Data-efficient semantic segmentation via extremely perturbed data augmentation

Rafal Pytel, Tomasz Motyka

Delft University of technology

1 Introduction

Data can be seen as fuel for machine/deep learning models. Some of them i.e. Google's BERT require immense amounts of both data and computing resources. Gathering and labelling data as well as funding huge computing clusters is quite expensive and only a few companies or universities in the world have the resources to acquire tons of data and make use of it. In our project, we aimed to find methods for improving the effectiveness of the learning process given the limited data resources. In particular, we tackled the task of efficient learning methods for image semantic segmentation (primarily in the premise of 1st Visual Inductive Priors Workshop[1]). In this report we present new variation of strong augmentation CutMix - Progressive Sprinkles, presenting improved results over our Baseline by 3.8%. Moreover, we investigate how to tune the hyper-parameters of these advanced augmentations for the area of scene understanding, as this analysis is not explored enough for the domain of semantic segmentation.

2 Related work

Methods removing part - CutOut [1,2] of an image were already introduced to enhance generalization of image classification. Regional Dropout techniques are also used in object localization [3,4] to improve localization ability of CNNs. The idea of mixing two images into one was introduced with Mixup [5,6], which later span large selection of its variants [7,8,9] performing various types of transforms, e.g. feature level interpolation. Out of the success of CutOut and Mixup combination of them was created: CutMix [10]. After that, the variation of CutMix - CowMix [11] was proposed, which have cow patch-like mask instead of rectangular shape.

3 Dataset

3.1 MiniCity

MiniCity is a subset of CityScapes [2] created for the purpose of 1st Visual inductive Priors for Data-Efficient Deep Learning ECCV workshop [3]. It consists of images from

[1] https://vipriors.github.io/

[2] https://www.cityscapes-dataset.com/

[3] https://vipriors.github.io/

the dashboard of the car from multiple German and Swiss cities. There are 30 classes, organized into 8 categories: flat (road, sidewalk etc), human, vehicle, construction, object, nature, sky, void (ignored classes). MiniCity consists of 3 folders: train, val and test, each having 200, 100 and 200 images respectively. Regarding the diversity of samples, images are takes during daytime in a variety of seasons (spring, summer and fall), during good or medium weather conditions.
Important thing to mention here is that, since the dataset is part of the challange, the labels for the test set are not provided. There is a possibility of sending the results to the evaluation server however one can only send the results 10 times with the limit of 1 per day. Therefore, the results regarding the Minicity dataset are based on validation set. Having this in mind we were changing only 2 learning parameters provided by the baseline being the *batch size* and *learning patience*.

4 Methods

4.1 Baseline

As mentioned earlier, the MiniCity is the dataset created specifically for the ECCV workshop on Visual inductive Priors for Data-Efficient Deep Learning. This workshop provides also the baseline UNet model implementation with the training pipeline. All the methods we tried in our project were integrated into that baseline implementation.

4.2 CutOut

The CutOut is a quite simple augmentation method, proposed in [1], where the centre position, width and height of the rectangular crop are drawn from a beta distribution. This method enforces classifier to focus on a larger selection of features, not only the most discriminate ones, to avoid mislabeling in the situation with missing or occluded part of the image.

4.3 CutMix

As stated before, CutMix came from a combination of two other regularization methods: Cutout and Mixup. It differs from Cutout as in cut area the different image is pasted. As in Cutout position of centre, width and height are drawn from a beta distribution. The distinction between CutOut, Mixup and CutMix can be seen in the Figure 1.

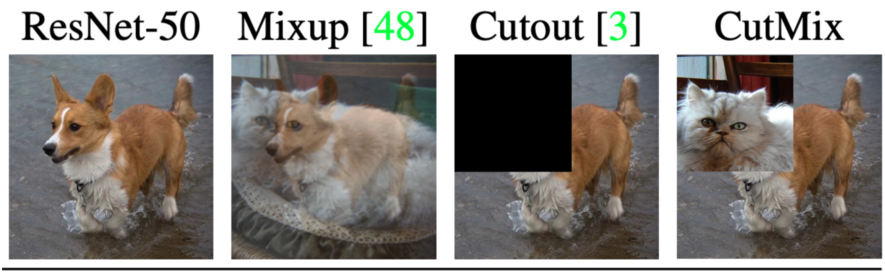

Fig. 1: Distinction between Mixup, Cutout and CutMix from CutMix paper

To understand how it looks on MiniCity dataset, please refer to Figure 2.

Fig. 2: Example of cutmix in semantic segmentation setup.

4.4 Progressive Sprinkles and Cutmix Sprinkles

Progressive Sprinkles is a variant of CutOut where instead of having one larger patch, there are multiple smaller blackout patches. Unfortunately, there is no research paper presenting this approach, however, according to blogpost [4] it produces SOTA results on Image classification NIH malaria dataset [5].
Out of the success of Progressive Sprinkles, we designed our method, called CutMix Sprinkles, which essentially is a variant of CutMix in which simple CutMix with smaller patches is invoked multiple times within certain image. In such a setup, the final image is a combination of multiple images, instead of only two images as it is in CutMix or CowMix case.
To understand how does CutMix Sprinkles look on MiniCity, refer to Figure 3.

[4] https://medium.com/@lessw/progressive-sprinkles-a-new-data-augmentation-for-cnns-and-helps-achieve-new-98-nih-malaria-6056965f671a
[5] https://lhncbc.nlm.nih.gov/publication/pub9932

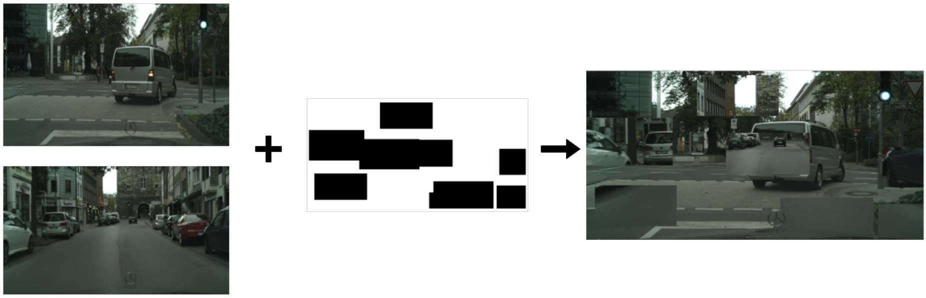

Fig. 3: Example of CutMix Sprinkles in semantic segmentation setup.

5 Results

5.1 Experimental setup

For experiments, UNet [12] with no pre-training was used, with SGD optimizer, cross-entropy loss, weight decay of 1e-4, trained for 300 epochs. In terms of the learning rate, ReduceLROnPlateau scheduler was used with starting learning rate on the level of 1e-2, finishing at 1e-5 and 15 epochs *learning patience*. The whole training was run for 300 epochs to provide stable results. Each image is firstly resized to 512x1024, then random scaling of (0.7,1.3), after that random crop of size 384x768 is produced and then random colour jitter on the level of 0.3 is applied.
Proposed method - CutMix Sprinkles had 0.5 probability of being applied.

5.2 MiniCity

Tuning the batch size According to our analysis (Table 1), we noticed that using different batch size influences the results, as with smaller batch size for MiniCity we observed higher results for proposed method. As a baseline, we use simple UNet with basic data augmentations we mentioned in Experimental setup. By adding CutMix Sprinkles we observed 3.8% improvement in performance.

Batch size	UNet	UNet + C. Sprinkles
2	41.4	45.2
5	41.0	44.5

Table 1: Fine-tuning the batch size for advanced augmentations.

Tuning the augmentations hyper-parameters While we investigated influence of batch size in MiniCity we also observed that proposed augmentation method had multiple parameters to be tuned (number of boxes, beta used for location and size of boxes).

Considering the results from previous section all experiments were conducted using batch size of 2.
In our method(Table 2) we checked the influence of two parameters, number of boxes, and Beta. We observed that our method performs the best with a smaller number of boxes (46.1 % for 4 boxes using beta 2.0). This could be since with a *batch size* of 2, with a large number of patches, target image which should be a combination of two images becomes fully dominated by one of the images. This behaviour is counter-productive as we want to create examples which are not available in the dataset. Moreover, we observe that this method provides significantly better results for betas different than 1.0 (non-Uniform distribution).

Beta	**Num boxes = 4**	**Num boxes = 8**	**Num boxes = 12**
0.5	45.8	45.9	45.3
1.0	44.8	43.1	43.3
2.0	46.1	44.2	44.1

Table 2: Fine-tuning hyper parameters of CutMix Sprinkles (beta and number of clusters).

5.3 Final submission

For our submission we used setup matching the one presented in Experimental setup, with addition of our method, called Cutmix Sprinkles, having 4 boxes and beta equal to 2.0. For that solution we got 46% on validation set and 43% on test set.

6 Conclusions

When tackling the problem of finding data-efficient methods for semantic segmentation, we tried a lot of different methods. In the end we managed to improve the performance by relatively significant amount using more advanced data augmentation. Although we are aware that all the methods we tried, did not have much chance to be super successful in this scenario and the problem of data-efficient learning would require some new innovative architectures or mechanisms introducing visual-prior knowledge, we still gained few insights. We learned that complex data augmentations, especially the ones with higher dimensionality of perturbations (CutMix Sprinkles), lead to improved generalization. During our research we were also able to get better understanding of how each hyper-parameter of augmentation influences the training results.
As our results show quite interesting insight into batch size with usage of strong augmentations we plan to extend that research to validate observed insight.

References

1. Devries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. CoRR **abs/1708.04552** (2017)

2. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. CoRR **abs/1708.04896** (2017)
3. Singh, K.K., Lee, Y.J.: Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). (2017) 3544–3553
4. Choe, J., Lee, S., Shim, H.: Attention-based dropout layer for weakly supervised single object localization and semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020) 1–1
5. Zhang, H., Cissé, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. CoRR **abs/1710.09412** (2017)
6. Tokozume, Y., Ushiku, Y., Harada, T.: Between-class learning for image classification. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2018) 5486–5494
7. Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D., Bengio, Y.: Manifold mixup: Better representations by interpolating hidden states. In Chaudhuri, K., Salakhutdinov, R., eds.: Proceedings of the 36th International Conference on Machine Learning. Volume 97 of Proceedings of Machine Learning Research., Long Beach, California, USA, PMLR (09–15 Jun 2019) 6438–6447
8. Summers, C., Dinneen, M.J.: Improved mixed-example data augmentation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). (2019) 1262–1270
9. Guo, H., Mao, Y., Zhang, R.: Mixup as locally linear out-of-manifold regularization. CoRR **abs/1809.02499** (2018)
10. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. CoRR **abs/1905.04899** (2019)
11. French, G., Oliver, A., Salimans, T.: Milking cowmask for semi-supervised image classification. ArXiv **abs/2003.12022** (2020)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015)