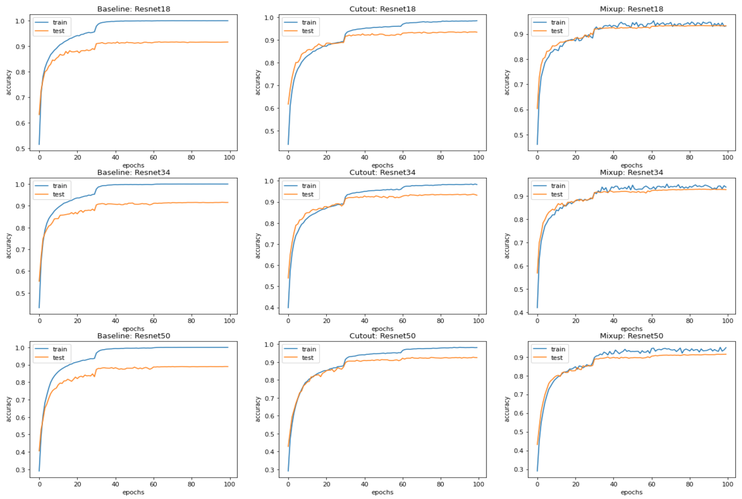
Improved Regularization of Convolutional Neural Networks

Xian Sun

Introduction & Background

Overfitting is a common problem for Convolutional Neural Networks. And this project is mainly exploring the combination of two regularization techniques: cutout and mixup to address overfitting problems on Resnet. I found out that these two methods have very little computational cost and can solve overfitting effectively for Resnet on CIFAR10 , SVHN and Fashion MNIST. There are 3 main following sections in this project:

- Implemented cutout and mixup to address overfitting problems for ResNet18,34 and 50.
- Do grid search for a good combination of cutout and mixup for ResNet18-CIFAR10.
- Transfer the best combination of cutout and mixup to SVHN and Fashion MNIST, trained on ResNet18



Implemented cutout and mixup on ResNet18, 34 and 50.

Cutout

To implement cutout, we simply apply a fixed-size zero- mask to a random location of each input image during each epoch of training.

Mixup

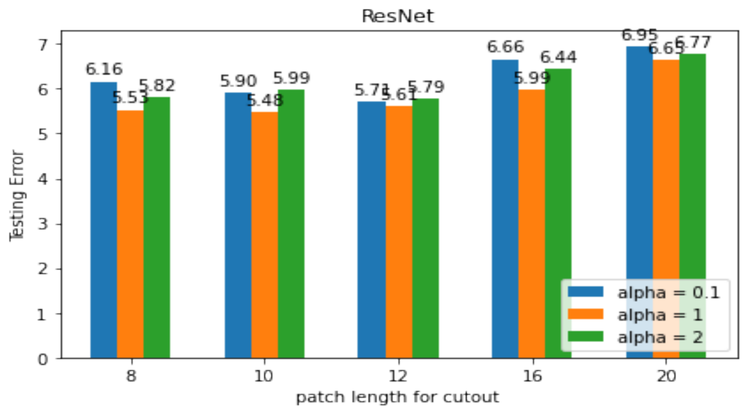
To implement Mixup, we simply do data transformation during each batch of training, $\tilde{x} = \lambda x_i + (1 - \lambda)x_j$, $\tilde{y} = \lambda y_i + (1 - \lambda)y_j$, where (x_i, y_i),(x_j, y_j) are two feature-target vectors drawn at random from the training data and $\lambda \sim Beta(\alpha, \alpha)$.

Experiments on CIFAR10-Resnet18,34,50

With fixed α = 0.1 for mixup and n_holes = 1 and length = 16 for cutout, all Resnets have a better performance on test accuracy than before. We can find the gap between train and test accuracy become smaller which means the overfitting is reduced. But for cutout there is not big difference between Resnets with different depths. For mixup, we can find the gap between train and test is becoming slightly larger as we improve the depth of Resnet, which means it is harder for mixup to solve overfitting problems as the Resnet become deeper.

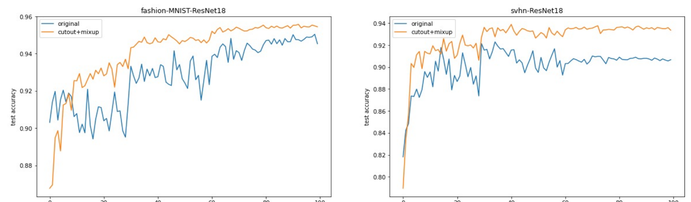
Grid search for a good combination of cutout and mixup for Resnet18-CIFAR10.

Based on CIFAR10, I did grid search on ResNet18 with alpha from [0.1,1, 2] for mixup and length from [8,10,12,16,18, 20] and tried to find the best combination where we can have the smallest test error. From the figure, we can notice that ResNet18 has the smallest test error 5.48% when I set the combination where patch length = 10 and alpha = 1.



Transfer the best combination of cutout and mixup to SVHN and Fashion MNIST

I trained ResNet18 on fashion MNIST and SVHN with the best combination of mixup and cutout, epochs = 100,and learning rate = 0.05. From the figure below, we can see after I transferred the best combination of mixup and cutout, the test accuracy is improved by 1% for Fashion-MNIST and 3% for svhn. Besides, we can see that this mixture of two techniques also make the objective curve more stable.



Conclusion

Cutout and mixup are two regularization techniques that can work well with each other together to solve overfitting problems for Resnet effectively on datasets like CIFAR10, SVHN and Fashion MNIST.

References

Moustapha Cisse et al. (2018). “mixup: Beyond Empirical Risk Minimization”. In: arXiv:1710.09412.

Terrance DeVries et al. (2017). “Improved Regularization of Convolutional Neural Networks with Cutout.” In: arXiv:1708.04552v2