Reproducibility of SCAN: Semantic Clustering by Adopting Nearest neighbors

Tasneem Naheyan

EECS 6322

SCAN: Learning to Classify Images without Labels

Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis,
Marc Proesmans, Luc Van Gool

SCAN

An approach for image classification in the absence of ground truth labels:

1. Pretext
2. SCAN Clustering
3. Self-labeling

- Previous end-to-end methods e.g. DeepCluster, IIC, rely on initial feature representations and are dependent on network initialization
- SCAN avoids this by using representation learning

Pretext Step

- Learn feature representations which are invariant to image transformations
- Minimize the distance between an image and its augmented form:

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

- Prevent clusters from latching on lower-level features which do not capture the content of an image
- Mine K nearest neighbors based on feature similarity, avoiding cluster degeneracy
- Use the neighbors as a prior for the semantic clustering step

SCAN Step: Semantic Clustering

- Train a new neural network initialized with weights from pretext task
- Classify each image and its nearest neighbors together by minimizing the novel objective function:

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

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

Probability of X being assigned to cluster c

- Entropy term spreads the predictions across the clusters preventing all samples from being assigned to a single cluster

Self-labeling Step

- Network initialized with weights from SCAN step
- Correct incorrect classifications due to noisy neighbors
- Threshold applied to filter most-confident predictions
- Network weights are updated using a weighted cross-entropy loss, calculated on strongly augmented versions of confident samples to prevent overfitting

Published Results

- Datasets: CIFAR10, CIFAR100-20, STL10 and ImageNet
- Best models outperform the state-of-the-art by:
  - 26.6% on CIFAR10
  - 25.0% on CIFAR100-20
  - 21.3% on STL10
- Evaluation metrics: clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI)
- Ablation studies on CIFAR10 found e.g. SCAN performs best when the pretext task imposes invariance between an image and its augmentations
- First unsupervised learning method to perform well on ImageNet

Reproducibility Goals

- Central Claim
  - SCAN outperforms state-of-the-art (IIC: classification accuracy of 61.7%)
  - Reproduce results on CIFAR10 within reported range of average results:
    - classification accuracy for SCAN without self-labeling: 81.8 ± 0.3
    - SCAN with self-labeling: 87.6 ± 0.4
- Secondary Claims
  - Applying strong augmentation to images improves SCAN performance
  - Successful self-labeling requires a shift in augmentation

Methodology

- Python reimplementation using PyTorch torchvision package, CIFAR10 dataset
- SimCLR instance discrimination used for the pretext step, SCAN step with the novel loss function & the self-labeling code using weighted cross-entropy loss reimplemented following official documentation and authors' public code as reference
- Sections of code not directly part of the author's novel contribution are reused e.g. code for strong augmentations (RandAugment & Cutout), hungarian matching evaluation code, faiss library for KNN mining
- Experiments run with SCAN reimplementation and original code with same hyperparameter settings as paper for comparison purposes

Methodology

- Single GPU with 12 GB RAM on Google Colab (NVIDIA Tesla K80) or Colab Pro (T4 or P100), 8-24 hour continuous runtimes
- Checkpointing
- Runtime
  - Pretext SimCLR training: 9-13 hours for 500 epochs
  - SCAN & self-labeling steps: 6-8 hours

Results

- Complete SCAN reimplementation achieved classification accuracy within the range reported in the paper
- Results support central claim
- However, evaluation metrics for the reimplementation without self-labeling do not fall within the corresponding published range

Evaluation Metric	ACC	NMI	ARI
Published SCAN	81.8± 0.3	71.2± 0.4	66.5± 0.4
Published SCAN + Self-label	87.6± 0.4	78.7± 0.5	75.8± 0.7
Reimplemented SCAN	79.7	67.4	62.3
Reimplemented SCAN + Self-label	87.2	79.7	75.8
Original SCAN	82.0	71.6	66.9
Original SCAN + Self-label	87.3	78.5	75.4

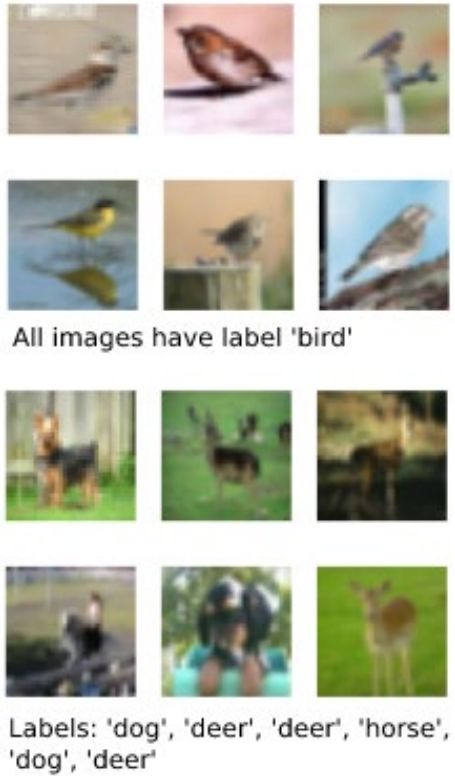
Table 1: Percentage clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI) scores for the SCAN approach with and without self-labeling. 'Original' refers to authors' code run directly. The 'reimplemented' scores reported are the average of two runs.

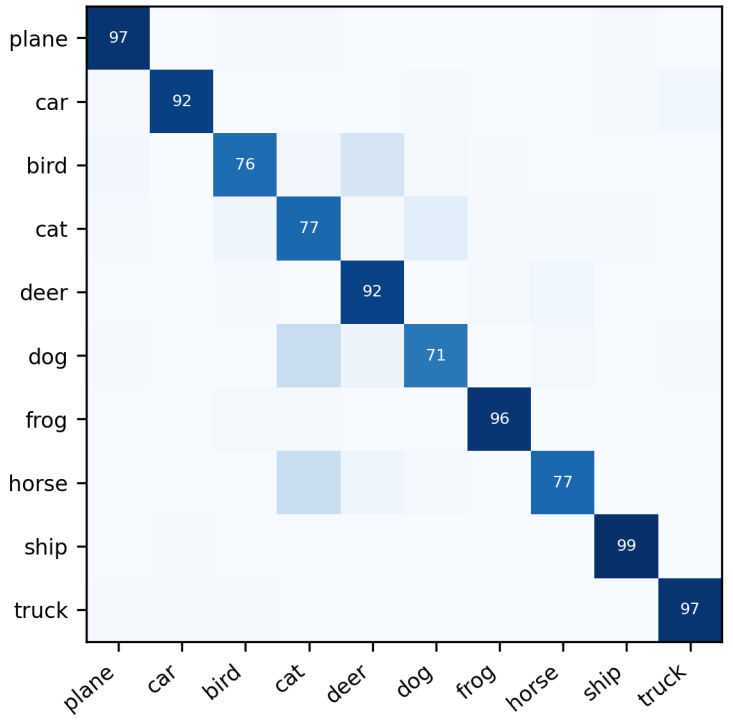
Model	Train Set	Test Set
Reimplemented SimCLR	80.2	75.0
Original SimCLR	79.8	77.9

Table 2: Percentage accuracy in mining k nearest neighbors. 'Train set' column shows the accuracy of the top 20 nearest neighbors in the train set and 'Test set' shows the accuracy of the top 5 nearest neighbors on the test set.

Results

Two sets of images with their nearest neighbors mined using the reimplemented SimCLR model





Confusion matrix created with the predictions after self-labeling

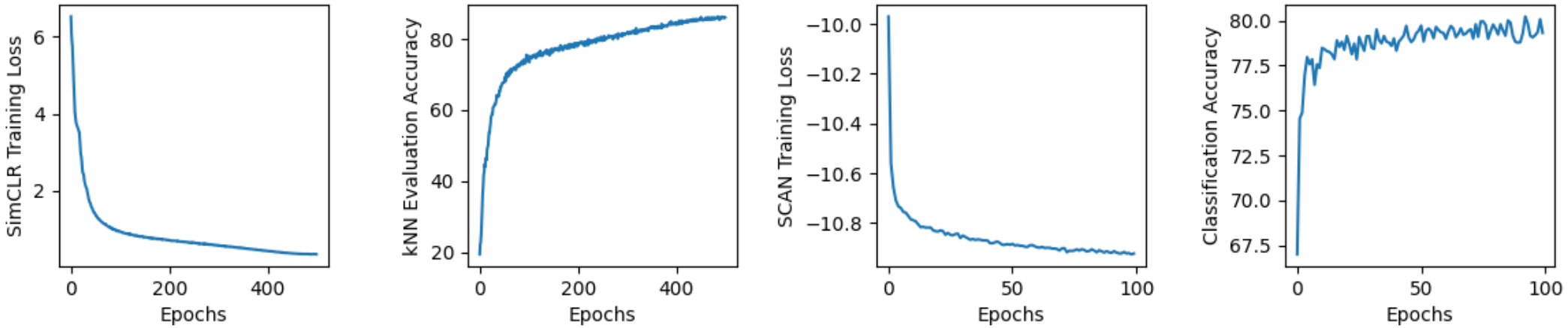
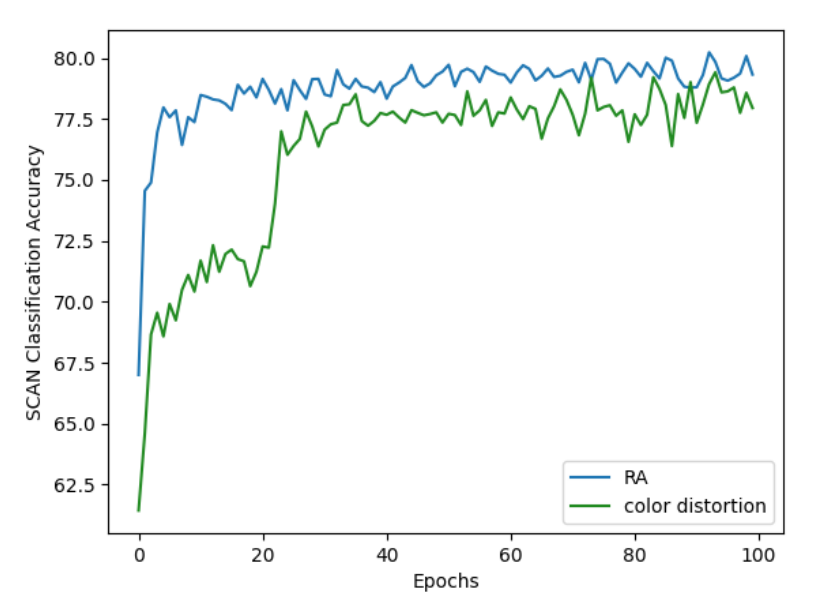
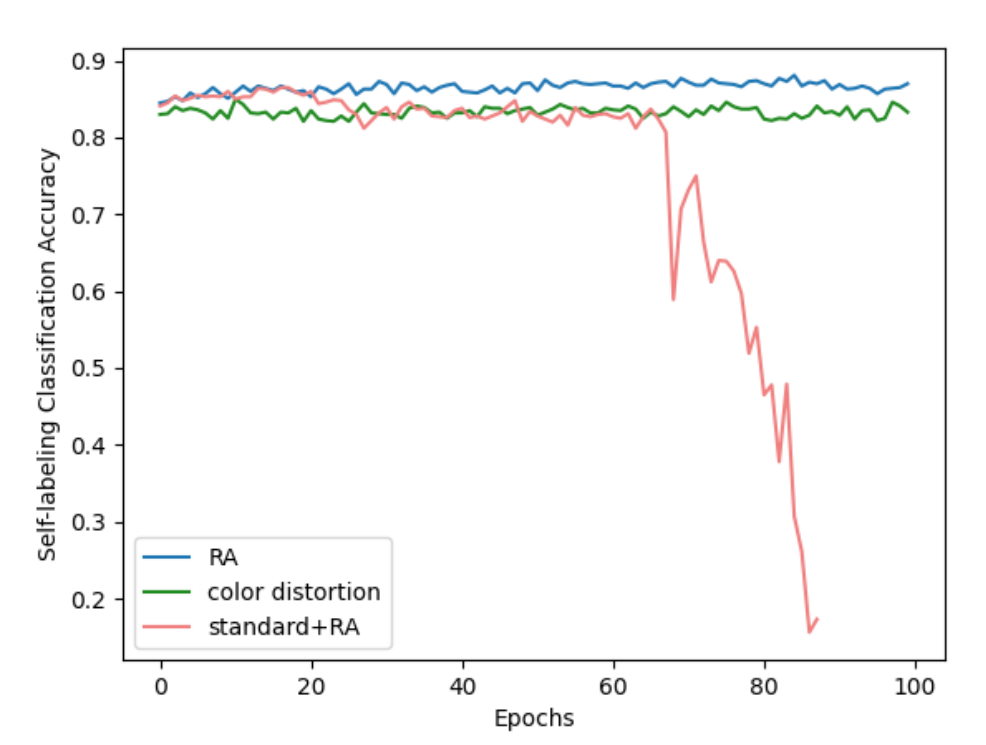


Figure 1: The two plots on the left show training loss and accuracy in identifying nearest neighbors during training in the pretext step. The two plots on the right show training loss and classification accuracy in the SCAN step.

Stretch Goal

- Claims:
  - Strong image augmentations improve performance by imposing additional invariances
    - Apply strong augmentations by selecting four transformations from RandAugment, then apply Cutout
  - A shift in augmentations is required for successful self-labeling, needed to prevent network from overfitting on already well-classified examples
    - Transformation parameters are sampled between fixed intervals
- Experiments:
  - Change augmentation strategy to random color distortion
  - Apply strong augmentation to one image in pair for self-labeling





Stretch Goal

- Higher accuracy reached in SCAN and self-labeling steps when using strong augmentations
- Performance drops in self-labeling when strong augmentation is applied to only one image
  - this differs from the authors' observations where the self-labeling performance drops when SimCLR transformations are applied to both images, and the accuracy is high when RandAug is applied to one image and either RandAug or SimCLR transforms are applied to the other image
- Note: Self-labeling model here initialized with weights from SCAN model trained with original augmentation strategy

Discussion

- Results support main claim of paper in case of CIFAR10
  - complete SCAN reimplementation achieved classification accuracy of 87.2% which is within published range i.e. 87.6±0.4 and above the state-of-the-art (IIC [3]) accuracy of 61.7%
- Accuracy measures for the pretext and SCAN (without self-labeling) steps slightly lower than published results.
  - Self-labeling step is effective in correcting mistakes caused by noisy nearest neighbors
- Results support claim that strong augmentations improve performance
- Results do not support successful self-labeling requires a shift in augmentation

What was easy

- Reimplementing SCAN novel loss function
- Running the authors' public code

What was difficult

- Somewhat complex code with parts specific to ImageNet
- Hyperparameter settings unclear
- Misleading runtime

Replication successful on CIFAR10 but to fully verify claims the method should be reimplemented on at least CIFAR100-20 & STL10.

References

1. Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Scan: Learning to classify images without labels, 2020.
2. Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *arXiv preprint arXiv:1702.08734,2017*.
3. Xu Ji, João F. Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation, 2019.
4. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.
5. Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical automated data augmentation with a reduced search space, 2019.
6. Terrance DeVries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout, 2017.
7. Alex Krizhevsky. Learning multiple layers of features from tiny images, 2009.

THANK YOU