Improved Regularization for Convolutional Neural Networks

Chen Dong,
Weisheng Jin,
Xinyi Xie

Duke
PRATT SCHOOL of ENGINEERING

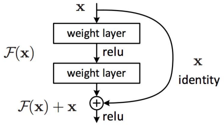
Introduction

- Explored advanced regularization techniques (cutout, mixup, self-supervised rotation) on ResNet-20 for CIFAR-10.
- Assessed performance against image corruptions and white-box adversarial attacks (FGSM, rFGSM, random noise, PGD).
- Identified the most effective strategy for enhancing CNN robustness and generalization.
- Provided insights into mitigating overfitting in complex neural network architectures.

ResNet Baseline Model

The ResNet-20 model on CIFAR-10 is employed for all subsequent experiments.

The ResNet model comprises:
Convolutional Layers
Shortcut Connections
Fully Connected Layer

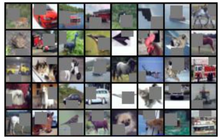


Residual Block for ResNet

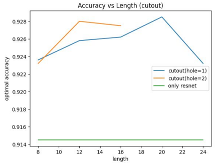
The model is trained with SGD, starting at a learning rate of 0.1, 1e-4 weight decay, and Cross-Entropy Loss. Learning rate drops at epochs 60, 120, and 160 during the 200-epoch training, using a batch size of 128. The best validation accuracy is 91.45%.

Cutout Regularizer

The dataset is augmented by randomly masking out patches from images, varying the number (n_holes) and size (length) of patches during training. The best validation accuracy is 92.85% with n_holes=1, length=20.



Cutout on the Image

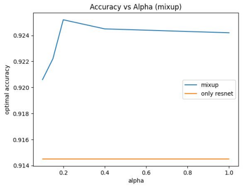


Test Accuracy with Different Lengths and N_holes

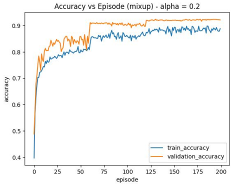
Mixup Regularizer

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j$$
$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j$$

The dataset is augmented by randomly selecting two images and combine them linearly, the combination ratio λ is sampled from Beta(α,α).
The model is trained with different mixup hyperparameter α to evaluate the performance of mixup regularizer. The best validation accuracy is 92.52% with α = 0.2.



Test Accuracy with Different Alphas

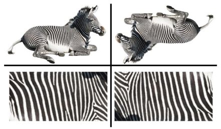


Training and Validation Accuracy with Best Parameter alpha=0.2

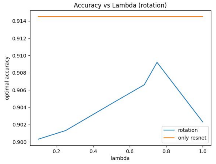
Self-supervised Learning with Auxiliary Rotation Head

The dataset is augmented with images rotated at 0, 90, 180, and 270 degrees. Training optimizes both image classification and rotation prediction, using a combined loss function with different values of hyperparameter λ. The best validation accuracy is 90.92% with λ = 0.75.

$$L(x, y; \theta) = L_{CE}(y, p(y|x; \theta)) + \sum_{r \in \{0°, 90°, 180°, 270°\}} L_{CE}(\text{one\_hot}(r), p_{\text{rot\_head}}(r | R_r(x); \theta))$$



Predicting rotation requires understanding an object's shape,
as texture alone can't reliably indicate if a zebra is inverted.
Using self-supervised rotations in training can thus improve robustness.



Test Accuracy with Different Lambdas

Robustness Testing for Corruptions

Five types of corruptions have been applied to the test dataset. The following are the accuracy values for each of the five models when subjected to these corruptions.

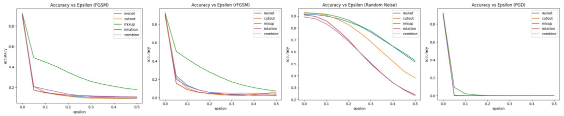
Gaussian Noise

Corruption	gaussian noise	shot noise	impulse noise	brightness	contrast
Original Resnet	0.299	0.2993	0.5814	0.9054	0.7881
Cutout	0.2132	0.2215	0.4695	0.9215	0.8214
Mixup	0.3647	0.3822	0.5938	0.918	0.8356
Rotation	0.1501	0.161	0.3579	0.9003	0.7804
Combination	0.1537	0.1648	0.3032	0.8878	0.7933

Test Accuracy of Five Models Subjected to Five Distinct Corruptions

Robustness Testing for Adversarial Attack

Attacks such as FGSM, rFGSM, Random Noise, and PGD have been implemented on the test dataset. To assess the robustness of the five models, epsilon values ranging from 0 to 0.5 were utilized.



Comparative Robustness of Five Models Against Four Adversarial Attacks

References

DeVries, T. et al. (2017). "Improved regularization of convolutional neural networks with cutout." In: arXiv preprint arXiv:1708.04552.

Hongyi Zhang et al. (2018). "Mixup: Beyond Empirical Risk Minimization." In: ICLR (Poster).

Dan Hendrycks et al. (2019). "Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty." In: NeurIPS.

He, K. et al. (2016). "Deep residual learning for image recognition." In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

ImageCorruptions Contributors. (GitHub Repository) "Bethge Lab Image Corruptions." In: https://github.com/bethgelab/imagecorruptions

Goodfellow, I. J. et al. (2014). "Explaining and harnessing adversarial examples." In: arXiv preprint arXiv:1412.6572.