*vitaliy.kinakh@unige.ch

UNIVERSITÉ DE GENÈVE

SIP stochastic information processing

ScatSimCLR: self-supervised contrastive learning with pretext task regularization for small-scale datasets

Vitaliy Kinakh*, Olga Taran, Sviatoslav Voloshynovskiy

Department of Computer Science, University of Geneva, Switzerland

Source code

Introduction 1

- Self-supervised learning (SSL) - learning of the data representations that are not based on labeled data, these methods demonstrate a classification performance close to their supervised counterparts.

- SSL methods are based on powerful neural networks with the number of parameters ranging from 5M to 500M.

- 'Small dataset' problem – SSL faces some problems with limited data, which lead to the overfitting of the big models.

Contributions

- Model with the reduced number of parameter at encoder while preserving the same classification performance. This is achieved by ScatNet – geometrically invariant network.

- Pretext task regularization based on the estimation of parameters of applied augmentation transform such as rotation and jigsaw permutation.

- Investigation of the role of augmentations.

- Achieved SOTA performance on STL10 and CIFAR-100-20 datasets.

Related work 2

- Contrastive learning is considered among SOTA technique for self-supervised learning. It is based on minimization of a distance between similar (positive) pairs and maximization of dissimilar (negative) ones.

- Hand-crafted geometrically invariant transform ScatNet is a class of CNNs designed with fixed weights with properties: (1) deformation stability; (2) sparce representations; (3) interpretable representations.

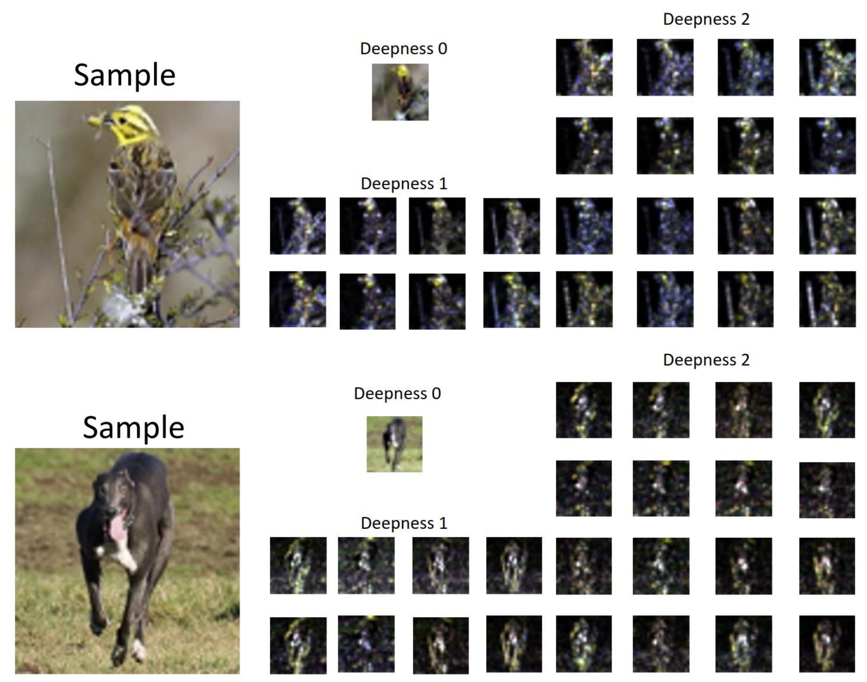

Fig. 1. Example of ScatNet feature vectors.

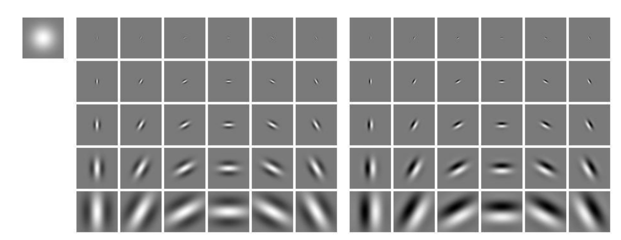

Fig. 2. ScatNet filter bank for J=5 (number of scales) and L = 6 (number of rotations).

ScatSimCLR 3

ScatSimCLR is based on SimCLR SSL method, where the base ResNet encoder is replaced by the hand-crafted ScatNet $f_{\phi_{\mathrm{Scat}}}$ and a small capacity adapter network $f_{\phi_{\mathrm{h}}}$ with a pretext task regularization. **t –** parameters of the transformation under the pretext task estimation: either rotation or jigsaw permutation, φ_{t} - transformation, $\tilde{\mathbf{x}} = \varphi_{\mathrm{t}}(\tilde{\mathbf{x}})$ - transformed view, $\mathbf{h} = f_{\phi_{\mathrm{Scat}}}(f_{\phi_{\mathrm{h}}}(\tilde{\mathbf{x}}))$ - embeddings.

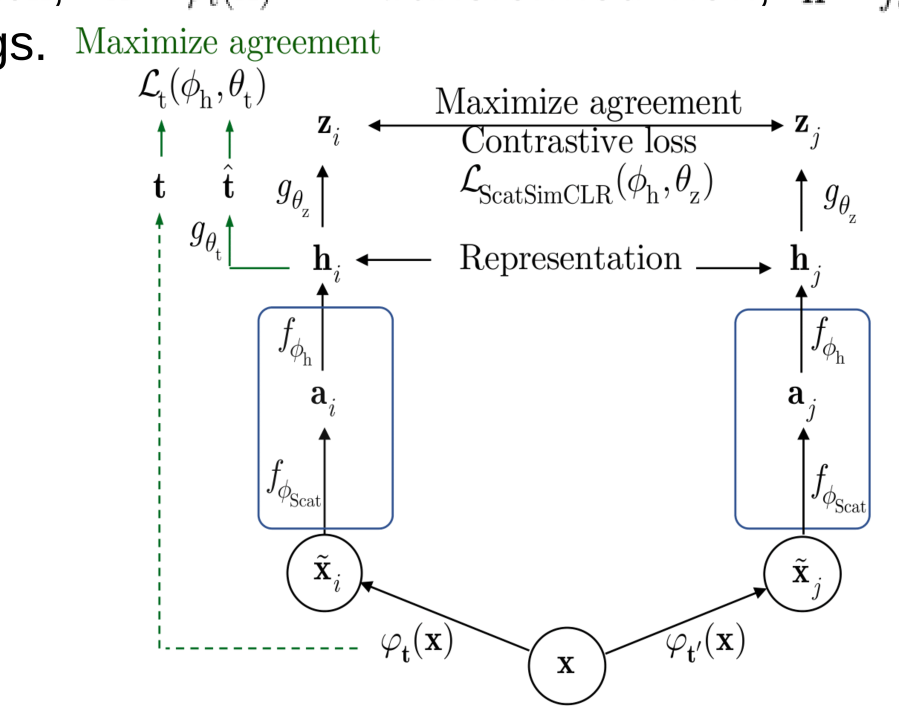

Fig. 3. Proposed ScarSimCLR system with contrastive loss and additional regullarization based on the estimation of augmentation transform.

The impact of augmentations 4

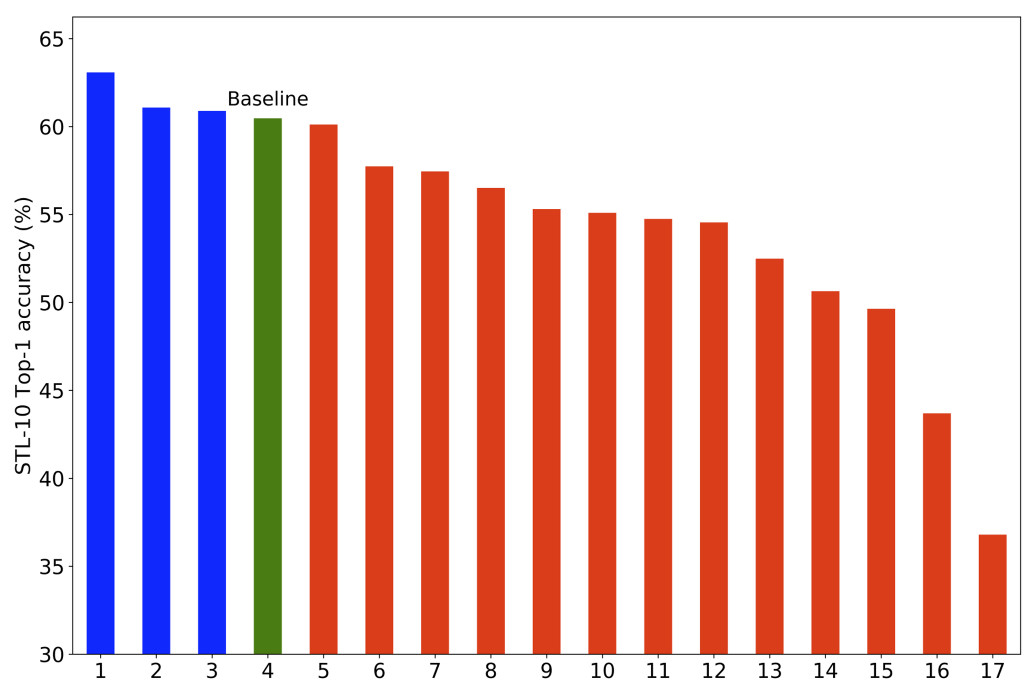

Fig. 4. Impact of removing the augmentations on the perfor-mance of ScatSimCLR for STL-10: ”Baseline” denotes ScatSim-CLR trained with all augmentations (cropping, flipping, color,grayscale, Gaussian blur and affine augmentations). The follow-ing labels denote: 1 - the baseline without the affine augmenta-tions; 2 - only cropping and color augmentations; 3 - the baselinewithout the horizontal flipping; 5 - the baseline without Gaussianblur augmentations; 6 - the baseline without cropping and Gaus-sian blur augmentations; 7 - the baseline without color and Gaus-sian blur augmentations; 8 - the baseline without grayscale andGaussian blur augmentations; 9 - the baseline without croppingand grayscale augmentations; 10 - the baseline without color aug-mentations; 11 - the baseline without cropping augmentations; 12- the baseline without grayscale augmentations; 13 - only crop-ping augmentations; 14 - the baseline without color and grayscaleaugmentations; 15 - only color augmentations; 16 - the baselinewithout crop and color augmentations; 17 - no augmentations.

Pretext task regularization 5

- The introduction of the pretext task improves the classification accuracy for both considered models ScatSimCLR and SimCLR. For all models rotation augmentation pretext task provides higher increase in classification performance in comparison to jigsaw.

Table 1. Impact of the pretext task regularization on the classification accuracy on STL-10 dataset.

Baseline model	Accuracy on STL-10: Without pretext	With pretext: Rotation	With pretext: Jigsaw	Num. of paramers
ScatSimCLR 8	74.78%	**77.86%**	76.36%	6.1 M
ScatSimCLR 12	76.57%	**78.43%**	77.78%	7.8 M
ScatSimCLR 16	77.03%	**78.5%**	77.91%	10.5 M
ScatSimCLR 30	77.86%	**79.11%**	78.4%	14.1 M
SimCLR (ResNet18)	71.90%	**76.36%**	75.22%	11.5 M

Results 6

ScatSimCLR achieves SOTA in unsupervised image classification on STL-10 – **85.11**% and on CIFAR-100-20 – **63.86**%.

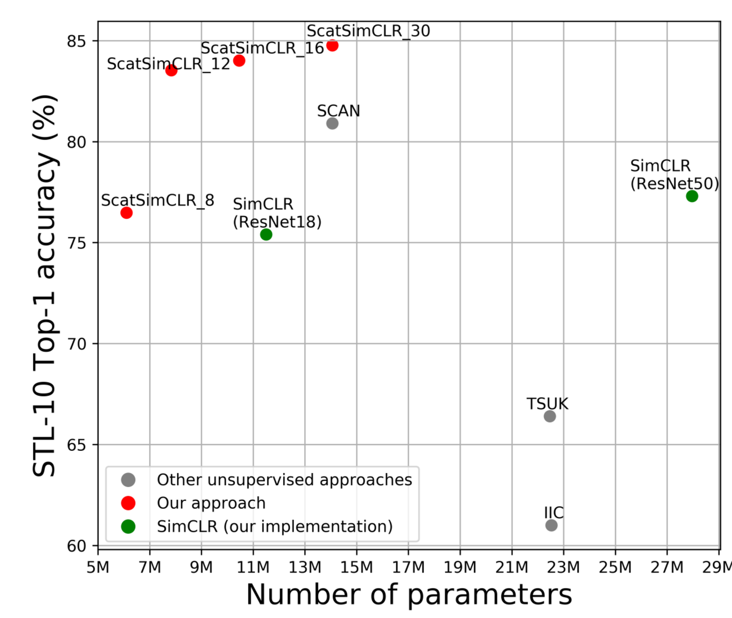

Fig. 5. STL-10 Top-1 accuracy of self-supervised methods. Gray dots indicete other self-supervised methods. Our method, ScatSimCLR, is shown in red. The results are obtained with models trained for 1000 epochs.

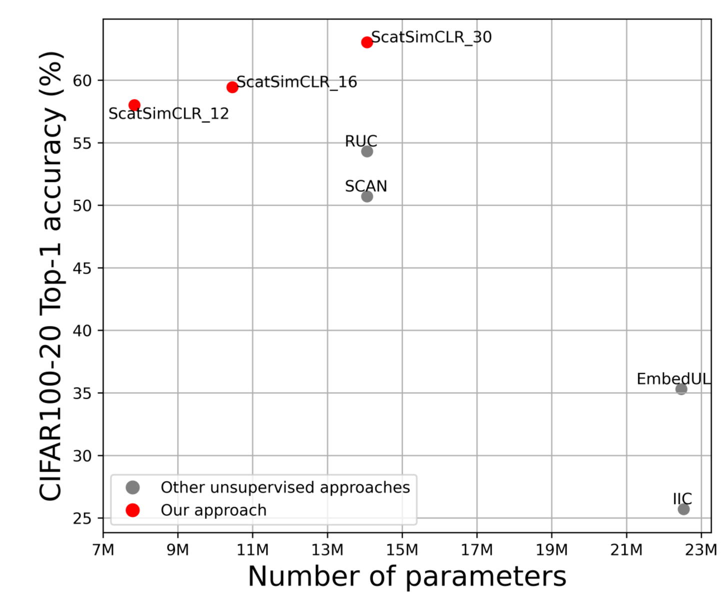

Fig. 6. CIFAR-100-20 Top-1 accuracy of self-supervised methods, Gray dots indicate other self-supervised methods. ScatSimCLR is shown in red.