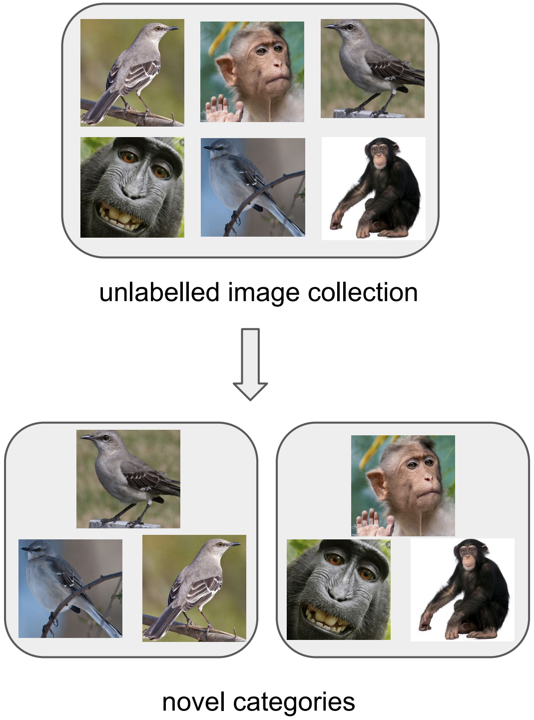
Novel Visual Category Discovery with Dual Ranking Statistics and Mutual Knowledge Distillation

Bingchen Zhao[1] and Kai Han[2,3,4]

[1]Tongji University [2]The University of Hong Kong [3]Google Research [4]University of Bristol

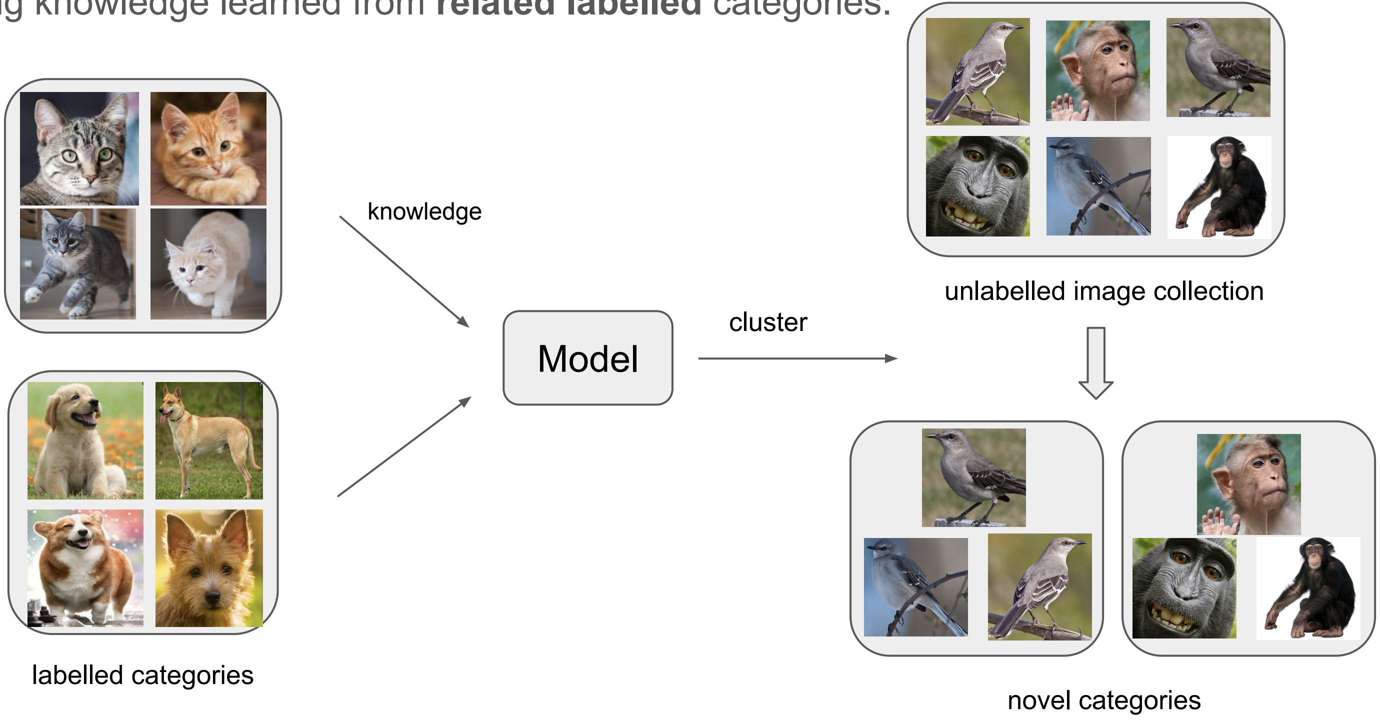
Problem definition: novel category discovery

- Discover/cluster **novel** categories in an unlabelled image collection

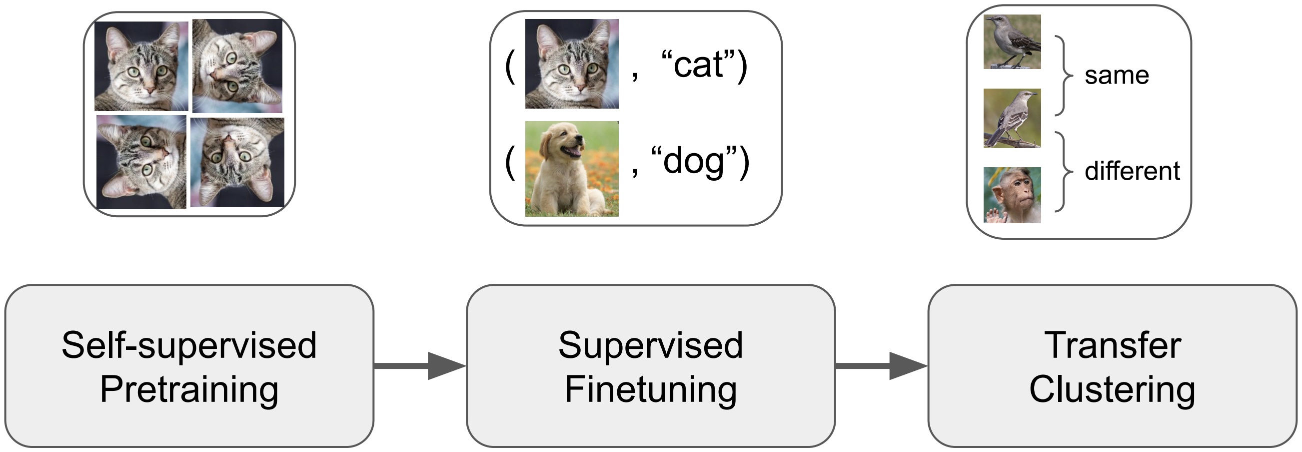


Problem definition: novel category discovery

- Discover/cluster **novel** categories in an unlabelled image collection
- Using knowledge learned from **related labelled** categories.



Pipeline



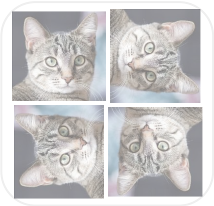
To learn class-agnostic **low-level** features.

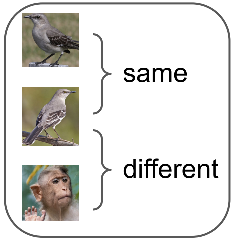
Finetune the network for meaningful **high-level** features.

Transfer the knowledge to discover **novel categories.**

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021
[2] Neighborhood Contrastive Learning for Novel Class Discovery, CVPR 2021

Pipeline





Self-supervised Pretraining → Supervised Finetuning → Transfer Clustering

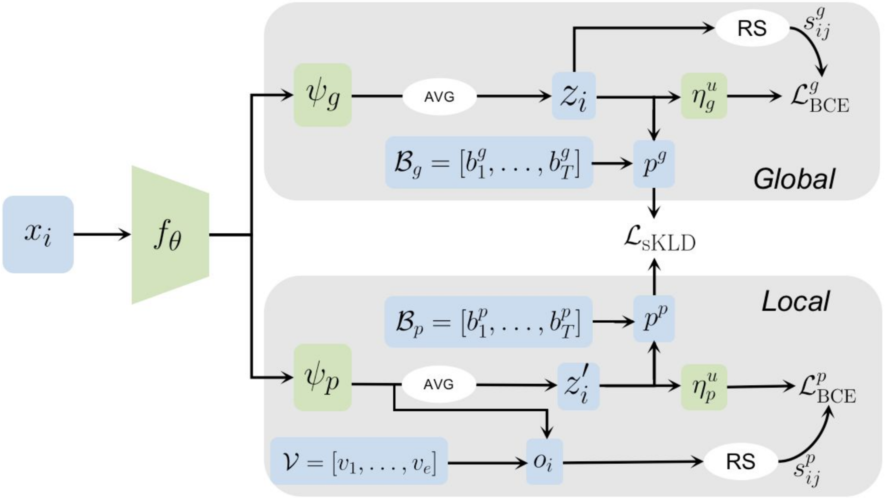
To learn class-agnostic **low-level** features.

Finetune the network for meaningful **high-level** features.

Transfer the knowledge to discover **novel categories.**

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021
[2] Neighborhood Contrastive Learning for Novel Class Discovery, CVPR 2021

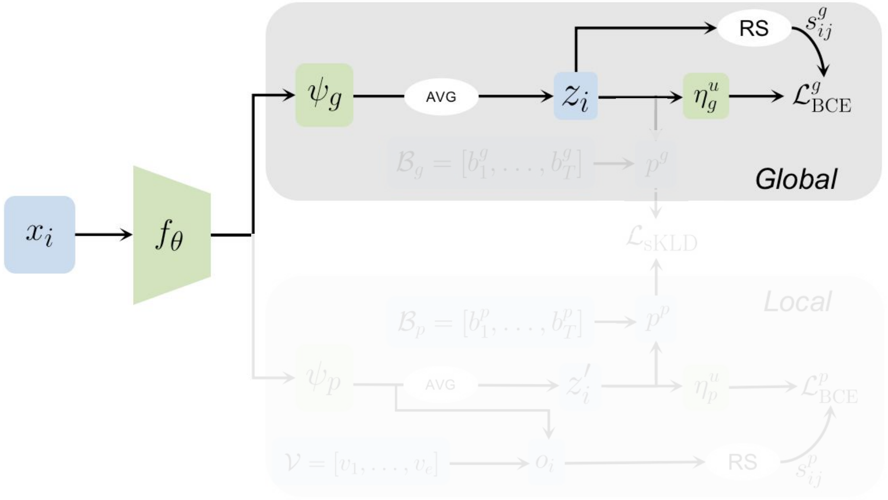
Method overview



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*. [1]
- **Mutual knowledge distillation**: allow *information exchange* between local and global.

[1] Unifying deep local and global features for image search, ECCV 2020

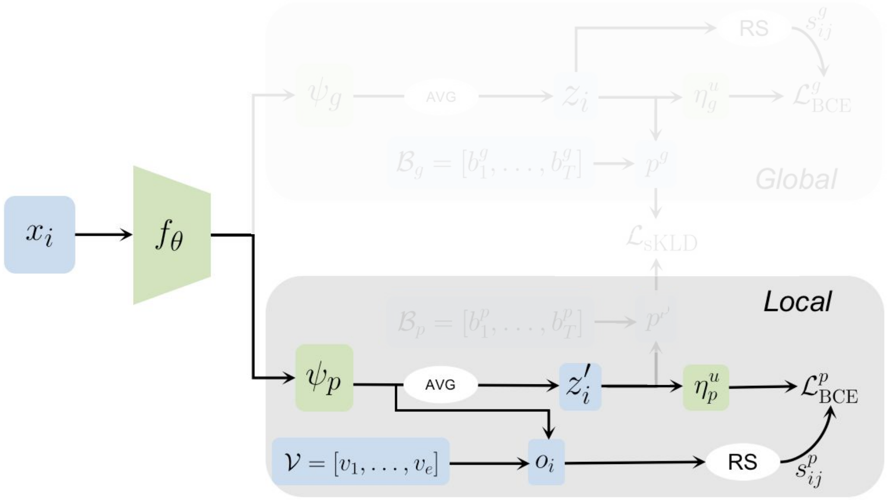
Method overview



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.

[1] Unifying deep local and global features for image search, ECCV 2020

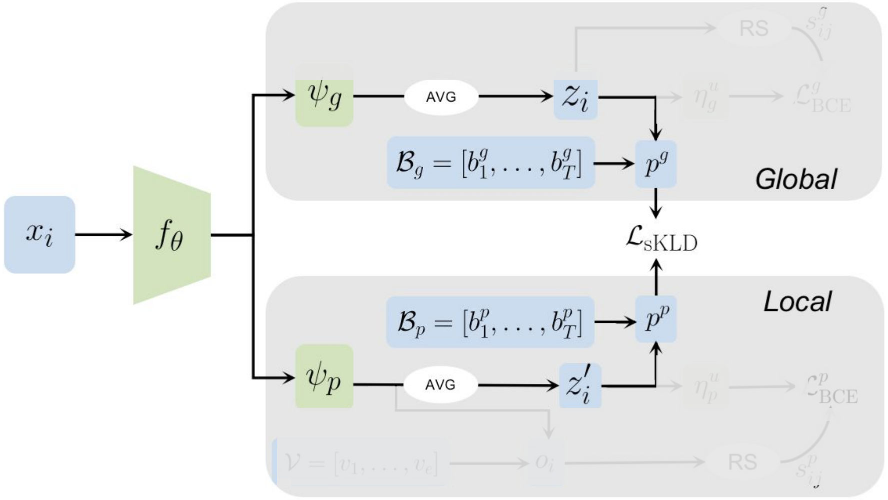
Method overview



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*. [1]

[1] Unifying deep local and global features for image search, ECCV 2020

Method overview

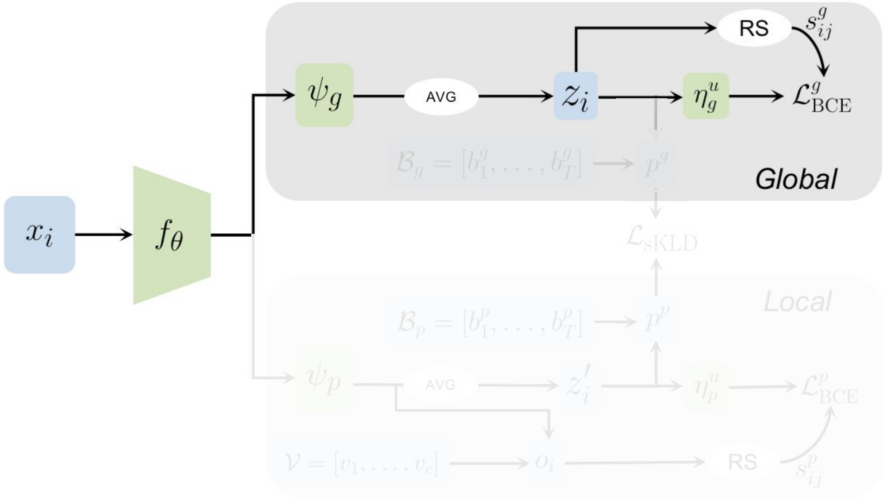


- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*. [1]
- **Mutual knowledge distillation**: allow *information exchange* between local and global.

[1] Unifying deep local and global features for image search, ECCV 2020

NEURAL INFORMATION PROCESSING SYSTEMS

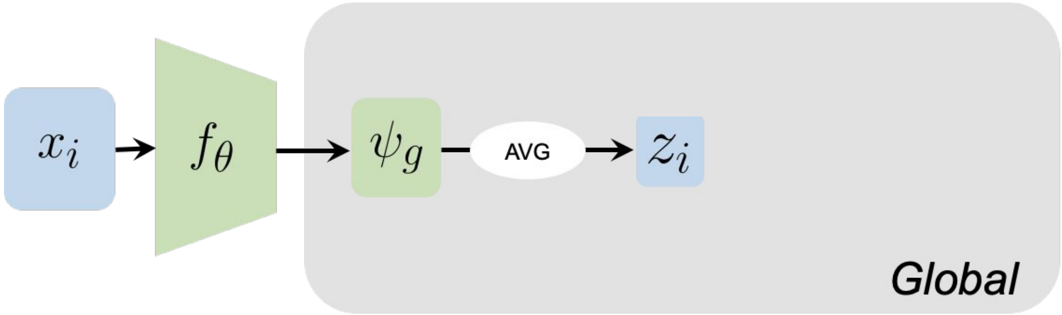
Method overview



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.

[1] Unifying deep local and global features for image search, ECCV 2020

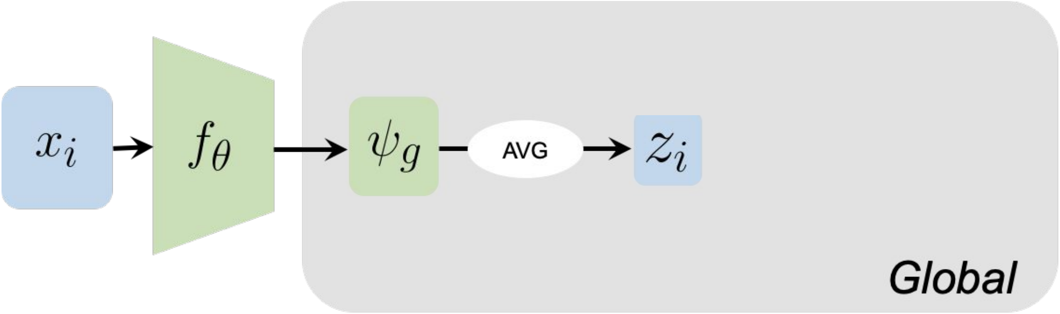
Dual ranking statistics: global branch



Global branch follows AutoNovel[1].

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021

Dual ranking statistics: global branch

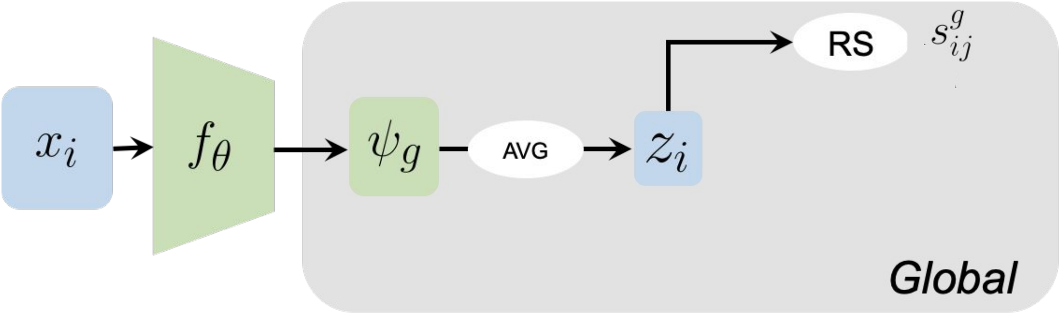


Global branch follows AutoNovel[1].

- Represent input as a single feature vector (use avgpool, AVG).

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021

Dual ranking statistics: global branch



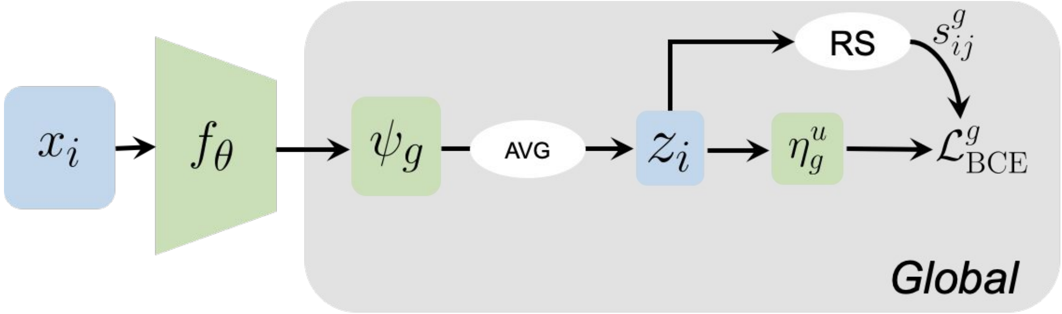
Global branch follows AutoNovel[1].

- Represent input as a single feature vector (use avgpool, AVG).
- Obtain pair-wise pseudo label using Ranking Statistics (RS).

$$s_{ij} = \mathrm{RS}(z_i, z_j) = \mathbb{1}\left\{\mathrm{top}_k(z_i) = \mathrm{top}_k(z_j)\right\}$$

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021

Dual ranking statistics: global branch



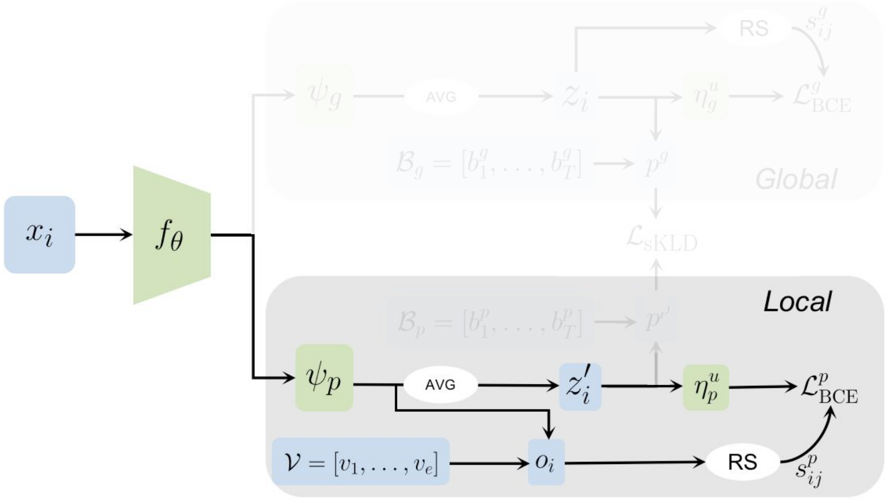
Global branch follows AutoNovel[1].

- Represent input as a single feature vector (use avgpool, AVG).
- Obtain pair-wise pseudo label using Ranking Statistics (RS).
- Train the network with binary cross-entropy loss (BCE).

[1] AutoNovel: Automatically Discovering and Learning Novel Visual Categories, TPAMI 2021

NEURAL INFORMATION PROCESSING SYSTEMS

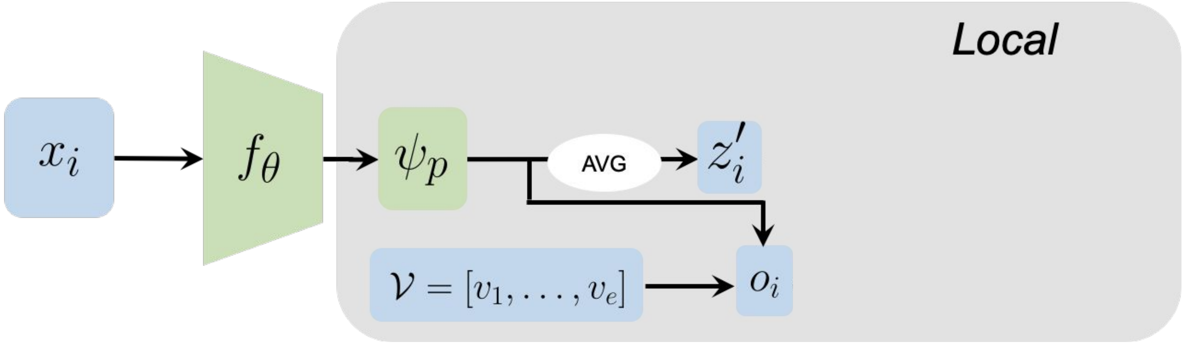
Method overview



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*. [1]

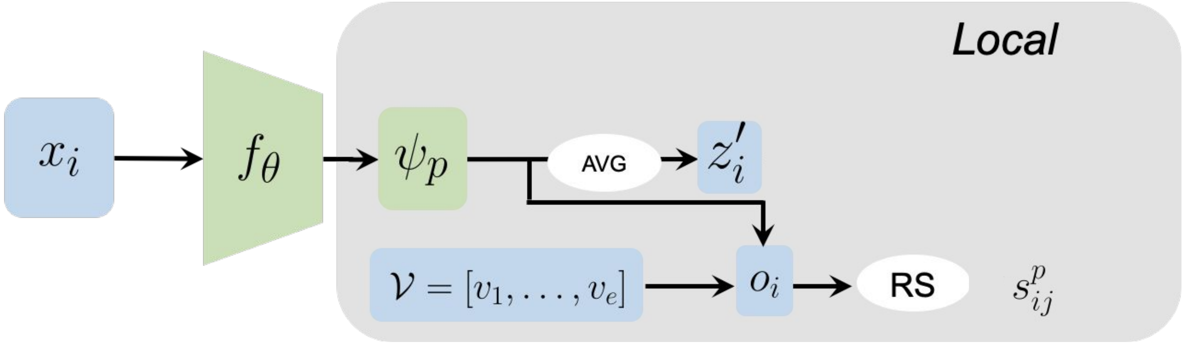
[1] Unifying deep local and global features for image search, ECCV 2020

Dual ranking statistics: local branch



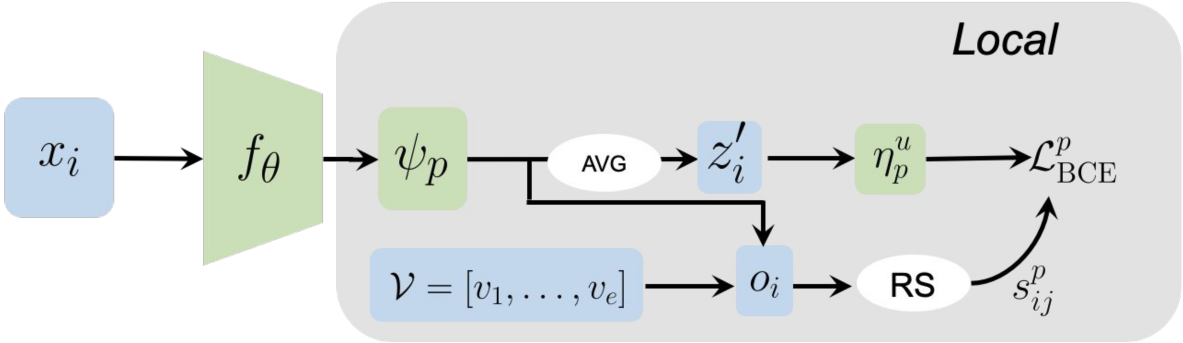
- Leverage the part dictionary to fuse local information on the feature map before avgpool.

Dual ranking statistics: local branch



- Leverage the part dictionary to fuse local information on the feature map before avgpool.
- Generate pair-wise pseudo labels using RS on the fused vectors.

Dual ranking statistics: local branch

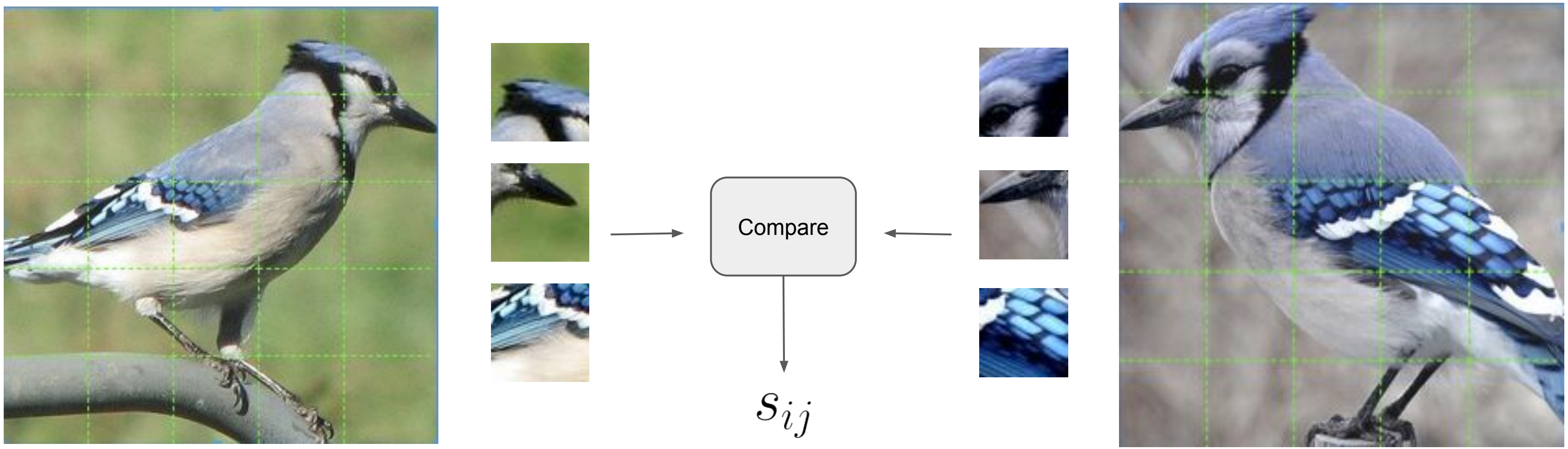


- Leverage the part dictionary to fuse local information on the feature map before avgpool.
- Generate pair-wise pseudo labels using RS on the fused vectors.
- Train the network with binary cross-entropy loss.

Dual ranking statistics: local branch

Generating more restrict pseudo-labels by **comparing local parts.**

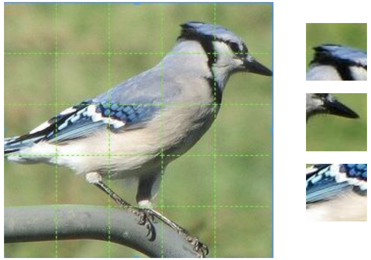
If **one vector** can contain all local informations, we can still use **RS.**



Dual ranking statistics: local branch

Part Dictionary

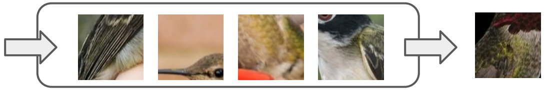
$$\mathcal{V} = \{v_1, v_2, ..., v_e\}$$



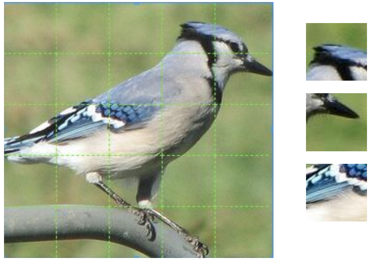
- Maintain a dictionary of local features
- Update it using local features of each new batches of image
- Update in a FIFO manner.

Dual ranking statistics: local branch

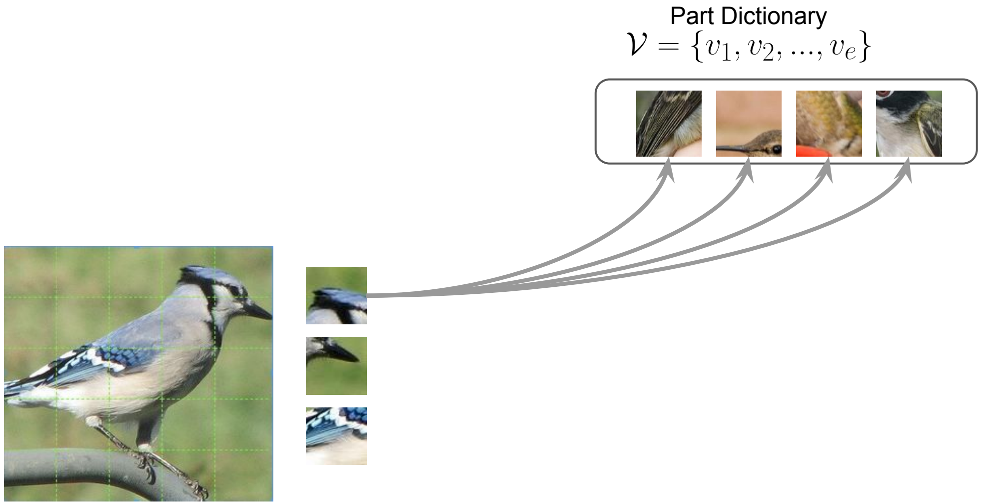
Part Dictionary
$\mathcal{V} = \{v_1, v_2, ..., v_e\}$



- Maintain a dictionary of local features
- Update it using local features of each new batches of image
- Update in a FIFO manner.

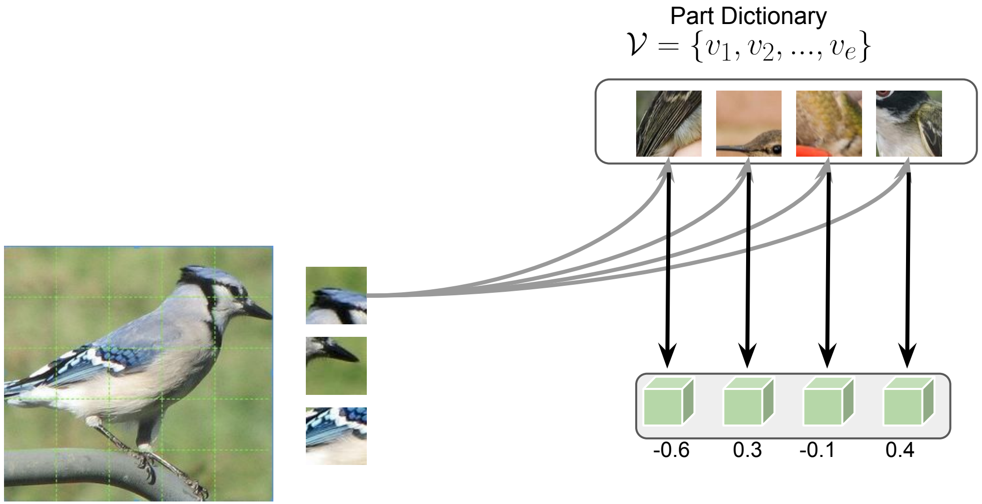


Dual ranking statistics: local branch



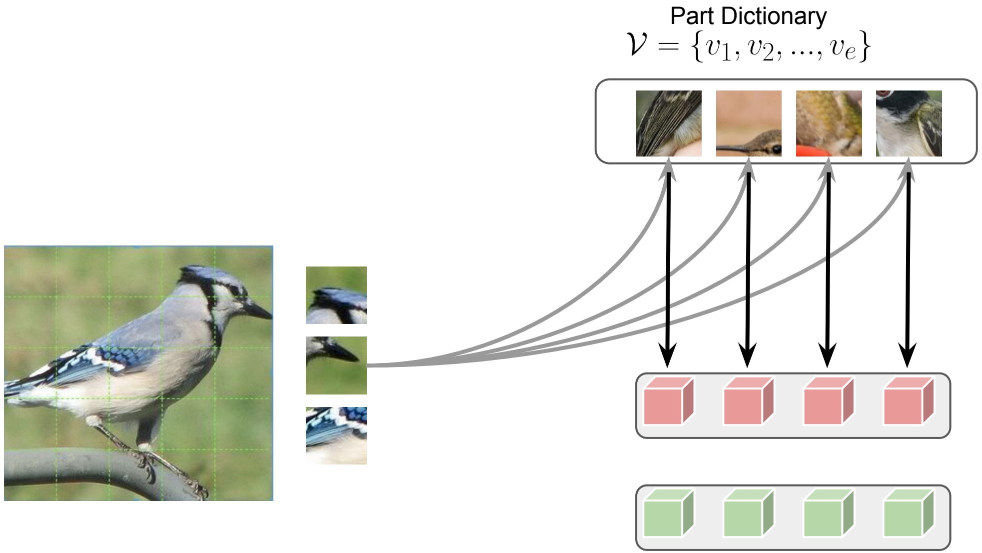
- Compare each part with the dictionary

Dual ranking statistics: local branch



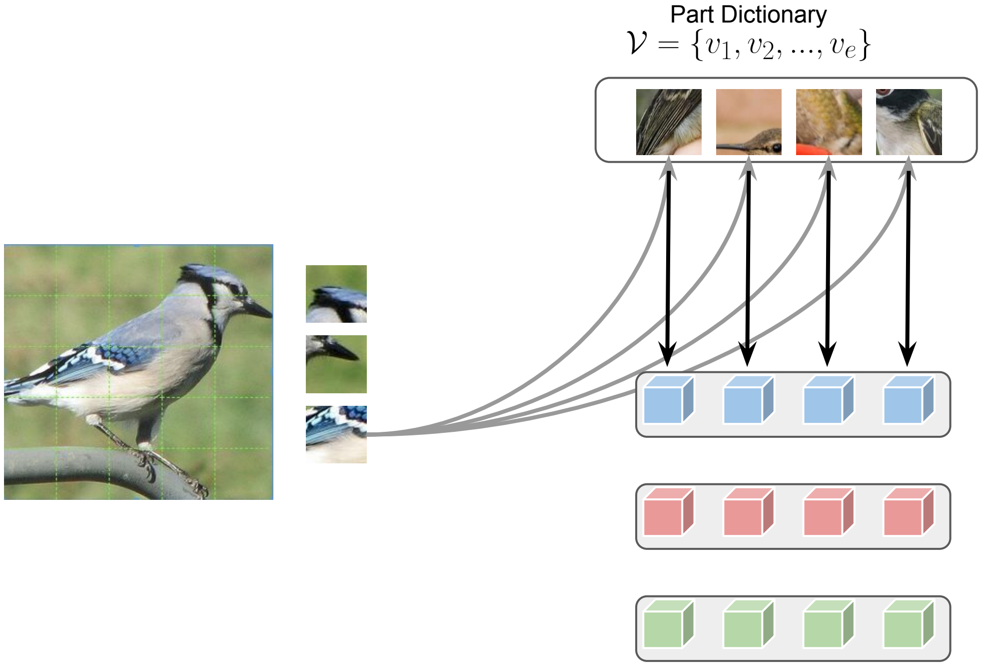
- Compare each part with the dictionary
- Get one similarity vector for each parts

Dual ranking statistics: local branch



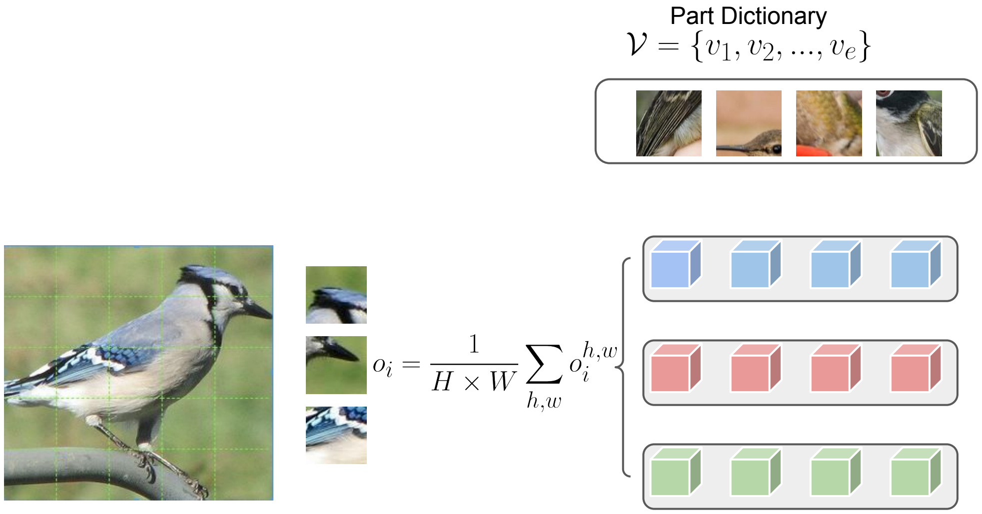
- Compare each part with the dictionary
- Get one similarity vector for each parts

Dual ranking statistics: local branch



- Compare each part with the dictionary
- Get one similarity vector for each parts

Dual ranking statistics: local branch

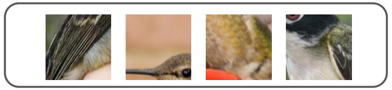


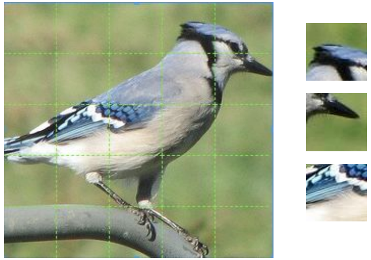
- Compare each part with the dictionary
- Get one similarity vector for each parts
- Fused all vectors together use avgpool

Dual ranking statistics: local branch

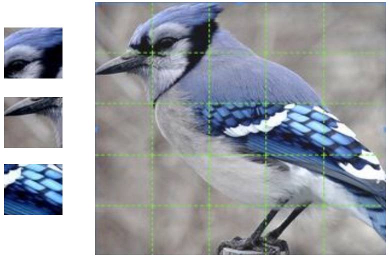
Part Dictionary

$$\mathcal{V} = \{v_1, v_2, ..., v_e\}$$



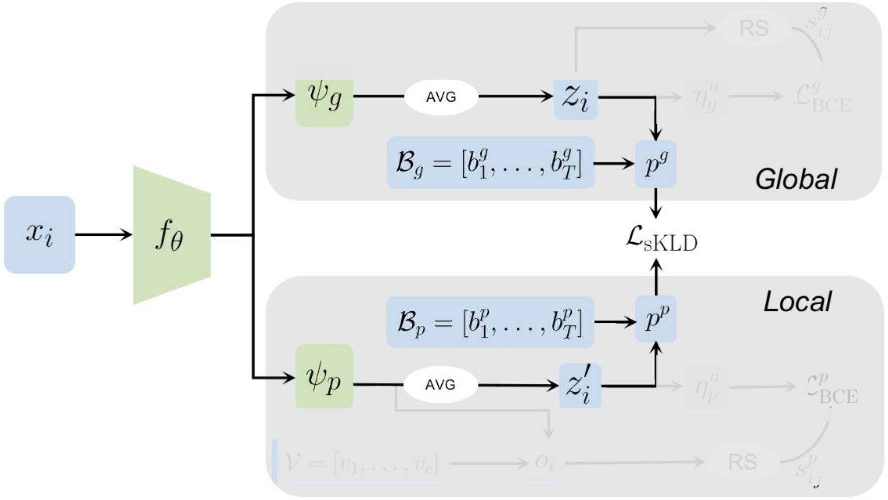


$$o_i \rightarrow \quad s_{ij} = \mathrm{RS}(o_i, o_j) \quad \leftarrow o_j$$



- Do the local information fusion for each image.
- Generate pair-wise pseudo label using RS on the fused vectors.

Method overview

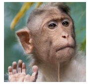


- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*. [1]
- **Mutual knowledge distillation**: allow *information exchange* between local and global.

[1] Unifying deep local and global features for image search, ECCV 2020

Mutual knowledge distillation

Different cluster assignment of the two classifier due to no labels

Cluster assignments	Global Classifier η_g^u	Local Classifier η_p^u
( , “brid”)	Cluster 1	Cluster 0
( , “dog”)	Cluster 2	Cluster 1
( , “monkey”)	Cluster 3	Cluster 2

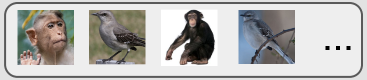
So conventional mutual learning [1] is not applicable.

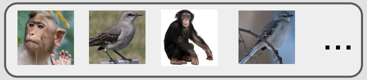
[1] Deep Mutual Learning, CVPR 2018

Mutual knowledge distillation

Global

$$\mathcal{B}_g = [b_1^g, \ldots, b_T^g]$$



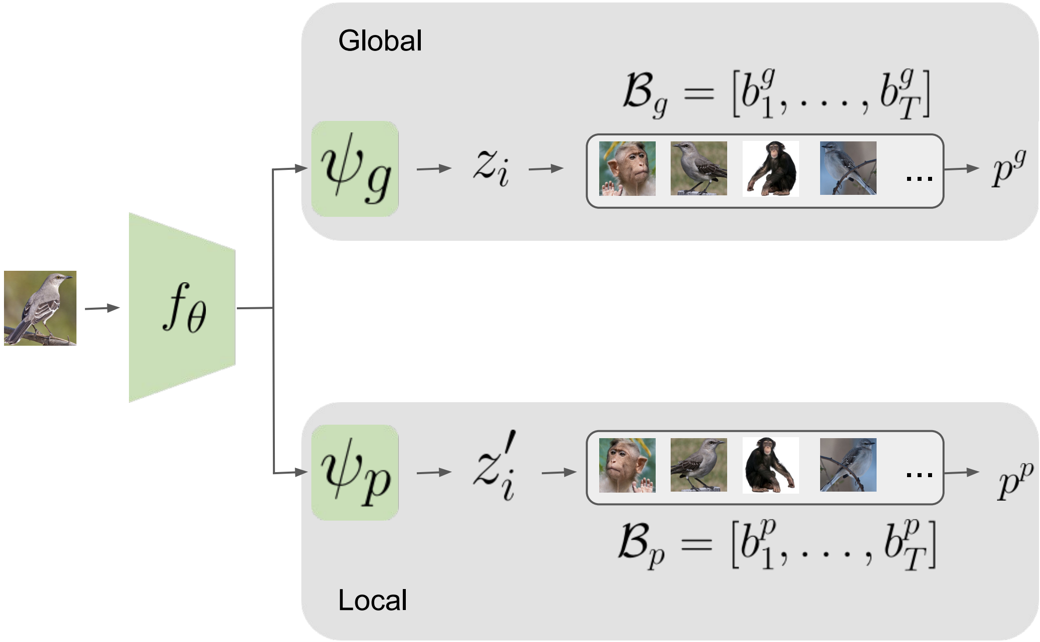


$$\mathcal{B}_p = [b_1^p, \ldots, b_T^p]$$

Local

- Maintain two FIFO feature banks for local and global branches.

Mutual knowledge distillation

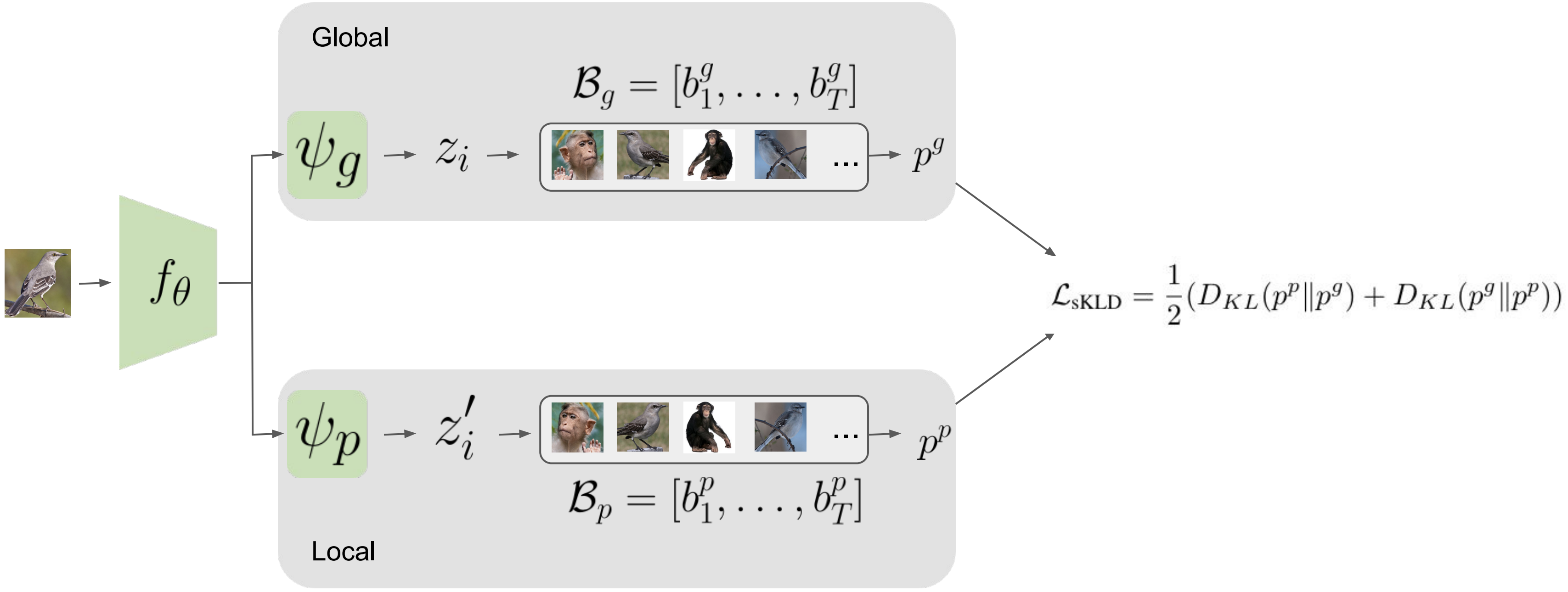


$$p^g(x_i^u, \mathcal{B}_g) = [p_1^g, \ldots, p_T^g], \quad p_j^g = \frac{exp(z_i^u \cdot b_j^g/\tau)}{\sum_{b_k^g \sim \mathcal{B}_g} exp(z_i^u \cdot b_k^g/\tau)}$$

$$p^p(x_i^u, \mathcal{B}_p) = [p_1^p, \ldots, p_T^p], \quad p_j^p = \frac{exp(z_i'^u \cdot b_j^p/\tau)}{\sum_{b_k^p \sim \mathcal{B}_p} exp(z_i'^u \cdot b_k^p/\tau)}$$

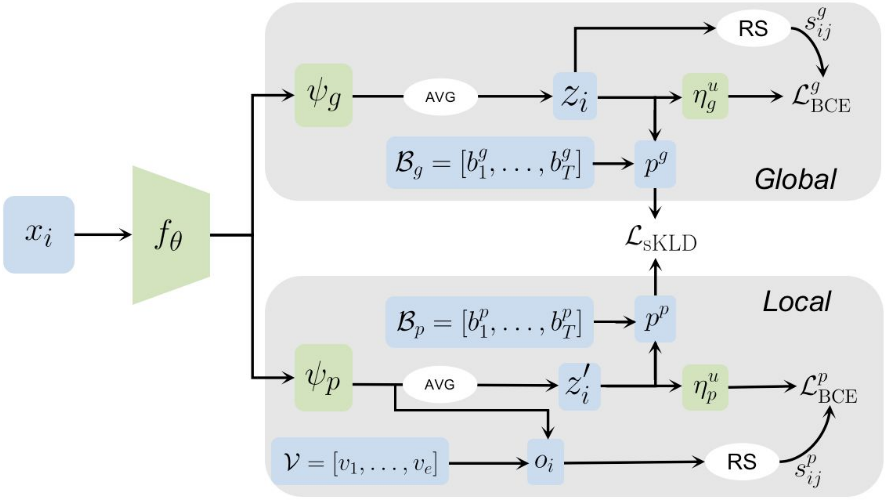
- Maintain two FIFO feature banks for local and global branches.
- Calculate the similarity score distribution over the banks.

Mutual knowledge distillation



- Maintain two FIFO feature banks for local and global branches.
- Calculate the similarity score distribution over the banks.
- Perform mutual learning on the score distributions.

Overall framework



- **Dual ranking statistics**:
  - Global comparison to have a better *recall*.
  - Local part comparison to have a better *precision*.
- **Mutual knowledge distillation**: allow *information exchange* between local and global.

Experiments

Metric: **clustering accuracy** for all the novel classes.

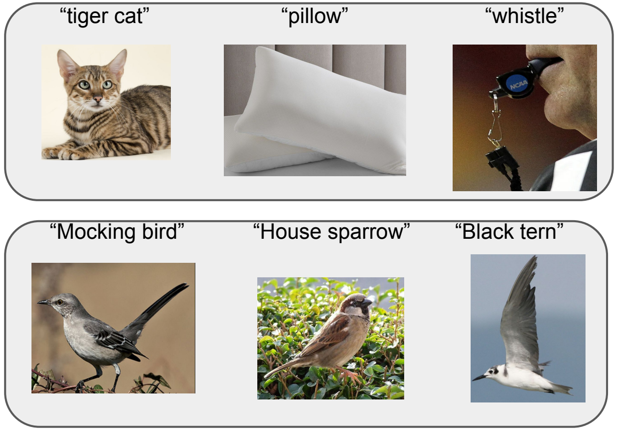
Datasets: generic image classification benchmark + fine-grained classification benchmark.

Table 1: **Data splits in the experiments.**

	labelled	unlabelled
CIFAR-10	5	5
CIFAR-100	80	20
ImageNet-1K	882	{30, 30, 30}
ImageNet-100	70	30
CUB-200	160	40
Stanford-Cars	156	40
FGVC-aircraft	81	21

Imagenet classes

CUB-200 classes



Experiments: generic datasets

Table 2: **Comparison of novel category discovery on generic classification datasets.** For fair comparison, our method uses ResNet18 [22] backbone initialized with RotNet [16] following [17].

No	Method	CIFAR-10	CIFAR-100	ImageNet-1K
(1)	k-means [33]	72.5±0.0%	56.3±1.7%	71.9%
(2)	KCL [24]	66.5±3.9%	14.3±1.3%	73.8%
(3)	MCL [25]	64.2±0.1%	21.3±3.4%	74.4%
(4)	DTC [19]	87.5±0.3%	56.7±1.2%	78.3%
(5)	RankStat [17]	90.4±0.5%	73.2±2.1%	82.5%
(6)	Ours	**91.6±0.6%**	**75.3±2.3%**	**88.9%**

Experiments: fine-grained datasets

Table 3: **Comparison of novel category discovery on fine-grained classification datasets.** "Ours w/o global" means our proposed method without global branch and mutual distillation.

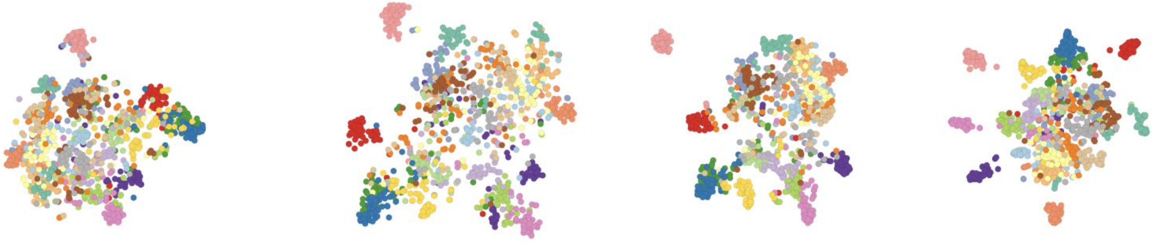
No	Method	CUB-200	Stanford-Cars	FGVC-Aircraft
(1)	DTC [19]	$33.6 \pm 0.7\%$	$46.5 \pm 2.4\%$	$58.7 \pm 1.2\%$
(2)	RankStat [17]	$39.5 \pm 1.7\%$	$53.8 \pm 2.0\%$	$66.3 \pm 0.7\%$
(3)	Ours w/o global	$43.1 \pm 2.3\%$	$56.8 \pm 2.3\%$	$67.3 \pm 1.0\%$
(4)	Ours full	$\mathbf{47.8 \pm 2.4\%}$	$\mathbf{61.9 \pm 2.5\%}$	$\mathbf{70.4 \pm 0.9\%}$

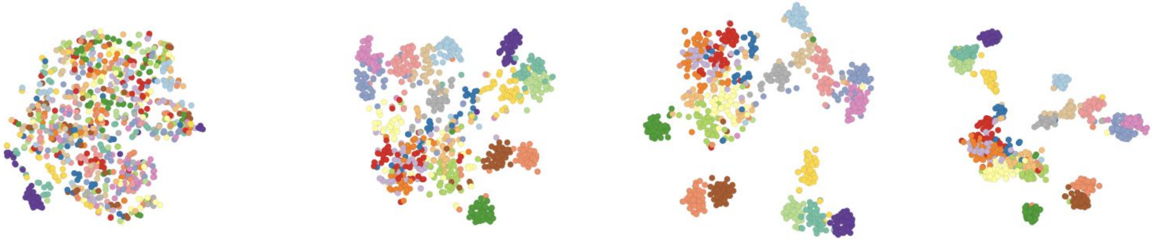
Experiments: ablation study

Table 4: **Effectiveness of different components of our method.** "MSE" means MSE consistency loss; "CE" means cross-entropy loss for training on labelled data; "BCE" means binary cross-entropy loss for training both global and local branches on unlabeled data; "sKLD" means the sKLD loss for mutual distillation between the two branches; "Self-sup." means self-supervised pre-training.

	CUB-200	Stanford-Cars	FGVC-Aircraft	ImageNet-100
Ours w/o BCE	$2.2 \pm 1.3\%$	$3.1 \pm 0.4\%$	$5.1 \pm 0.4\%$	$3.0 \pm 0.3\%$
Ours w/o sKLD	$39.8 \pm 1.8\%$	$50.6 \pm 2.1\%$	$60.8 \pm 1.5\%$	$58.2 \pm 1.2\%$
Ours w/o CE	$41.2 \pm 2.4\%$	$52.4 \pm 4.3\%$	$60.2 \pm 2.7\%$	$59.1 \pm 2.7\%$
Ours w/o MSE	$37.9 \pm 4.5\%$	$50.6 \pm 6.2\%$	$58.9 \pm 5.7\%$	$57.2 \pm 3.6\%$
Ours w/o Self-sup.	$44.3 \pm 3.5\%$	$58.2 \pm 1.8\%$	$67.4 \pm 1.3\%$	$65.3 \pm 1.3\%$
Ours full	$\mathbf{47.8 \pm 2.4\%}$	$\mathbf{61.9 \pm 2.5\%}$	$\mathbf{70.4 \pm 0.9\%}$	$\mathbf{69.4 \pm 2.1\%}$

Experiments


IM-100: (a) init (MoCo v2) (b) epoch 40 (c) epoch 120 (d) epoch 200


SCars: (a) init (MoCo v2) (b) epoch 40 (c) epoch 120 (d) epoch 200

Top row: ImageNet-100
Bottom row: Stanford Cars

Summary

- We tackles the task of novel category discovery.
- A dual ranking statistics framework is proposed.
  - The local branch focuses on local comparison
  - The global branch focuses on global information
- A mutual learning scheme is proposed to allow information exchange between the two branches.
- State-of-the-art results on both generic and fine-grained benchmarks.

Thanks for listening!

https://github.com/DTennant/dual-rank-ncd