some image processing operations to train a text recog- nizer, such as rotation, perspective distortion, motion blur and Gaussian noise. Those processes are referred to as BDA (Base Data Augmentation) for short. They are randomly added to the training images. The experiment shows that BDA also is useful for the direction classifier training. Be- sides BDA, some new data augmentation operations are proposed recently for improving the effect of image clas- sification, for example, AutoAugment (Cubuk et al.2019) RandAugment (Cubuk et al.2020), CutOut DeVries and Taylor2017), RandErasing (Zhong et al.2020), HideAnd- Seek (Singh and Lee[2017), GridMask (Chen 2020), Mixup (Zhang et al.|2017) and Cutmix (Yun et al.|2019). But the experiments show that most of them don't work for the direction classifier training except for RandAugment and RandErasing. RandAugment works best. Eventually, we add BDA and RandAugment to the training images of the direc- tion classification.

tion classification. Input Resolution In general, when the input resolution of a normalized image is increased, accuracy will also be improved. Since the backbone of the direction classifier is very light, increasing the resolution properly will not lead to the computation time raise obviously. In the most of the previous text recognition methods, the height and width of a normalized image is set as 32 and 100, respectively. How- ever, in PP-OCR, the height and width is set as 48 and 192, respectively, to improve the accuracy of the direction classi- fer. PACT Quantization Quantization allows the neural net

PACT Quantization Quantization allows the neural net work model to have lower latency, smaller volume and lower computational power consumption. At present, quantiza- tion is mainly divided into two categories: offline quanti- zation and online quantization. Offline quantization refers to a fixed-point quantization method that uses methods such as KL divergence and moving average to determine quan- tization parameters and does not require retraining. Online quantization is to determine quantization parameters dur ing the training process, which can provide less quantization loss than offline quantization mode. PACT (PArameterized Clipping acTivation) (Choi et al.

loss than offline quantization mode. PACT (PArameterized Clipping acTivation) (Choi et al. 2018) is a new online quantification method that removes some outliers from the activations in advance. After remov- ing the outliers, the model can learn more appropriate quan- titative scales. The formula for PACT to preprocess the acti- vations is as follows:

The preprocessing of the activation value of the ordinary PACT method is based on the ReLU function. All activation values greater than a certain threshold are truncated. How- ever, the activation functions in MobileNetV3 are not only ReLU, but also hard swish. Using ordinary PACT quantiza- tion leads to a higher quantization loss. Therefore, we mod- ify the formula of the activations preprocessing as follows to reduce the quantization loss.

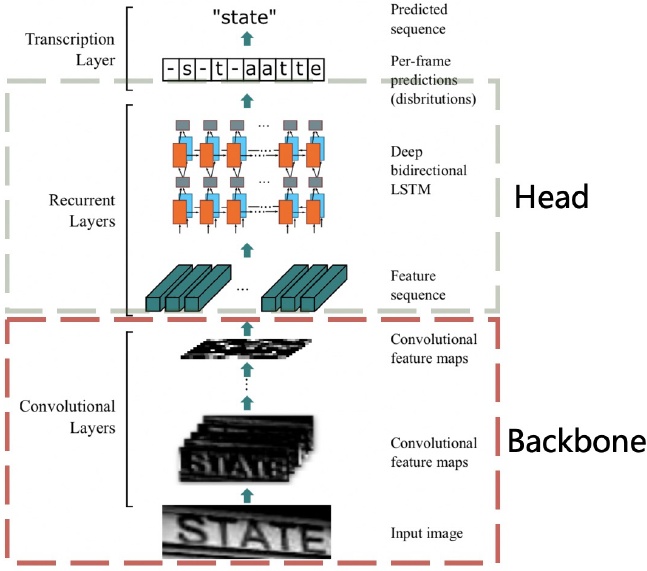


Figure 10: Architecture of the text recognizer CRNN. This figure comes from the paper (Shi, Bai, and Yao 2016). The red and gray rectangles show the backbone and head of the text recognizer separately.

We used the improved PACT quantification method to quantify the direction classifier model. In addition, L2 reg. ularization with a coefficient of O.001 is added to the PACT parameters to improve the model robustness. The implementation of the above FPGM Pruner and

parameters to improve the model robustness. The implementation of the above FPGM Pruner and PACT quantization is based on PaddleSlim PaddleSlim is a toolkit for model compression. It contains a collection of compression strategies, such as pruning, fixed point quan- tization, knowledge distillation, hyperparameter searching neural architecture search.

# 2.3Text Recognition

In this section, the details of nine strategies for enhancing the model ability or reducing the model size of a text recognizer will be introduced. Figure10shows the architecture of the text recognizer CRNN.

text recognizer CRNN. Light Backbone We also adopt MobileNetV3 as the backbone of the text recognizer which is the same as the text detection. MobileNetV3\_small\_x0.5 is selected to bal- ance accuracy and efficiency empirically. If you're not that sensitive to the model size, MobileNetV3\_small\_x1.0 is also a good choice. The model size is just increased by 2M, the accuracy is improved obviously Data Augmentation Besides BDA (Base Data Augmen-

Data Augmentation Besides BDA (Base Data Augmen- tation) which is often used in text recognition as mentioned earlier, TIA (Luo et al. 2020) also is an effective data aug- mentation method for text recognition. As shown in Figure