EVEN A SINGLE SIMPLE AUGMENTATION WITH SELF-SUPERVISED LEARNING CAN BE HELPFUL FOR THE DOWNSTREAM TASKS

Evgenii Pishchik
Independent researcher
Moscow, Russia
jenjapishhik@gmail.com

ABSTRACT

This paper explores some unexpected capabilities of Self-Supervised Learning (SSL) and shows that even a single cutout augmentation for SSL can achieve better results in downstream tasks compared to traditional supervised approaches. These unusual properties of SSL can be used for further research in this area.

1 INTRODUCTION

Self-Supervised Learning or SSL has been a huge area of research in the last few years. It has been shown to have many advantages over traditional supervised approaches. The biggest advantage is that it doesn't need a lot of labeled data, it can be used on any data from the Internet. Augmentations play a very important role in the training of some SSL approaches, which depend on generating different views of the same sample and trying to get a similar vector representation from both views. The augmentations and their roles are the main focus of this work.

2 METHOD

Two SSL methods were chosen - SimCLR (Chen et al., 2020) and SimSiam (Chen & He, 2020). SimCLR requires a lot of resources to achieve a good quality. The original paper says that we should use larger batch sizes and more epochs than for the supervised approaches. SimSiam overcomes this problem and allows us to use smaller batch sizes and fewer epochs. Both methods train models to predict similar vectors for two different views of the same image.

2.1 DATASET

The GTSRB (Houben et al., 2013) and STL-10 (Coates et al., 2011) datasets were used. The reason why these two datasets were chosen is quite simple. GTSRB is a sign dataset, and the proposed augmentations simulate the process of putting stickers on signs in real life, but without any perspective and with unnatural colors. STL-10 is a special dataset for unsupervised training. It's inspired by the CIFAR-10 (Krizhevsky, 2012) dataset, but with some modifications.

2.2 AUGMENTATIONS

The main idea is that even a simple augmentation such as a white box cutout with SSL training can be useful for training the model on the downstream task.

Four similar cutout augmentations have been proposed for SSL training. White box / random color box - cutout a single white box / random color box, but two views have the same augmentations before it. Advanced white box / advanced random color box - cutout a single white box / random color box, but two views have various popular augmentations before it, such as color jitter, random cropping, random flipping, etc.

For the downstream tasks, the models were trained on clean data without cutout augmentations, but were tested with such augmentations to simulate different attacks. The test set used a single white

Table 1: Test metrics for the three random color boxes augmentation. SSL training configuration - SimSiam, epochs = 200, batch size = 256, GTSRB dataset. SimSiam is the augmentation method used in the original paper. ImageNet pretrained is the supervised approach.

method	**top-1 accuracy**	**top-5 accuracy**
White box	0.388	0.679
Advanced white box	**0.470**	**0.731**
SimSiam	0.367	0.603
ImageNet pretrained	0.383	0.643

box cutout, a single random color box cutout, and three different sized box cutouts. The approach without test cutout augmentations was also tested.

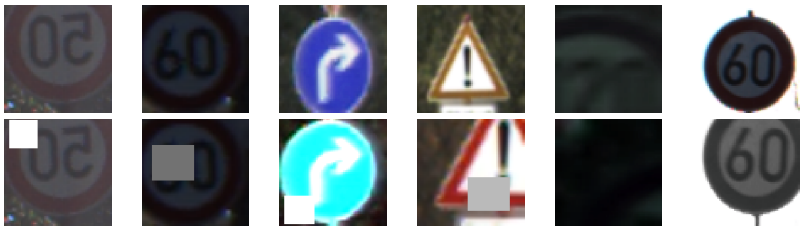

Figure 1: Augmentations of views for the SSL training on the GTSRB dataset. The augmentations are arranged from left to right - white box, random color box, advanced white box, advanced random color box, SimCLR, SimSiam.

3 Experiments

The SimSiam was trained on the GTSRB dataset for 200 epochs with a batch size of 256. The model was then trained on clean data for a downstream classification task, and metrics were measured on the data with test augmentations. See Table 1. The important thing is that such changes don't degrade the performance on the clean test data. SimSiam, advanced white box and ImageNet (Deng et al., 2009) pretrained methods have the similar 0.9 top-1 accuracy on the clean data, but the supervised approach with random weight initialization loses with each test augmentation, even on the clean data, where it achieves 0.73 top-1 accuracy.

4 Conclusions

It has been shown that even a single white box augmentation can outperform the usual pretrained supervised approaches and previous SSL methods in some cases. SSL is a new, rapidly developing area, and it has been shown that many new opportunities can be found using SSL. We hope that this article will show that augmentation for SSL methods is a huge area for future research.

URM Statement

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020.

Xinlei Chen and Kaiming He. Exploring simple siamese representation learning, 2020.

Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, pp. 215–223, 2011.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel. Detection of traffic signs in real-world images: The german traffic sign detection benchmark. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2013. doi: 10.1109/IJCNN.2013.6706807.

Alex Krizhevsky. Learning multiple layers of features from tiny images. 2012.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017.

Yang You, Igor Gitman, and Boris Ginsburg. Large batch training of convolutional networks, 2017.

A SSL CONFIGURATION

The configurations for the SimCLR and SimSiam SSL methods are taken from the original papers.

SimCLR was trained using the LARS (You et al., 2017) optimizer with a learning rate = $0.3 \times$ BatchSize$/256$, weight decay = 10^{-6}, and the 128-dimensional latent space projection. The learning rate schedule is cosine decay without restarts with warmup for 10 epochs.

SimSiam was trained using the SGD optimizer with a learning rate = $0.05 \times$ BatchSize$/256$, momentum = 0.9, weight decay = 10^{-4}, and the 2048-dimensional latent space projection. For the large batch sizes ($>= 1024$), cosine decay without restarts was used with 10 epochs of warmup (Loshchilov & Hutter, 2017). Warmup was not used for the smaller batch sizes.

B METRICS

It has been found that such augmentations can outperform supervised approaches trained from scratch on the most test augmentations. See Table 2 and Table 3. This can be useful when pretrained ImageNet weights can't be used in a very specific domain.

Table 2: Test metrics for the three random color boxes augmentation. SSL training configuration - SimCLR, epochs = 1000, batch size = 2048, GTSRB dataset.

method	**top-1 accuracy**	**top-5 accuracy**
Advanced white box	0.433	0.701
Advanced random color box	**0.498**	**0.781**
Random initialization	0.339	0.618

Table 3: Test metrics on the clean test dataset. SSL training configuration - SimCLR, epochs = 1000, batch size = 2048, GTSRB dataset.

method	**top-1 accuracy**	**top-5 accuracy**
Advanced white box	0.834	0.975
Advanced random color box	**0.850**	**0.982**
Random initialization	0.733	0.956

Even when compared to standard SSL, suggested augmentations may be useful in some cases with specific test augmentations. Approach with suggested augmentations can achieve comparable performance on the clean test dataset and it achieves better performance on the specific test augmentations. See Table 4 and Table 5.

Table 4: Test metrics for the three random color boxes augmentation. SSL training configuration - SimSiam, epochs = 200, batch size = 256, STL-10 dataset. SimSiam is the augmentation method used in the original paper.

method	**top-1 accuracy**	**top-5 accuracy**
Advanced white box	0.698	0.984
Advanced random color box	**0.727**	**0.986**
SimSiam	0.598	0.957

Table 5: Test metrics on the clean test dataset. SSL training configuration - SimSiam, epochs = 200, batch size = 256, STL-10 dataset. SimSiam is the augmentation method used in the original paper.

method	**top-1 accuracy**	**top-5 accuracy**
Advanced white box	0.760	0.989
Advanced random color box	**0.773**	0.991
SimSiam	0.759	**0.992**

C AUGMENTATIONS

The images below show different types of augmentations for SSL training and for downstream training.

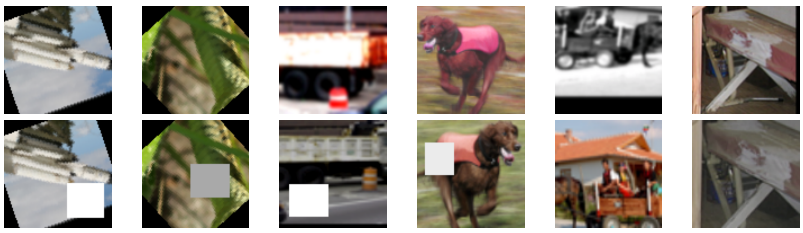

Figure 2: Augmentations of views for the SSL training on the STL-10 dataset. The augmentations are arranged from left to right - white box, random color box, advanced white box, advanced random color box, SimCLR, SimSiam.

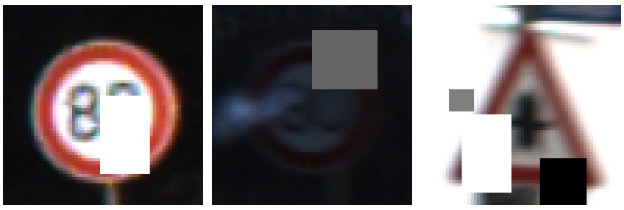

Figure 3: Augmented examples from the GTSRB test dataset for the downstream task. Types of test augmentation are arranged from left to right - white box, random color box, three random color boxes.

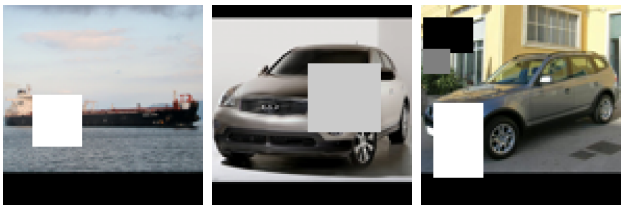

Figure 4: Augmented examples from the STL-10 test dataset for the downstream task. Types of test augmentation are arranged from left to right - white box, random color box, three random color boxes.