# LA3: Efficient Label-Aware AutoAugment

Mingjun Zhao[1], Shan Lu[1], Zixuan Wang[2], Xiaoli Wang[2] and Di Niu[1]

[1] University of Alberta
[2] Platform and Content Group, Tencent
{zhao2,slu1,dniu}@ualberta.ca, {zackiewang,evexlwang}@tencent.com

## 1 Training Details of *LA3*

**Table 1.** Training hyperparameters for different networks on CIFAR-10, CIFAR-100 and ImageNet.

| Dataset | Model | Batch Size | LR | WD | Epoch |
|---|---|---|---|---|---|
| CIFAR-10 | WRN-40-2 | 128 | 0.1 | $5e-4$ | 600 |
| | WRN-28-10 | 128 | 0.1 | $5e-4$ | 200 |
| | Shake-Shake (26 2x96d) | 128 | 0.01 | $1e-3$ | $1,800$ |
| | Shake-Shake (26 2x112d) | 128 | 0.01 | $1e-3$ | $1,800$ |
| | PyramidNet+ShakeDrop | 128 | 0.1 | $5e-4$ | $1,800$ |
| CIFAR-100 | WRN-40-2 | 128 | 0.1 | $5e-4$ | 600 |
| | WRN-28-10 | 128 | 0.1 | $5e-4$ | 200 |
| | Shake-Shake (26 2x96d) | 128 | 0.05 | $5e-4$ | $1,800$ |
| | PyramidNet+ShakeDrop | 128 | 0.05 | $5e-4$ | $1,800$ |
| ImageNet | ResNet-50 | $1,024$ | 0.4 | $1e-4$ | 270 |
| | ResNet-50 (BA) | $1,024 \times 4$ | 0.4 | $1e-4$ | 270 |

In this section, we present the details of training hyperparameters of different target networks on CIFAR-10, CIFAR-100 and ImageNet.

For CIFAR-10 and CIFAR-100, we follow previous work and apply our searched policies on top of the baseline augmentations including random cropping the input image to $32 \times 32$ from the padded image, horizontally flipping it with 0.5 probability, and a Cutout operation with $16 \times 16$ pixels. For ImageNet, the searched policies are applied after random cropping, resizing to $224 \times 224$, and horizontal flipping with 0.5 probability.

All the networks are trained with SGD optimizer and cosine learning rate decay. In the training of ResNet-50 model, label smoothing is set to 0.1. Other training hyperparameters are shown in Table 1. In the Batch Augment (BA) version of ResNet-50, a training batch is composed of 4 copies of augmented $1,024$ samples.

## 2 Choice of $\alpha$

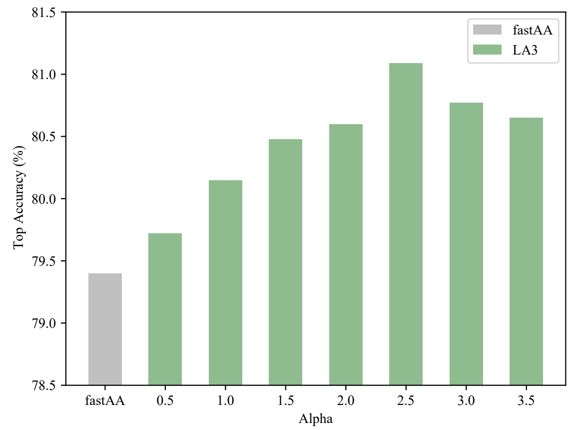


**Fig. 1.** The top-1 test accuracy of WRN-40-2 on CIFAR-100 verses different $\alpha$ values.

In our method, $\alpha$ is a hyperparameter in score calculation of augmentation triples to adjust the weight between the reward value and the redundancy value. To choose the optimal $\alpha$, we evaluate the performance of our proposed method with WRN-40-2 network on CIFAR-100 verses different $\alpha$ values from 0.5 to 3.5. From Figure 1, we can observe that the test accuracy increases with $\alpha$ before $\alpha = 2.5$ and decreases after. Therefore, we chose $\alpha = 2.5$ in our experiments. Note that our *LA3* method constantly beats FastAA with all choices of $\alpha$ values, which again confirms the effectiveness of our design.

## 3 Results on ViT

We have also conducted an experiment on ViT-Tiny to evaluate our *LA3* method. Due to unavailability of many baselines, we only include comparison results with two static methods, AA and FastAA. As shown in the following Table 2, *LA3* is effective and outperforms AA and FastAA on ViT-Tiny.

**Table 2.** ViT-Tiny top-1 accuracy on CIFAR-10 and CIFAR-100.

| | base | AA | FastAA | LA3 |
|---|---|---|---|---|
| CIFAR-10 | 86.07 | 87.39 | 86.83 | **87.83** |
| CIFAR-100 | 97.49 | 98.03 | 97.93 | **98.08** |