COLORFUL CUTOUT: ENHANCING IMAGE DATA AUGMENTATION WITH CURRICULUM LEARNING

Juhwan Choi and Youngbin Kim
Chung-Ang University, Seoul, Republic of Korea
{gold5230, ybkim85}@cau.ac.kr

ABSTRACT

Data augmentation is one of the regularization strategies for the training of deep learning models, which enhances generalizability and prevents overfitting, leading to performance improvement. Although researchers have proposed various data augmentation techniques, they often lack consideration for the difficulty of augmented data. Recently, another line of research suggests incorporating the concept of curriculum learning with data augmentation in the field of natural language processing. In this study, we adopt curriculum data augmentation for image data augmentation and propose **colorful cutout**, which gradually increases the noise and difficulty introduced in the augmented image. Our experimental results highlight the possibility of curriculum data augmentation for image data. We publicly released our source code to improve the reproducibility of our study.

1 INTRODUCTION

Data augmentation is an important regularization trick to train the deep learning model that aims to improve generalization ability and prevent overfitting (Yang et al., 2022). From the basic manipulation of input images, such as cropping, rotating, and jittering, data augmentation techniques for image data have evolved. For example, cutout and random erasing (DeVries & Taylor, 2017; Zhong et al., 2020) augmentation suggested a dropout strategy on the input image level by erasing a portion of a given image. After mixup (Zhang et al., 2018) introduced the concept of vicinal risk minimization through the mixture of two images, cutmix (Yun et al., 2019) proposed a strategy that combines cutout and mixup.

However, previous approaches have limited considerations about the difficulty of augmented data. It is widely accepted that a well-defined training procedure with the consideration of the difficulty of given data can enhance the performance of the trained model (Bengio et al., 2009; Soviany et al., 2022). Recently, researchers have been exploring the combination of data augmentation and curriculum learning in the context of curriculum data augmentation (Wei et al., 2021; Ye et al., 2021; Lu & Lam, 2023). Nonetheless, these approaches are mainly performed for the text data.

In this paper, we propose a novel curriculum data augmentation technique for image data. Specifically, it first introduces the colorization into cutout, which originally erases the portion of a given image. Additionally, through the division of the erasure box and filling the sub-regions with different colors, we are allowed to adjust the difficulty of the augmented image. To the best of our knowledge, this is the first study that pioneers curriculum data augmentation in the computer vision field. Our comprehensive experiment on various models and datasets demonstrates the effectiveness of our method, highlighting the advantage of curriculum data augmentation.

2 METHOD

We first briefly explain the procedure of traditional cutout. From the given image $\boldsymbol{x}$, it randomly selects a box region with a fixed size. After the selection, the box is erased and filled with zero value. Instead of simple erasure, our proposed colorful cutout fills the given box with a random color. This colorization establishes additional variation in augmented images, addressing a common limitation of previous methods and contributing to performance gain (Zhang & Ma, 2022).

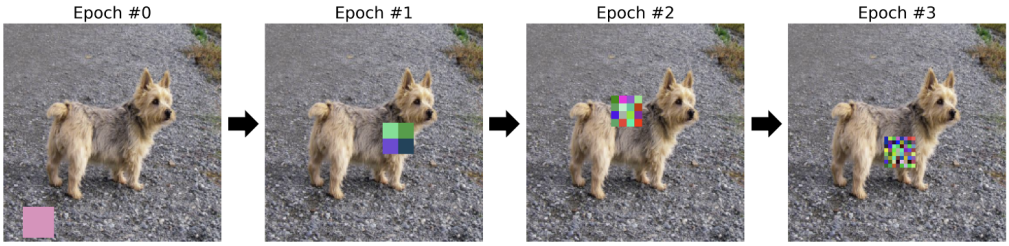

Figure 1: As the training procedure progresses, colorful cutout introduces more complex and difficult noise into augmented images.

Table 1: Accuracy (%) for each model and augmentation techniques across three datasets. "C10", "C100", and "TI" symbolizes CIFAR-10, CIFAR-100, and Tiny ImageNet, respectively.

	ResNet50			**EfficientNet-B0**			**ViT-B/16**		
Dataset	C10	C100	TI	C10	C100	TI	C10	C100	TI
Baseline	94.82	80.56	73.09	96.48	82.38	78.25	95.58	83.94	81.54
Cutout	95.49	80.97	73.52	96.56	82.53	78.41	96.08	84.21	81.49
Mixup	95.56	81.15	73.24	96.63	82.50	78.26	96.45	84.25	81.48
CutMix	95.67	81.45	73.63	96.67	82.96	78.53	96.27	84.32	81.82
Ours w/o Curr.	95.16	81.15	73.61	96.72	82.92	78.32	96.35	84.20	82.15
Ours	**95.70**	**81.57**	**73.81**	**96.81**	**83.37**	**78.65**	**96.55**	**84.36**	**82.36**

Additionally, colorful cutout introduces the concept of curriculum data augmentation through the division of the erasure box into sub-regions. Each of these sub-regions could have different colors. As the number of sub-regions increases, the erasure box becomes more tangled, resulting in more difficult samples as the training progresses. Figure 1 demonstrates the gradual increment of difficulty as the training progresses. Please refer to Appendix C for the pseudo-code of colorful cutout.

3 Experiment

We conducted an experiment to evaluate the effectiveness of our proposed method. First, we adopted three different datasets, CIFAR-10, CIFAR-100 (Krizhevsky et al., 2009), and Tiny ImageNet (Le & Yang, 2015) for evaluation. Second, we compared our methods against various previous augmentation techniques, including traditional cutout (DeVries & Taylor, 2017), mixup (Zhang et al., 2018), and cutmix (Yun et al., 2019). Last, we applied these methods on three different models, CNN-based ResNet50 (He et al., 2016) and EfficientNet-B0 (Tan & Le, 2019), and Transformer-based ViT-B/16 (Dosovitskiy et al., 2020). Please refer to Appendix A for more details.

Table 1 displays the experimental result. The results demonstrate a significant improvement in model performance with colorful cutout compared to other methods, particularly traditional cutout. Additionally, our ablation experiment on colorful cutout without the curriculum data augmentation shows similar performance to cutout, which suggests the curriculum data augmentation plays an important role for enhancing the performance of the model. This shows the potentiality of curriculum data augmentation in image data augmentation.

4 Conclusion

In this paper, we proposed a simple yet effective augmentation strategy that incorporates the concept of curriculum data augmentation into the computer vision field. The experimental results highlight the effectiveness of our approach and the possibility of curriculum image augmentation. Future research could investigate applying curriculum data augmentation to other image augmentation strategies and introducing soft labels to augmented data considering its difficulty (Choi et al., 2023).

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2022R1C1C1008534), and Institute for Information & communications Technology Planning & Evaluation (IITP) through the Korea government (MSIT) under Grant No. 2021-0-01341 (Artificial Intelligence Graduate School Program, Chung-Ang University).

URM STATEMENT

First author Juhwan Choi meets the URM criteria of ICLR 2024 Tiny Papers Track. He is outside the range of 30-50 years, non-white researcher.

REFERENCES

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

Juhwan Choi, Kyohoon Jin, Junho Lee, Sangmin Song, and YoungBin Kim. Softeda: Rethinking rule-based data augmentation with soft labels. In *ICLR 2023 Tiny Papers*, 2023.

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

Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7), 2015.

Quentin Lhoest, Albert Villanova del Moral, Yacine Jernite, Abhishek Thakur, Patrick von Platen, Suraj Patil, Julien Chaumond, Mariama Drame, Julien Plu, Lewis Tunstall, et al. Datasets: A community library for natural language processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 175–184, 2021.

Hongyuan Lu and Wai Lam. Pcc: Paraphrasing with bottom-k sampling and cyclic learning for curriculum data augmentation. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 68–82, 2023.

TorchVision maintainers and contributors. Torchvision: Pytorch's computer vision library. `https://github.com/pytorch/vision`, 2016.

Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey. *International Journal of Computer Vision*, 130(6):1526–1565, 2022.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Jason Wei, Chengyu Huang, Soroush Vosoughi, Yu Cheng, and Shiqi Xu. Few-shot text classification with triplet networks, data augmentation, and curriculum learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5493–5500, 2021.

Suorong Yang, Weikang Xiao, Mengcheng Zhang, Suhan Guo, Jian Zhao, and Furao Shen. Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*, 2022.

Seonghyeon Ye, Jiseon Kim, and Alice Oh. Efficient contrastive learning via novel data augmentation and curriculum learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1832–1838, 2021.

Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6023–6032, 2019.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018.

Linfeng Zhang and Kaisheng Ma. A good data augmentation policy is not all you need: A multi-task learning perspective. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 13001–13008, 2020.

A Implementation Details

This section describes implementation details and setups for reproduction. Please refer to our source code for more detailed information. [1]

Model Implementation. Every three models were based on the pre-trained checkpoints on ImageNet offered by TorchVision (maintainers & contributors, 2016) library. After the feature extraction from the pre-trained models, a two-layer classification with a dropout layer of $p = 0.2$ and ReLU activation is followed.

Every input image is resized to 256×256 and randomly cropped into 224×224 size in the training procedure. In the validation and test procedure, an image with 224×224 size is obtained from the center of the original image.

Augmentation Implementation. For traditional cutout, cutmix, and our method, the size of the box w is set to 32×32 for every model. For the mixup and cutmix method, we used $\alpha = 0.2$, where α denotes the bound of the beta distribution that determines the mixup ratio.

Colorful cutout increases the number of sub-regions as the training epoch increases. Specifically, the number of sub-regions is defined as $2^{N_{epoch}}$. Each sub-region is assigned different random colors. We set the initial N_{epoch} starts at 0, indicating that there is no sub-region in the first epoch. Please refer to Figure 1 as an example.

Datasets. CIFAR-10 Krizhevsky et al. (2009) is an image classification dataset consisting of 50,000 training images and 10,000 test images in 10 classes. CIFAR-100 is an extended version of CIFAR-10, which is composed of 100 classes. Tiny ImageNet (Le & Yang, 2015) is a subset of ImageNet (Deng et al., 2009), which has 200 classes and 100,000 training images. Three datasets were downloaded from Datasets library (Lhoest et al., 2021) operated by Hugging Face. As there is no predefined validation set exists, we randomly selected 10% of the training data as the validation set.

Hyperparameters. Adam (Kingma & Ba, 2015) has been deployed as the optimizer, with a learning rate of 5e-5. We trained each model for 5 epochs with a batch size of 32. We applied label smoothing (Szegedy et al., 2016) with a smoothing factor 0.05 for every model.

[1] `https://github.com/c-juhwan/colorful-cutout-aug`

Further Details. Every experiment was performed using a single NVIDIA RTX 3090 GPU. We trained the model with our method for 75.7 minutes on Tiny ImageNet, while cutout baseline took 74.2 minutes.

B COMPARISON BETWEEN OTHER TECHNIQUES

We provide an example of colorful cutout compared to other methods on the same image.

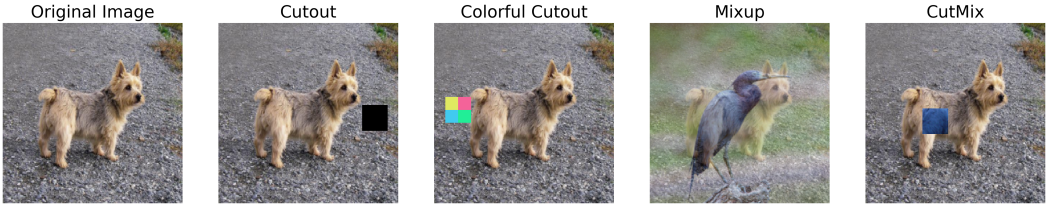

Figure 2: An example of our proposed colorful cutout compared to previous data augmentation methods.

C ALGORITHM OF COLORFUL CUTOUT

We provide a pseudo-code for colorful cutout.

Algorithm 1 The procedure of colorful cutout.

Require: Given image $\boldsymbol{x}$, pre-defined size of erasure box w, current epoch index N_{epoch}

1: Randomly generate erasure box B with size of $w \times w$ from $\boldsymbol{x}$
2: Get the number of sub-region $N_{region} = 2^{N_{epoch}}$
3: Divide B into N_{region} squared sub-regions
4: Fill divided B with N_{region} random colors
5: Return augmented image $\hat{\boldsymbol{x}}$