

Table 9: Ablation study of data augmentation for direction classification.

|  |  |  |  |
| --- | --- | --- | --- |
| Backbone MobileNetV3 small\_x0.35 MobileNetV3\_ small\_x0.5 | Accuracy 0.6288 0.6556 | Model Size (M) 22 23 | Inference Time (CPU, ms) 17 17.27 |
| MobileNetV3\_ small\_x1 | 0.6933 | 28 | 19.15 |
|  |  |  |  |
|  |  |  |  |

Table 10: Compares the performance of the different back. bones for text recognition. The number of channel in the head is 256.

and RandAugment are useful for text direction classifica- tion. Therefore, in PP-OCR, we use BDA (base data aug- mentation) and RandAugment to train a direction classifier. Table [8|shows the ablation study of input resolution and

mentation) and RandAugment to train a direction classifier. Table [8|shows the ablation study of input resolution and PACT quantization for direction classification. When the in- put resolution is adjusted from 3 32 100 to 3 48 192 The classification accuracy has improved but the prediction speed is basically unchanged. Furthermore, we also verified quantization strategy is effective in accelerating the predic. tion speed of the text direction classifier. The model size is reduced 45.9% and the inference time has accelerated 25.86%. Accuracy is slight promotion.

# 3.4Text Recognition

Table [10] compares the performance of the different back. bones for text recognition. The accuracy, the model size and the inference time of the different scales of Mo- bileNetV3 change greatly. In PP-OCR, we choose Mo- bileNetV3\_small\_x0.5 to balance accuracy and efficiency.

Table 8: Ablation study of input resolution and PACT quantization for direction classification.

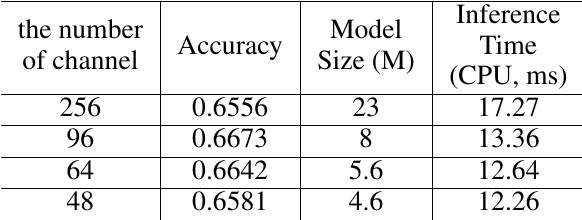


Table 11: Ablation study of the number of channel in the head for text recognition. The data augmentation is only used BDA.

Table [11compares the number of channel in the CRNN head for text recognition. Reduce the number of channe. from 256 to 48, the model size is reduced from 23M to 4.6M and the inference time has accelerated nearly 30%. However, the accuracy will not be affected. We can see the number of. channel in the head has a great influence on the model size of a lightweight text recognizer.

of a lightweight text recognizer. Tabel12 shows the ablation study of data augmentation, cosine learning rate decay, the stride of the second down sampling feature map, regularization parameters L2\_decay and learning rate warm-up for text recognition. To verify the advantages of each strategy, the setting of

To verify the advantages of each strategy, the setting of the basic experimental is the strategy S1. When using BDA, the accuracy will be improved 3.12%. Data augmentation is very necessary for text recognition. When we adopt the cosine learning rate decay further, the accuracy will be im- proved 1.47%. The cosine learning rate is an effective strat- egy for text recognition. Next, when we increase the fea- ture map resolution and reduce the stride of the second down sampling feature map from (2,1) to (1,1), the accuracy will be improved 5.27%. Then, when we adjust the regulariza- tion parameters L2\_decay from 0 to 1e - 5 further, the accu racy will be improved 3.4%. The feature map resolution and L2.decay have a great influence on the performance. Final. using learning rate warm-up, the accuracy will be improved 0.62%. Using TIA data augmentation, the accuracy will be improved 0.91%. Learning rate warm-up and TIA also are effective strategies for text recognition.

effective strategies for text recognition. Tabel13]shows the ablation study of PACT quantization for text recognition. When we use PACT quantization, the model size is reduced 67.39% and the inference time has ac- celerated 8.3%. Since there was no quantification on LSTM, The acceleration is not obvious. However, accuracy achieves a significant improvement. Therefore, PACT quantization also is an effective strategy for reducing the model size of. a text recognizer.