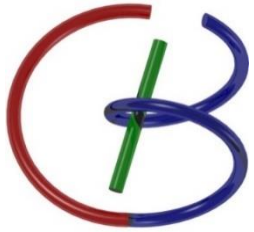
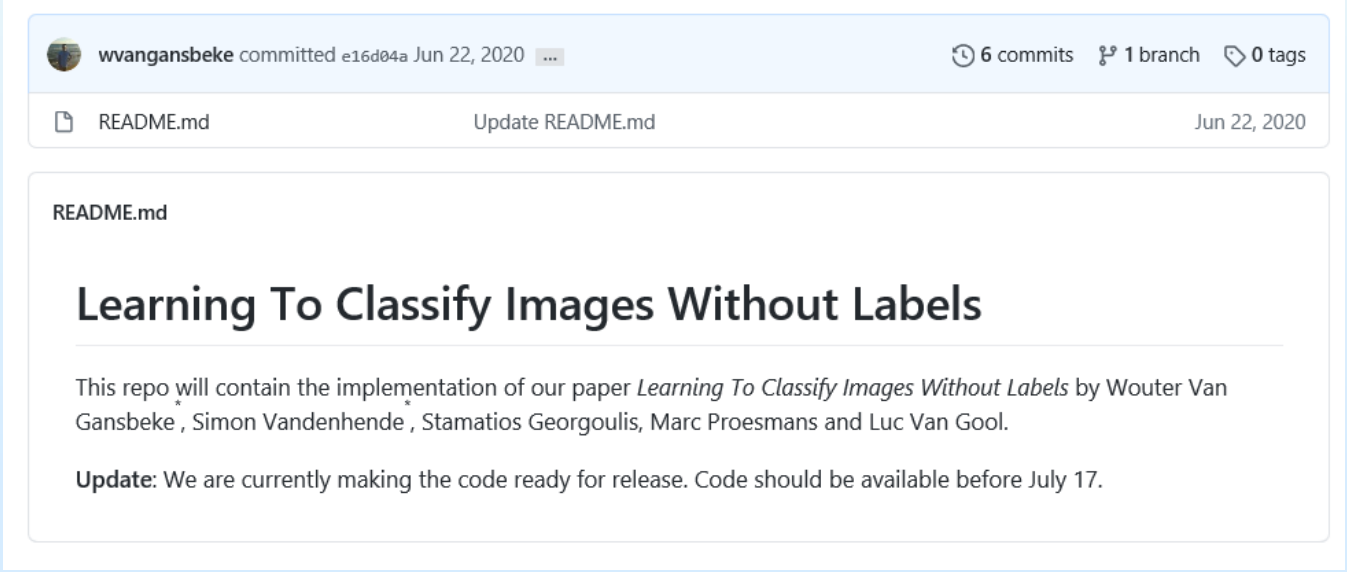
Rensselaer



Learning To Classify Images Without Labels

Wouter Van Gansbeke[1⋆] Simon Vandenhende[1⋆] Stamatios Georgoulis[2]
Marc Proesmans[1] Luc Van Gool[1,2]

[1]KU Leuven/ESAT-PSI [2]ETH Zurich/CVL, TRACE



wvangansbeke committed e16d04a Jun 22, 2020 ... 6 commits 1 branch 0 tags

README.md Update README.md Jun 22, 2020

README.md

Learning To Classify Images Without Labels

This repo will contain the implementation of our paper *Learning To Classify Images Without Labels* by Wouter Van Gansbeke*, Simon Vandenhende*, Stamatios Georgoulis, Marc Proesmans and Luc Van Gool.

Update: We are currently making the code ready for release. Code should be available before July 17.

Chuang Niu, July 1, 2020

Main ideas

- Learning to classify images without label = unsupervised classification = clustering
- In this paper, the authors deviate from recent works of end-to-end learning, and advocate a two-step approach where feature learning and clustering are decoupled.

(1) A self-supervised task from representation learning is employed to obtain semantically meaningful features.

(2) They use the obtained features as a prior in a learnable clustering approach.

A good prior representation can help avoid learning the low-level feature during clustering, which is presented in current end-to-end learning.

Method

Step 1: Representation learning for semantic clustering

A pretext task learns in a self-supervised fashion an embedding function (a neural network) that maps images into feature representations.

To learn appropriate representations for clustering, they also minimize the distance between images and their augmentations.

$$\min_{\theta} d(\Phi_\theta(X_i), \Phi_\theta(T[X_i])).$$

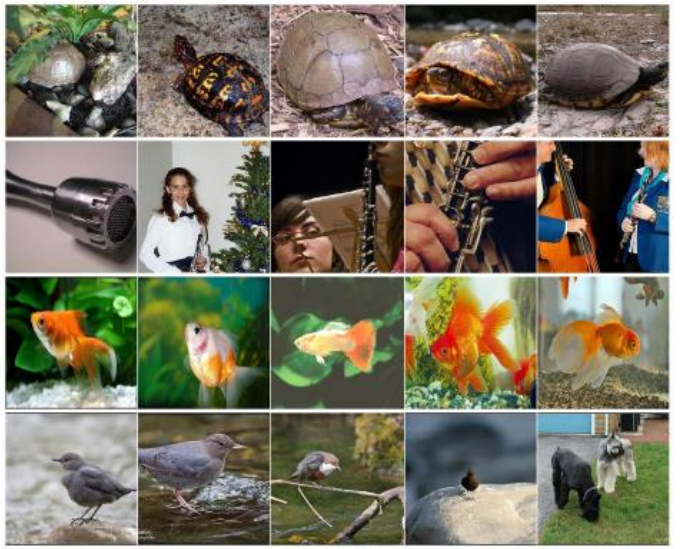


Fig. 1: Images (first column) and their nearest neighbors (other columns) sampled under pretext task [49].

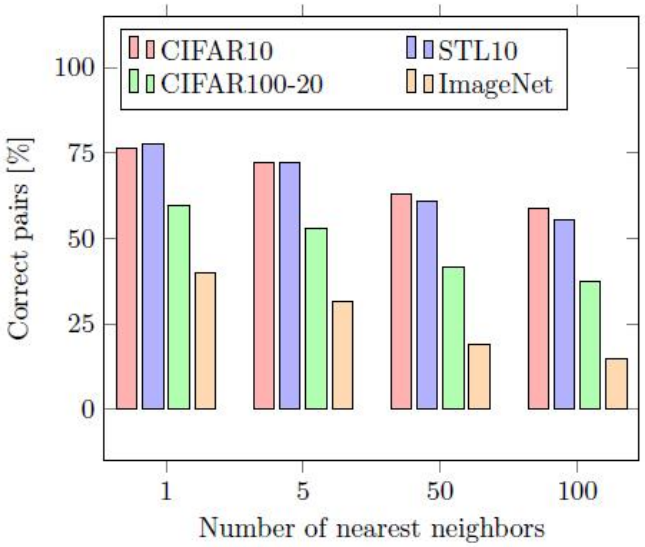


Fig. 2: Neighboring samples tend to be instances of the same semantic class.

Method

Step 2: Semantic Clustering

$$\Lambda = -\frac{1}{|\mathcal{D}|}\sum_{X\in\mathcal{D}}\sum_{k\in\mathcal{N}_X}\log\langle\Phi_\eta(X),\Phi_\eta(k)\rangle + \lambda\sum_{c\in\mathcal{C}}\Phi'^{c}_{\eta}\log\Phi'^{c}_{\eta}, \tag{2}$$

$$\text{with } \Phi'^{c}_{\eta} = \frac{1}{|\mathcal{D}|}\sum_{X\in\mathcal{D}}\Phi^{c}_{\eta}(X).$$

The first term imposes the clustering network to make consistent predictions for a sample and its neighboring.

The second term is to avoid assigning all samples to a single cluster by maximizing the entropy.

Method

Step 3: Fine-tuning through self-labeling

During training, confident samples are selected by thresholding the probability at the output, i.e. $p_{max} > threshold$.

For every confident sample, a pseudo label is obtained by assigning the sample to its predicted cluster. A cross-entropy loss is used to update the weights for the obtained pseudo labels.

To avoid overfitting, they calculate the cross-entropy loss on strongly augmented versions of the confident samples.

Method

Algorithm 1 Semantic Clustering by Adopting Nearest neighbors (SCAN)

Input: Dataset $\mathcal{D}$, Clusters $\mathcal{C}$, Task τ, Neural Nets Φ_θ and Φ_η, Neighbors $\mathcal{N}_\mathcal{D} = \{\}$.
Optimize Φ_θ with task τ. ▷ Pretext Task Step, Sec. 2.1
for $X_i \in \mathcal{D}$ **do**
    $\mathcal{N}_\mathcal{D} \leftarrow \mathcal{N}_\mathcal{D} \cup \mathcal{N}_{X_i}$, with $\mathcal{N}_{X_i} = K$ neighboring samples of $\Phi_\theta(X_i)$.
end for
while SCAN-loss decreases **do** ▷ Clustering Step, Sec. 2.2
    Update Φ_η with SCAN-loss, i.e. $\Lambda(\Phi_\eta(\mathcal{D}), \mathcal{N}_\mathcal{D}, C)$ in Eq. 2
end while
while $Len(Y)$ increases **do** ▷ Self-Labeling Step, Sec. 2.3
    Y $\leftarrow$ ($\Phi_\eta(\mathcal{D}) >$ threshold)
    Update Φ_η with cross-entropy loss, i.e. $H(\Phi_\eta(\mathcal{D}), Y)$
end while
Return: $\Phi_\eta(\mathcal{D})$ ▷ $\mathcal{D}$ is divided over C clusters

Experiments

Ablation studies

Table 1: Ablation Method on CIFAR10

Setup	ACC (Avg ± Std)
Pretext + K-means	35.9 ± 3.6
Sample + Batch Entropy Loss	19.2 ± 0.9
SCAN-Loss (Standard Augs) (Ours)	62.7 ± 3.3
SCAN-Loss (Strong Augs) (Ours)	72.5 ± 3.5
SCAN-Loss + Self-Labeling (Ours)	83.5 ± 4.1

Table 2: Ablation Pretext Task on CIFAR10

Pretext Task	ACC (Avg ± Std)
RotNet [15]	74.3 ± 3.9
Feature Decoupling [14]	83.0 ± 3.4
NCE [49]	83.5 ± 4.1

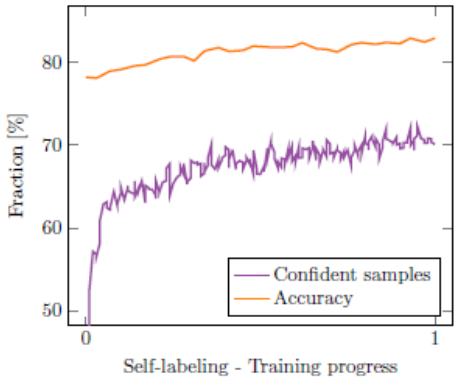


Fig. 6: Accuracy and the number of confident samples increase during self-labeling.

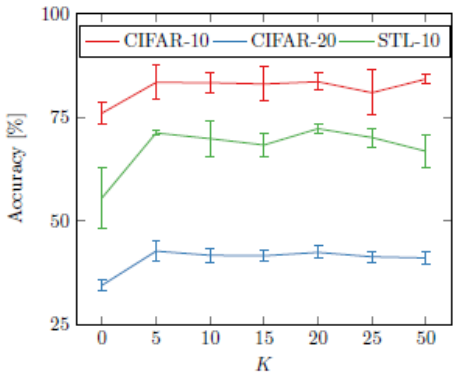


Fig. 7: Influence of the used number of neighbors K.

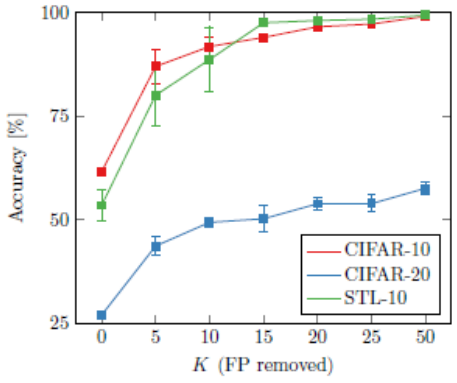


Fig. 8: Results without false positives in the nearest neighbors.

Experiments

Comparison results

Table 3: State-of-the-art comparison: We report average results after 100 epochs (*), and the results of the best model after 100 (†) and 1000 (‡) epochs, following the validation criterion from Section 3.1.

Dataset	CIFAR10			CIFAR100-20			STL10		
Metric	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
K-means [48]	22.9	8.7	4.9	13.0	8.4	2.8	19.2	12.5	6.1
SC [53]	24.7	10.3	8.5	13.6	9.0	2.2	15.9	9.8	4.8
Triplets [42]	20.5	–	–	9.94	–	–	24.4	–	–
JULE [52]	27.2	19.2	13.8	13.7	10.3	3.3	27.7	18.2	16.4
AEVB [26]	29.1	24.5	16.8	15.2	10.8	4.0	28.2	20.0	14.6
SAE [34]	29.7	24.7	15.6	15.7	10.9	4.4	32.0	25.2	16.1
DAE [47]	29.7	25.1	16.3	15.1	11.1	4.6	30.2	22.4	15.2
SWWAE [56]	28.4	23.3	16.4	14.7	10.3	3.9	27.0	19.6	13.6
AE [2]	31.4	23.4	16.9	16.5	10.0	4.7	30.3	25.0	16.1
GAN [40]	31.5	26.5	17.6	15.1	12.0	4.5	29.8	21.0	13.9
DEC [50]	30.1	25.7	16.1	18.5	13.6	5.0	35.9	27.6	18.6
ADC [17]	32.5	–	–	16.0	–	–	53.0	–	–
DeepCluster [4]	37.4	–	–	18.9	–	–	33.4	–	–
DAC [6]	52.2	40.0	30.1	23.8	18.5	8.8	47.0	36.6	25.6
IIC [24]	**61.7**	**51.1**	**41.1**	**25.7**	**22.5**	**11.7**	**59.6**	**49.6**	**39.7**
SCAN* (Ours)(Avg)	83.5 ± 4.1	73.1 ± 2.5	69.7 ± 3.9	42.8 ± 2.3	42.5 ± 1.1	26.3 ± 1.3	70.2 ± 1.3	63.2 ± 0.9	54.8 ± 1.6
SCAN† (Ours)	85.2	74.3	71.5	44.9	43.5	27.5	71.3	64.9	57.4
SCAN‡ (Ours)	**88.6**	**80.3**	**77.9**	**47.2**	**45.5**	**30.3**	**71.3**	**64.9**	**57.4**
Absolute [%]	+26.9	+29.2	+36.8	+21.5	+23.0	+18.6	+11.7	+15.3	+17.7
Relative [%]	+43.6	+57.1	+89.5	+83.7	+102.2	+159.0	+19.6.8	+30.8	+44.6
SCAN† (Overcluster)	**83.8**	**73.9**	**68.9**	**52.5**	**45.8**	**31.6**	**75.0**	**63.8**	**57.2**

Experiments

Results on ImageNet

Table 4: Validation set results for 50, 100 and 200 randomly selected classes from ImageNet. (*) Results obtained by running the publicly available code from [24]. (†) Results of applying K-means to the pretext task features from [49]

ImageNet	50 Classes				100 Classes				200 Classes			
Metric	Top-1	Top-5	NMI	ARI	Top-1	Top-5	NMI	ARI	Top-1	Top-5	NMI	ARI
Supervised	86.5	97.6	84.9	74.9	83.5	96.4	84.8	70.5	76.7	92.9	83.0	63.3
K-means†	27.1	-	23.1	11.9	22.8	-	23.1	9.6	16.7	-	22.9	6.6
IIC*	14.4	34.3	29.0	4.1	10.7	27.3	31.1	3.0	7.0	14.1	32.7	1.1
SCAN (Ours)	81.9	95.2	80.9	68.0	78.6	93.0	81.5	63.9	69.3	85.5	78.7	52.6

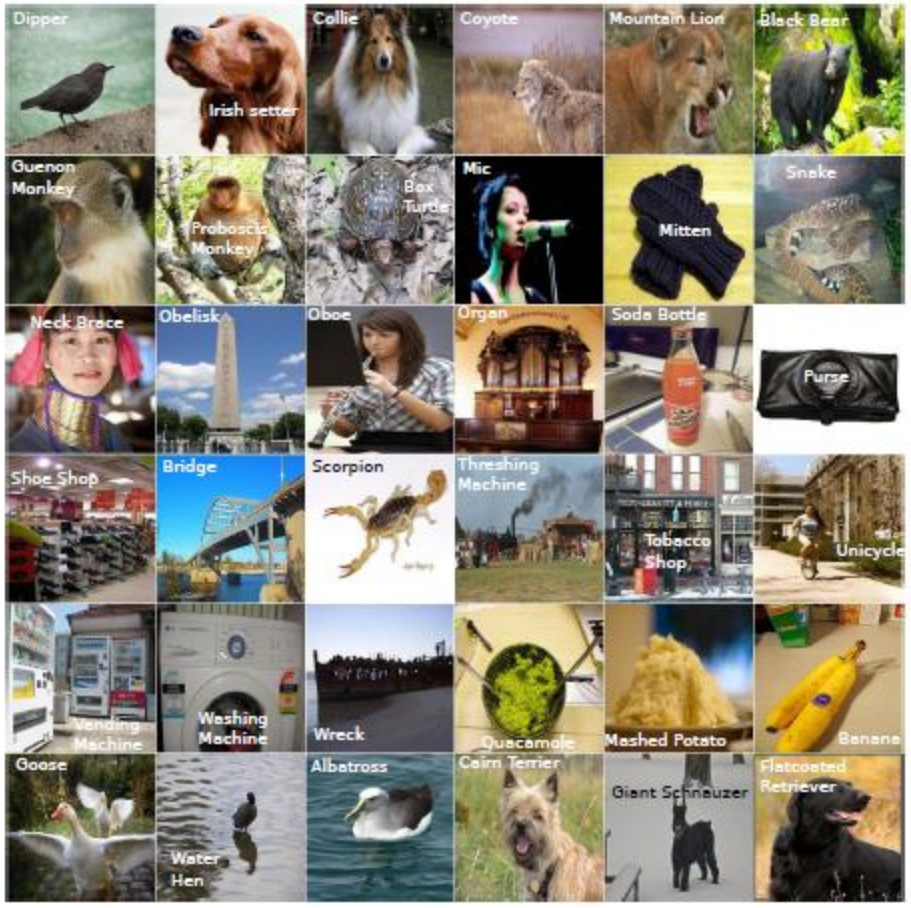


Fig. 9: Prototypes obtained by sampling the confident samples.

Experiments

Results on ImageNet: 1000 classes

Based on the ImageNet ground truth annotations, not all sample pairs should have been assigned to the same cluster. For example, the annotations discriminate between different primates, e.g. chimpanzee, baboon, langur, etc. They argue that there is not a single correct way of categorizing the images according to the semantics in case of ImageNet. Even for a human annotator, it is not straightforward to cluster each image according to the ImageNet classes without prior knowledge. Taking this into consideration, a quantitative analysis can be misguiding. However, they still evaluate the proposed model using the ground-truth annotations for completeness **(Top-1: 22%, Top-5: 39%, NMI: 60%, ARI: 11%)**.

Experiments

Results on ImageNet: 1000 classes

Based on the ImageNet hierarchy they select class instances of the following super-classes: dogs, insects, primates, snake, clothing, buildings and birds. Figure 10 shows a confusion matrix of the selected classes. The confusion matrix has a block diagonal structure. The result show that the misclassified examples tend to be assigned to other clusters from within the same superclass, e.g. the model confuses two different dog breeds. They conclude that the model has learned to group images with similar semantics together, while its prediction errors can be attributed to the lack of annotations which could disentangle the fine-grained differences between some classes.

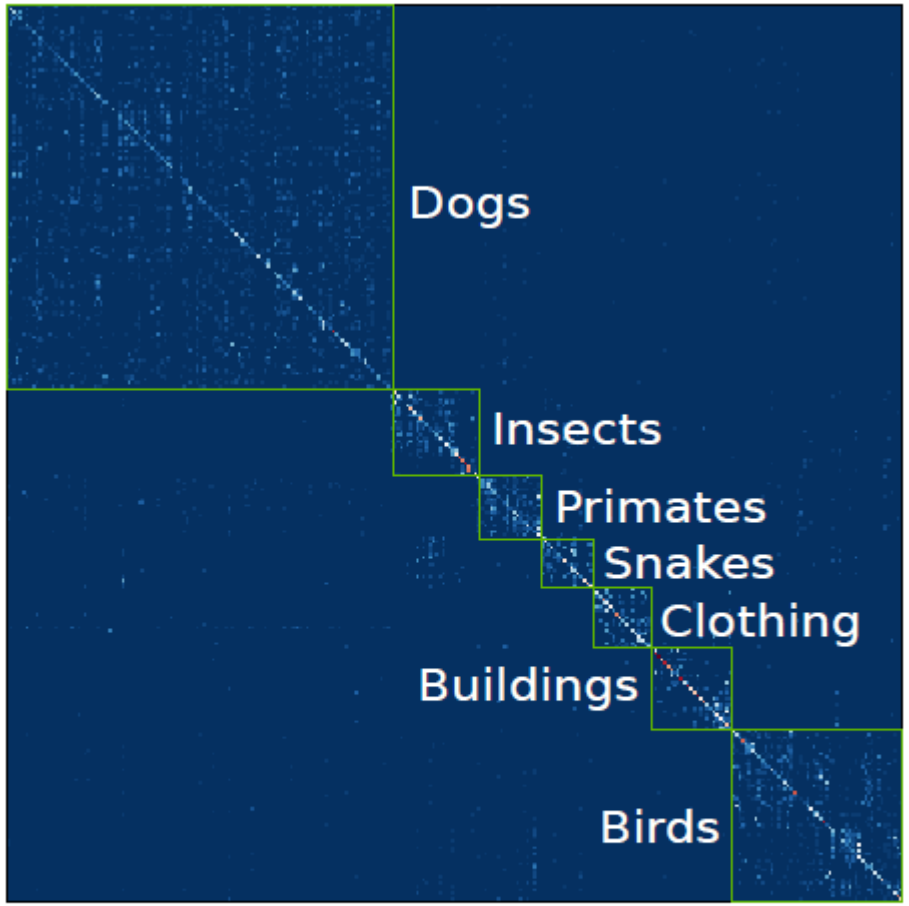


Fig. 10: Zoom on seven superclasses in the confusion matrix on ImageNet.

Conclusion

4 Conclusion

We presented a new framework to unsupervised image classification. The proposed approach comes with several advantages relative to recent works which adopted an end-to-end strategy. Experimental evaluation shows that the proposed method outperforms prior work by large margins, for a variety of datasets. Furthermore, positive results on ImageNet demonstrate that semantic clustering can be applied to large-scale datasets. Encouraged by these findings, we believe that our approach admits several extensions to other domains, e.g. semantic segmentation, semi-supervised learning and few-shot learning.

Thanks for your attention!