Representations. Chen et al. 2020.

Contrastive Learning: a supervised formulation



Supervised Contrastive Learning. Khosla et al. 2020.

Contrastive Learning: a supervised formulation



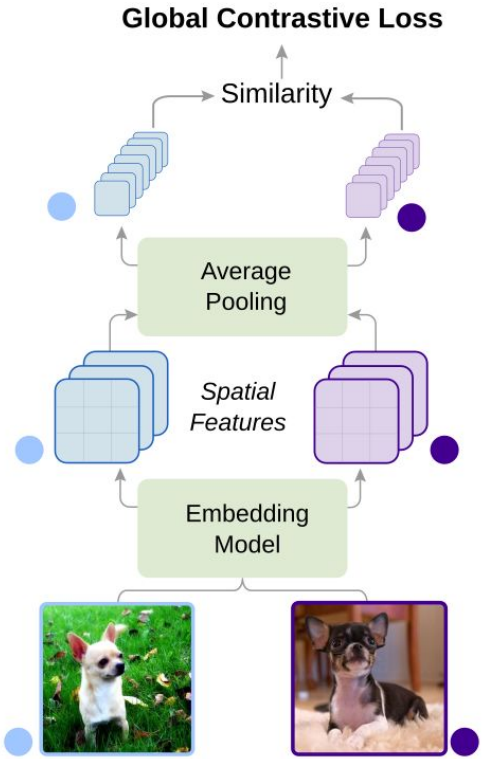
Supervised contrastive loss:

The positives are examples of the same class as the anchor.

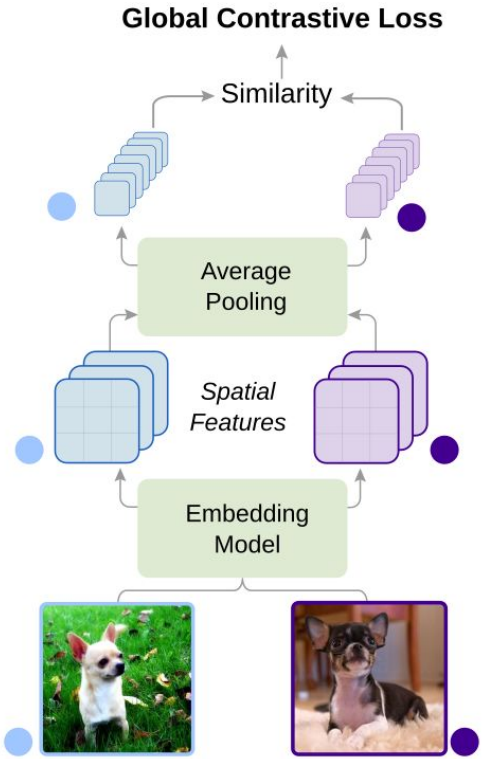
The negatives are examples of different class than the anchor.

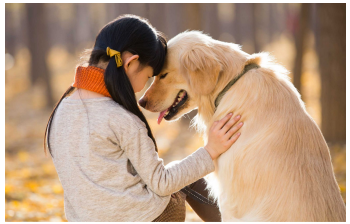
Supervised Contrastive Learning. Khosla et al. 2020.

Contrastive Learning: leveraging the spatial features



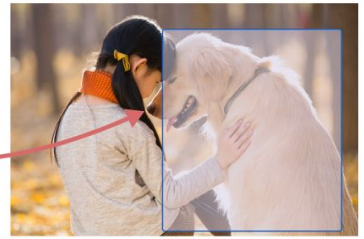
Contrastive Learning: leveraging the spatial features





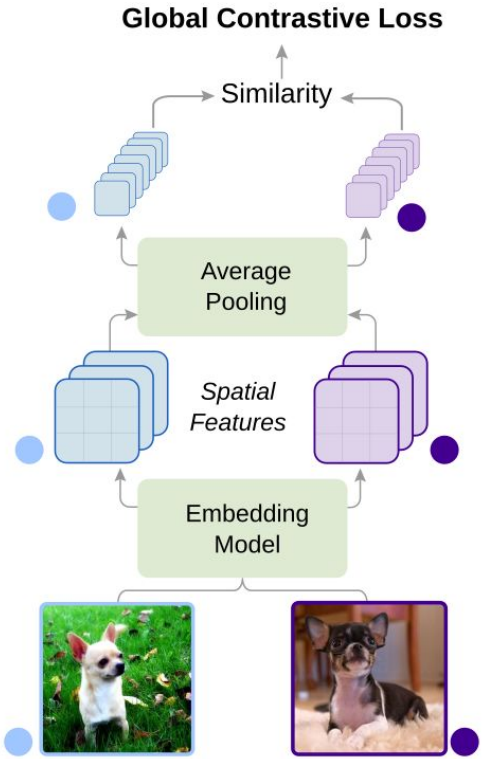
We might align irrelevant objects:

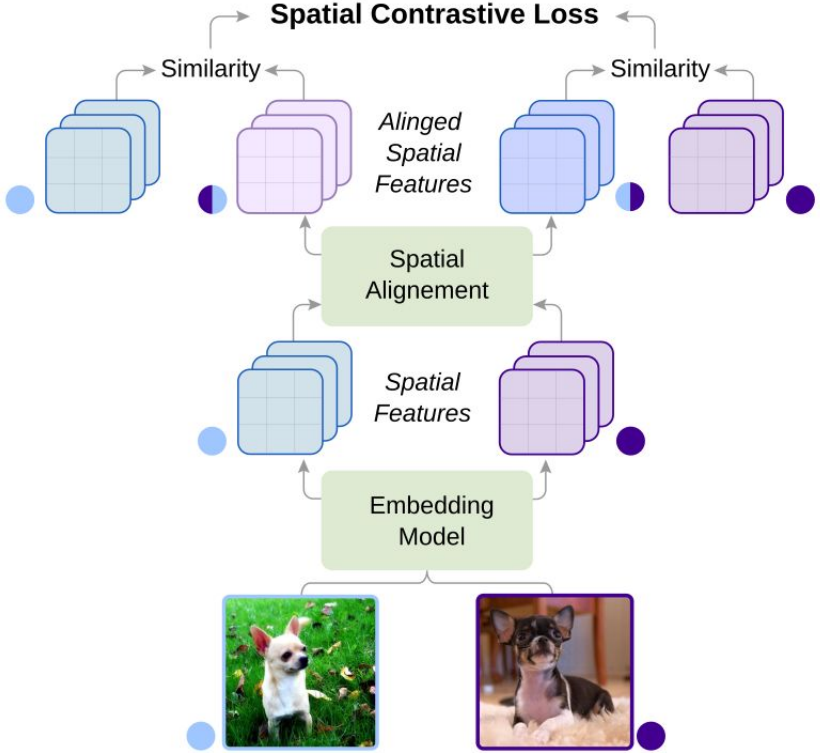
- Dog & person
- Dog & grass



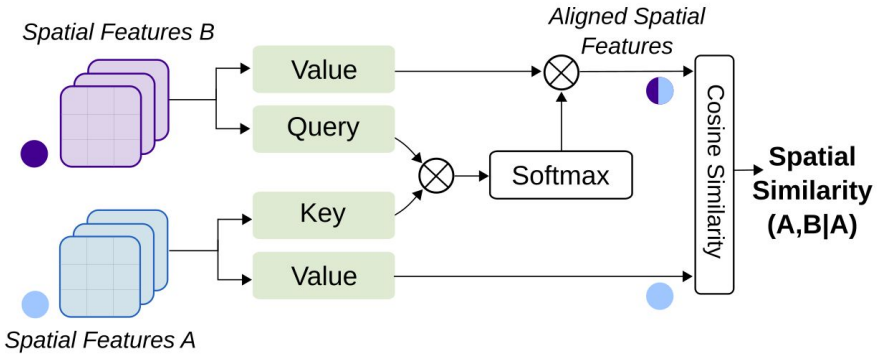
-> Use the spatial features

Spatial Contrastive Learning: introduction

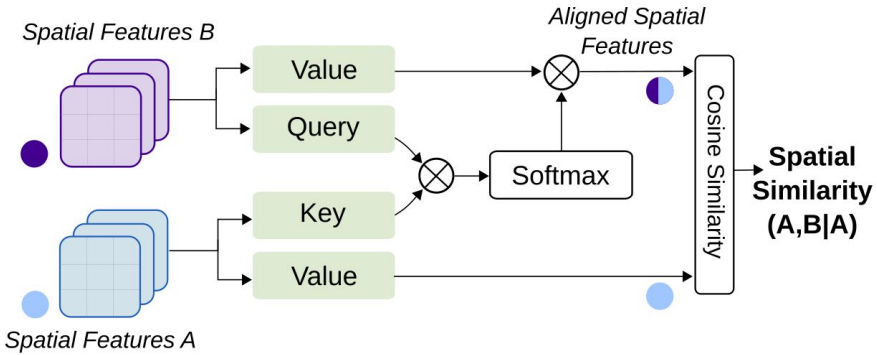


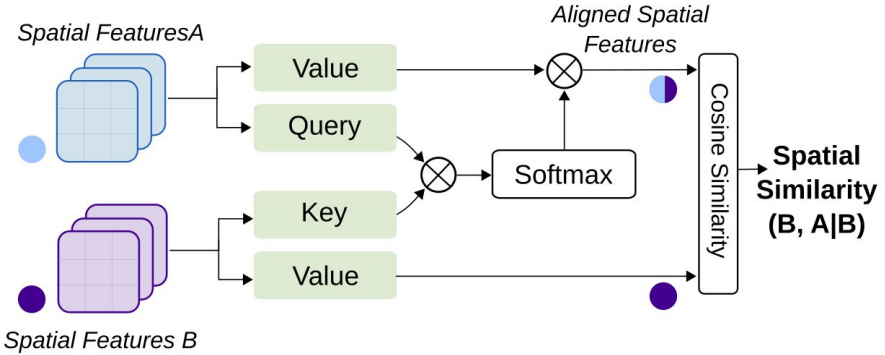


Spatial Contrastive Learning: spatial alignment

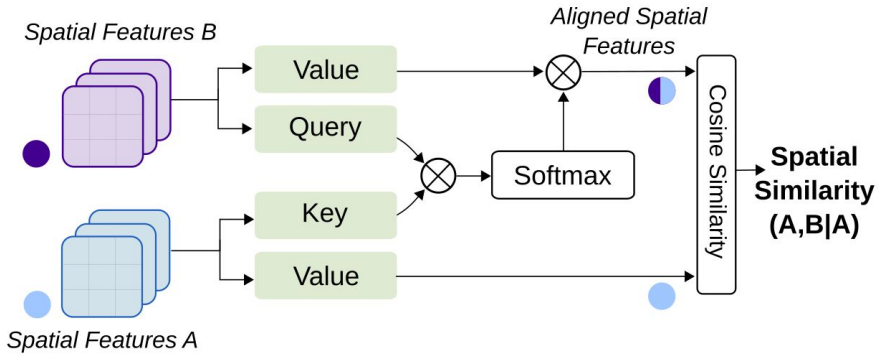


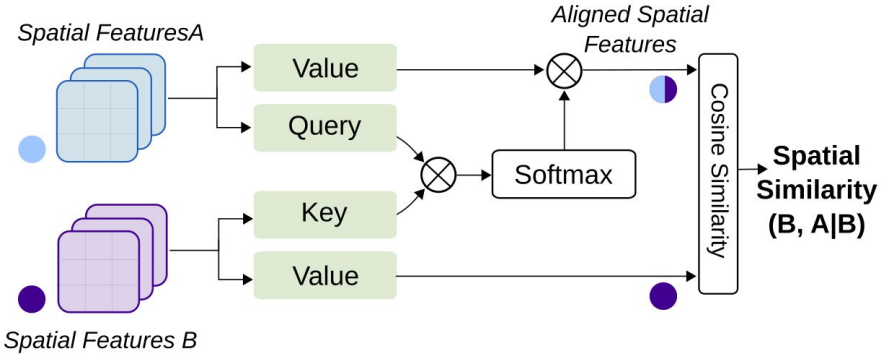
Spatial Contrastive Learning: spatial alignment





Spatial Contrastive Learning: spatial alignment

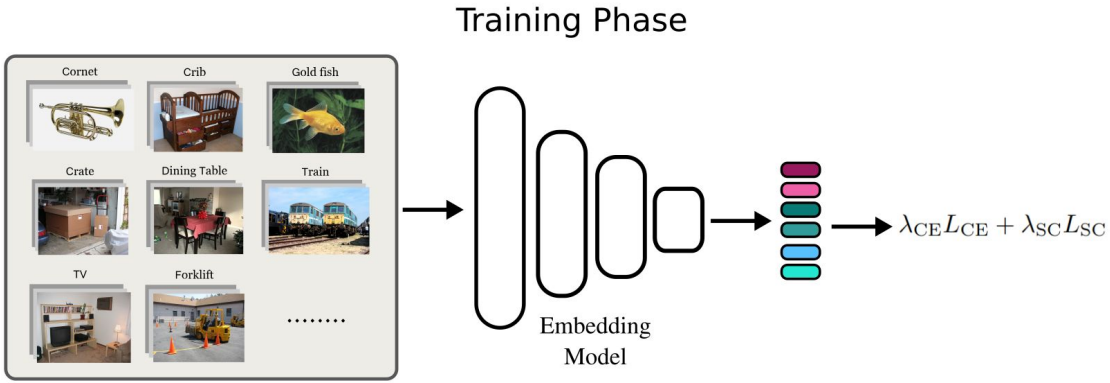




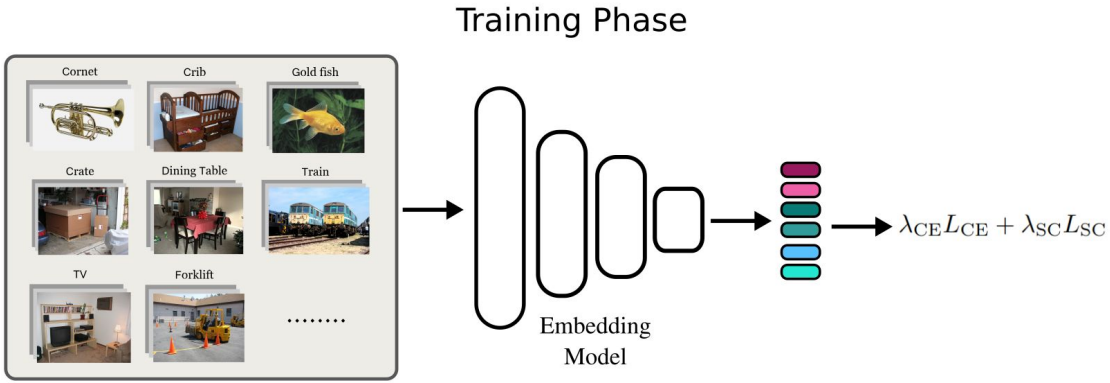
Spatial Similarity (A, B) = Spatial Similarity (A, B|A) + Spatial Similarity (B, A|B)

$$\ell_{ij} = -\log \frac{\exp(\mathrm{sim}_{\mathrm{s}}(\mathbf{z}_i^{\mathrm{s}}, \mathbf{z}_j^{\mathrm{s}})/\tau')}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(\mathrm{sim}_{\mathrm{s}}(\mathbf{z}_i^{\mathrm{s}}, \mathbf{z}_k^{\mathrm{s}})/\tau')},$$

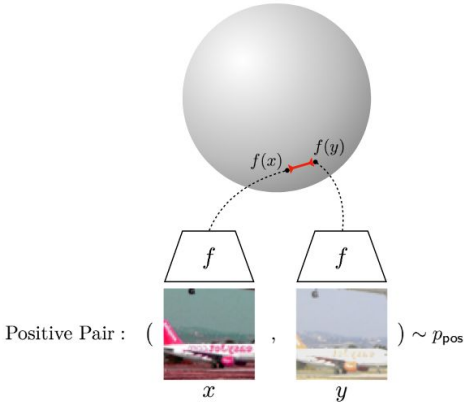
Spatial Contrastive Learning: standard transfer learning



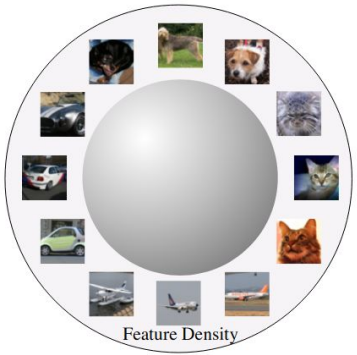
Spatial Contrastive Learning: standard transfer learning



Contrastive Distillation: an additional training step



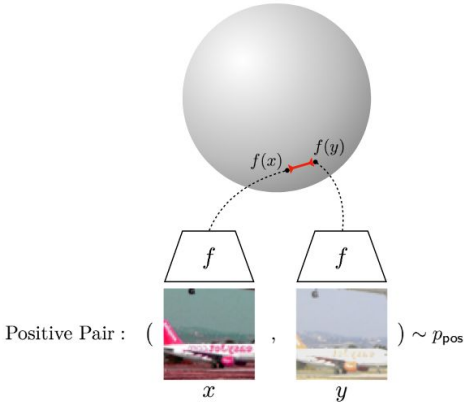
Alignment: Similar samples have similar features.



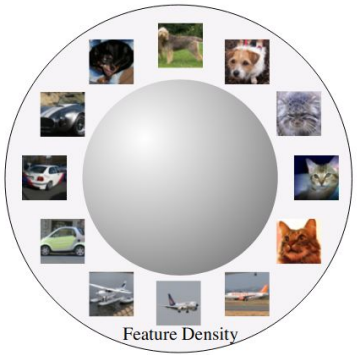
Uniformity: Preserve maximal information.

Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. Wang et al. 2020.

Contrastive Distillation: an additional training step



Alignment: Similar samples have similar features.

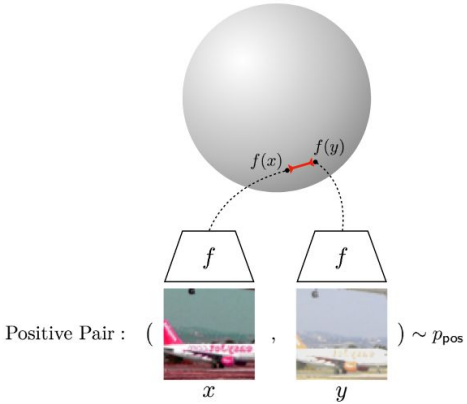


Uniformity: Preserve maximal information.

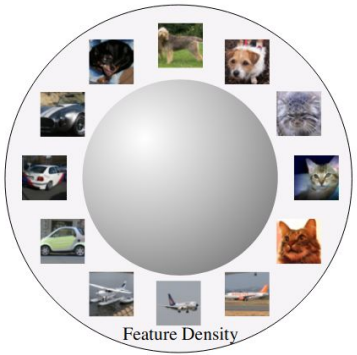
A possible over clustering of the features of the same class.

Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. Wang et al. 2020.

Contrastive Distillation: an additional training step



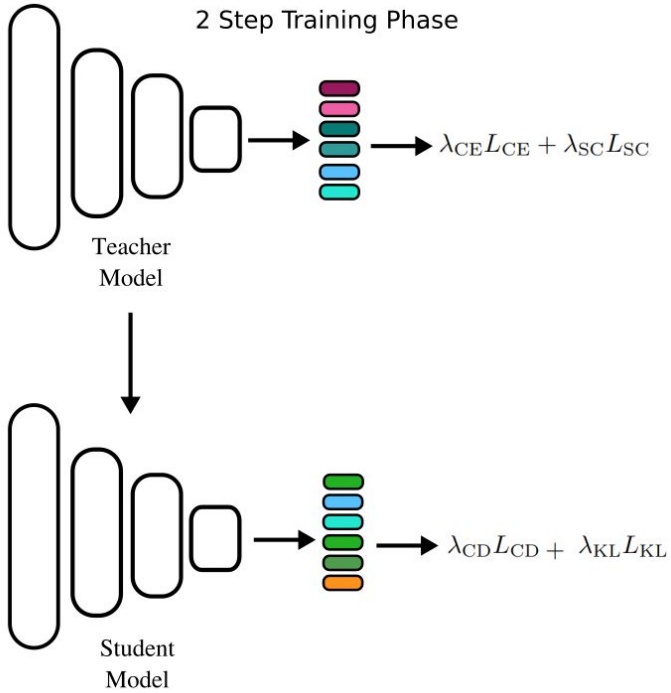
Alignment: Similar samples have similar features.



Uniformity: Preserve maximal information.

Possible over clustering of the features of the same class.

-> Contrastive alignement distillation without uniformity $L_{\text{CD}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{z}_i^{gt} - \mathbf{z}_i^{gs}\|_2^2$



Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. Wang et al. 2020.

Results: ablations & miniImageNet

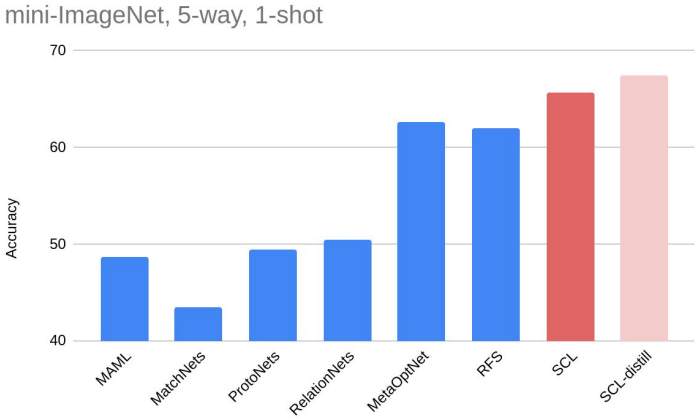
Loss Function	Aug.	*mini*-ImageNet, 5-way 1-shot	*mini*-ImageNet, 5-way 5-shot	CIFAR-CS, 5-way 1-shot	CIFAR-CS, 5-way 5-shot
CE		61.8 ± 0.7	79.7 ± 0.6	71.3 ± 0.9	86.1 ± 0.6
CE	✓	61.8 ± 0.8	78.6 ± 0.5	71.9 ± 0.9	86.3 ± 0.5
CE + SS-GC	✓	62.7 ± 0.7	81.0 ± 0.6	70.9 ± 0.9	84.5 ± 0.6
CE + SS-SC	✓	64.0 ± 0.8	81.5 ± 0.5	72.1 ± 0.8	86.2 ± 0.6
CE + SS-GC + SS-SC	✓	62.8 ± 0.8	81.1 ± 0.6	69.0 ± 0.9	85.0 ± 0.6
CE + GC	✓	65.0 ± 0.8	81.6 ± 0.5	74.0 ± 0.8	87.3 ± 0.6
CE + SC	✓	$\mathbf{65.7 \pm 0.8}$	$\mathbf{82.5 \pm 0.5}$	75.0 ± 0.9	87.4 ± 0.6
CE + GC + SC	✓	65.0 ± 0.8	81.3 ± 0.5	$\mathbf{76.0 \pm 0.7}$	$\mathbf{87.5 \pm 0.5}$

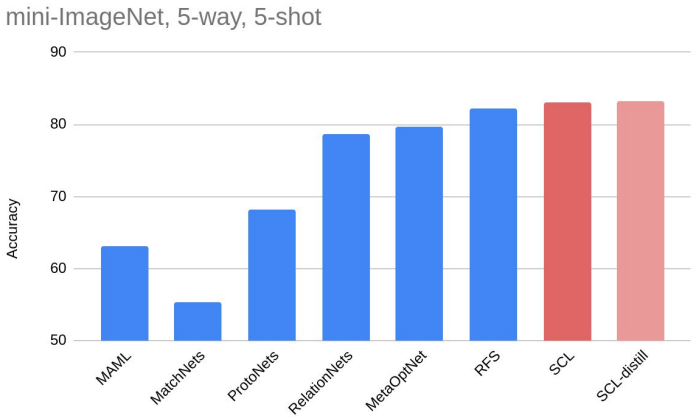
-> Optimizing the spatial features yields in better results.

Results: ablations & miniImageNet

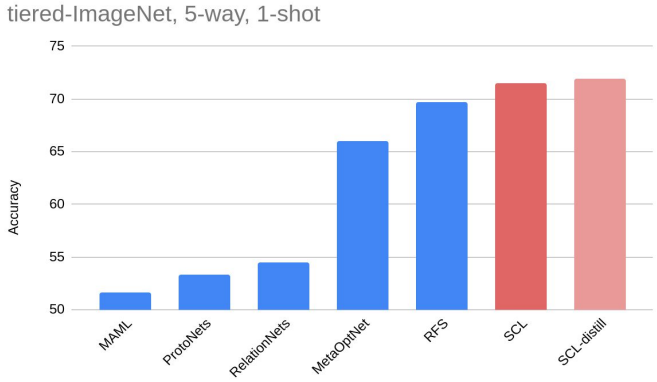
Loss Function	Aug.	*mini*-ImageNet, 5-way 1-shot	*mini*-ImageNet, 5-way 5-shot	CIFAR-CS, 5-way 1-shot	CIFAR-CS, 5-way 5-shot
CE		61.8 ± 0.7	79.7 ± 0.6	71.3 ± 0.9	86.1 ± 0.6
CE	✓	61.8 ± 0.8	78.6 ± 0.5	71.9 ± 0.9	86.3 ± 0.5
CE + SS-GC	✓	62.7 ± 0.7	81.0 ± 0.6	70.9 ± 0.9	84.5 ± 0.6
CE + SS-SC	✓	64.0 ± 0.8	81.5 ± 0.5	72.1 ± 0.8	86.2 ± 0.6
CE + SS-GC + SS-SC	✓	62.8 ± 0.8	81.1 ± 0.6	69.0 ± 0.9	85.0 ± 0.6
CE + GC	✓	65.0 ± 0.8	81.6 ± 0.5	74.0 ± 0.8	87.3 ± 0.6
CE + SC	✓	$\mathbf{65.7 \pm 0.8}$	$\mathbf{82.5 \pm 0.5}$	75.0 ± 0.9	87.4 ± 0.6
CE + GC + SC	✓	65.0 ± 0.8	81.3 ± 0.5	$\mathbf{76.0 \pm 0.7}$	$\mathbf{87.5 \pm 0.5}$

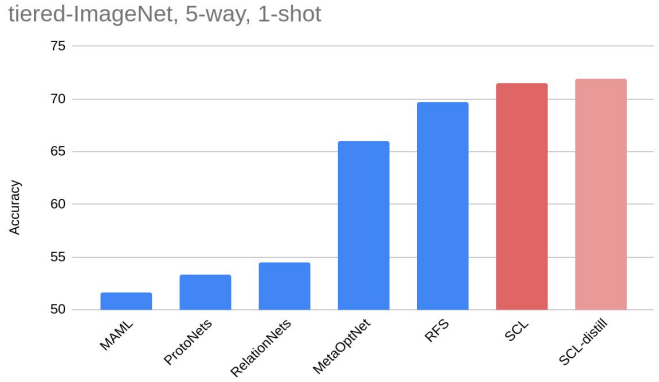
-> Optimizing the spatial features yields in better results.



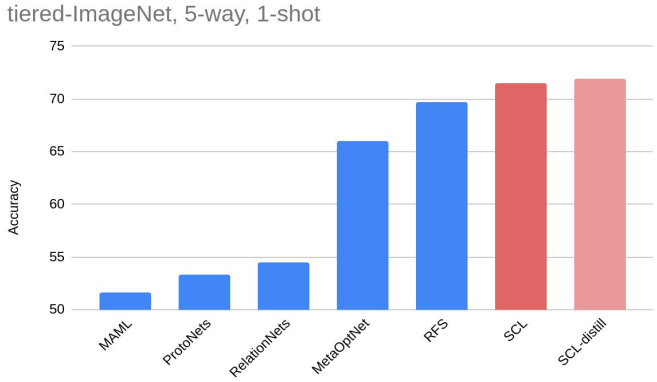


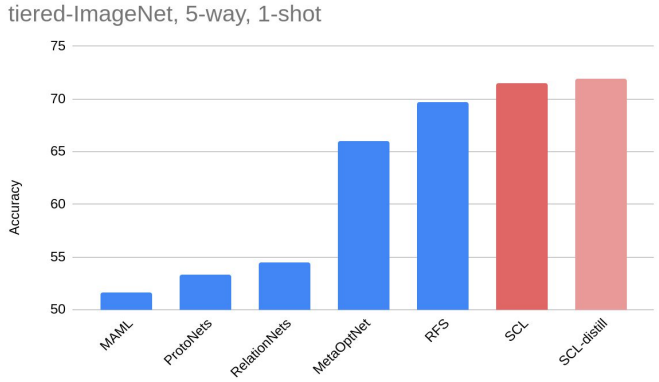
Results: tieredImageNet & cross-domain few-shot



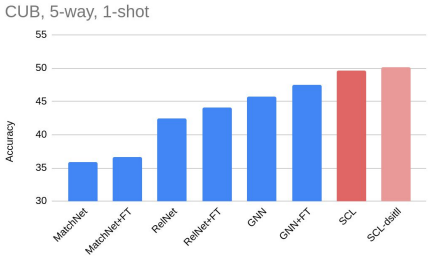


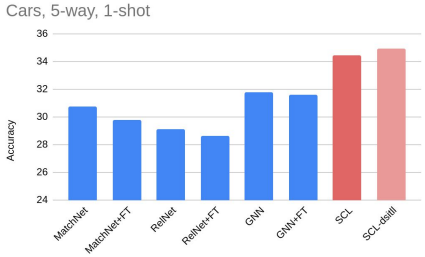
Results: tieredImageNet & cross-domain few-shot

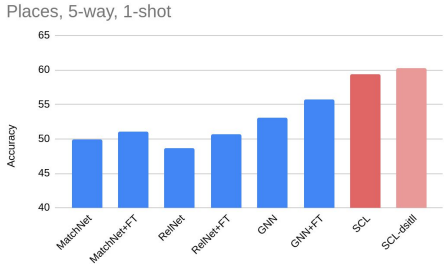


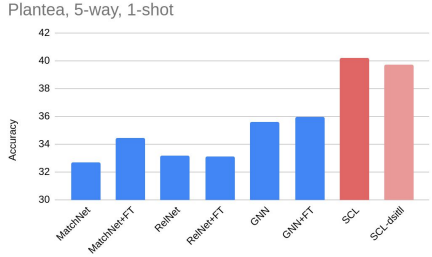


Cross-domain few-shot classification: Train on miniImageNet and test on a new domain (Cub, Cars, Places, or Plantae)

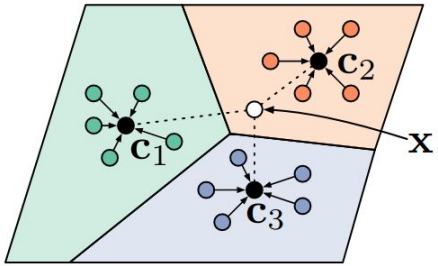






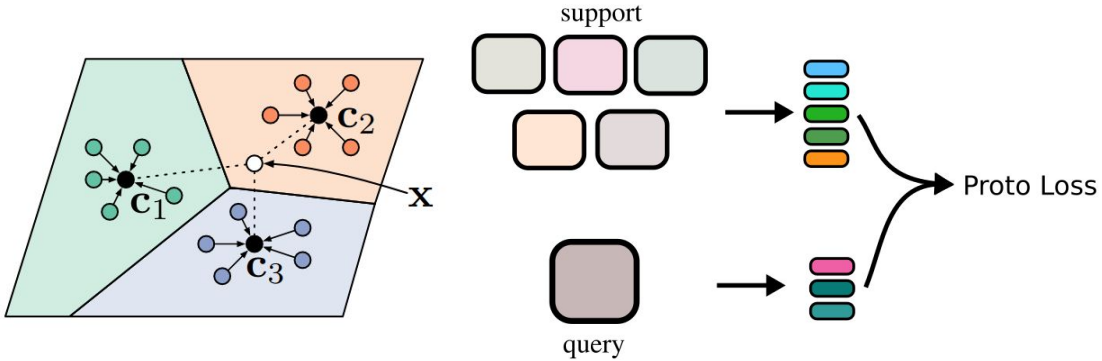


Results: prototypical networks



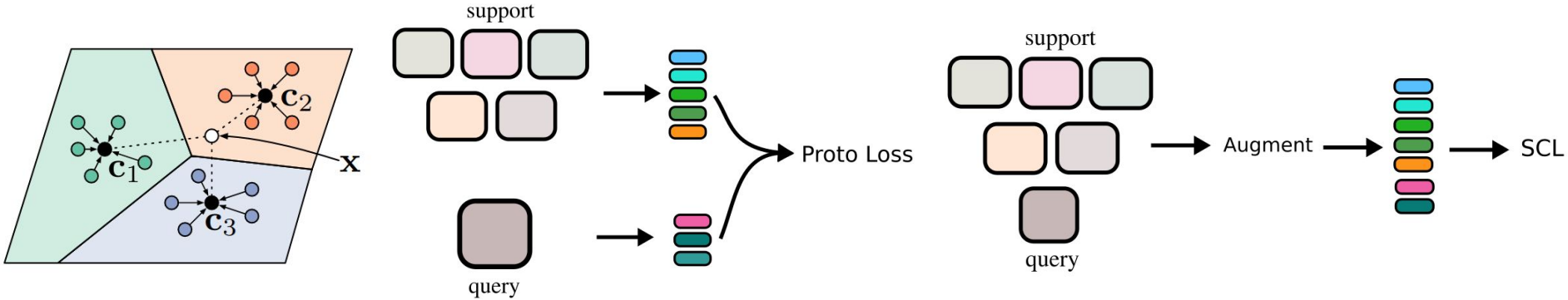
Prototypical Networks for Few-shot Learning. Snell el al. 2017.

Results: prototypical networks



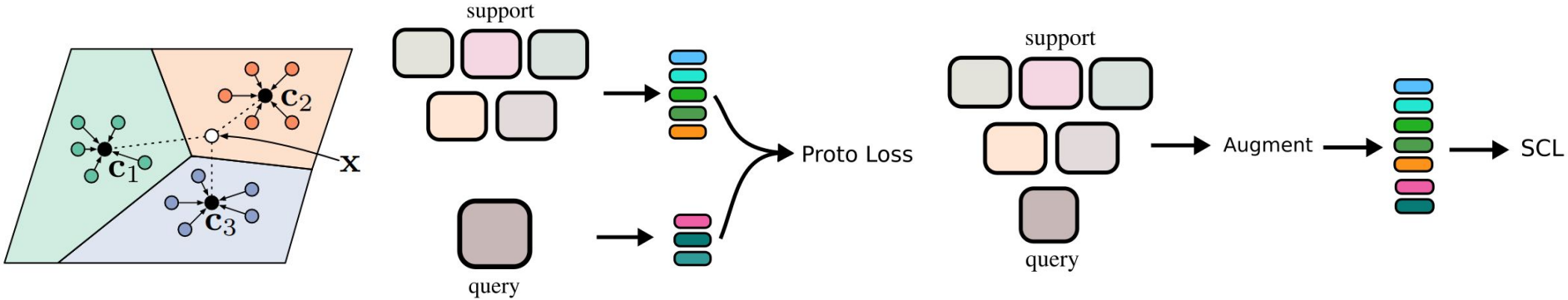
Prototypical Networks for Few-shot Learning. Snell el al. 2017.

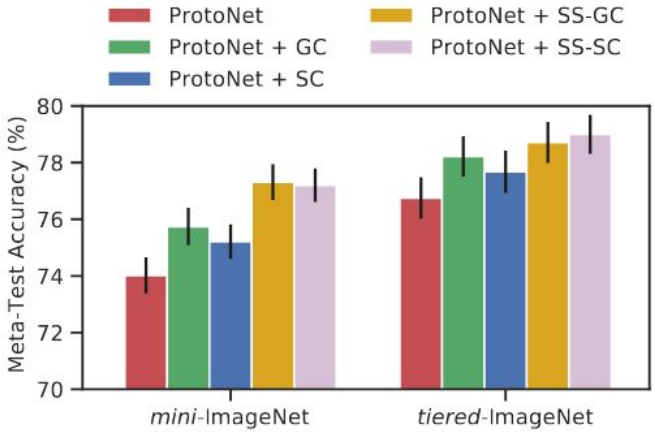
Results: prototypical networks

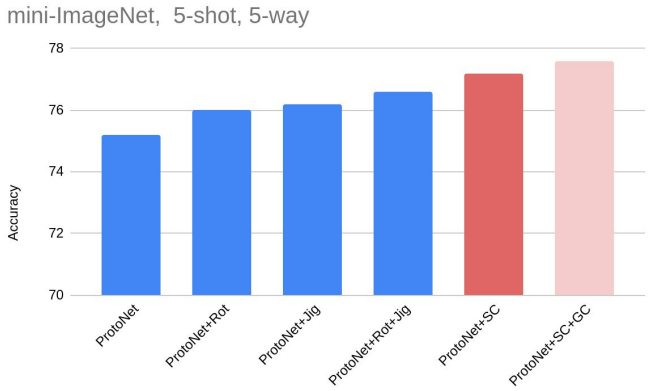


Prototypical Networks for Few-shot Learning. Snell el al. 2017.

Results: prototypical networks







Prototypical Networks for Few-shot Learning. Snell el al. 2017.

Conclusion

In this work, we:

- Explored the usage of contrastive learning for few-Shot classification to learn more general purpose representations.

- Proposed a novel Spatial contrantrasitve loss to further promotes class-independent discriminative patterns.

- Proposed a contrastive distillation step to relax the placement of the embeddings and enhance the learned representations.

- Demonstrated the effectiveness of the proposed method on different datasets, settings and frameworks.

For more details please see the paper / code.

Thank You!

For more details please visit the project's webpage:

https://yassouali.github.io/SCL_page/

Paris-Saclay University, CentraleSupélec, MICS