Figure 7: Architecture of the SE block. This figure comes from the paper (Hu, Shen, and Sun|2018)

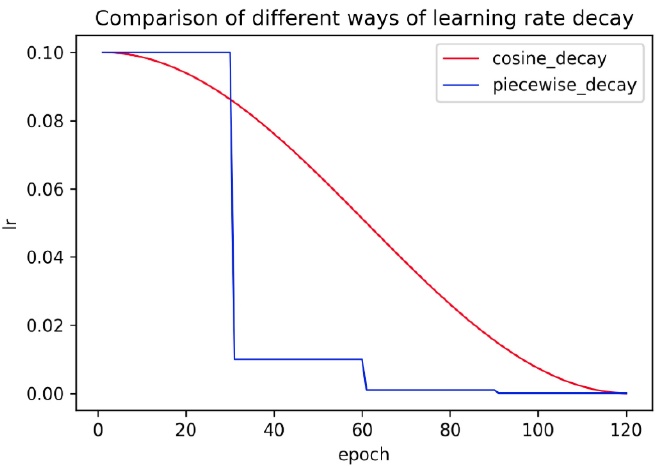


Figure 8: Comparison of different ways of learning rate de- cay.

short). The probability map and the threshold map are gen- erated from the fused feature map with convolutions which are also associated with the above inner channels. Thus in- ner channels has a great influence on the model size. When inner channels is reduced from 256 to 96, the model size is reduced from 7M to 4.1M, but the accuracy declines slightly

reduced from 7M to 4.1M, but the accuracy declines slightly Remove SE SE is the short for squeeze-and-excitation (Hu, Shen, and Sun 2018). As shown in Figure7 SE blocks model inter-dependencies between channels explicitly and re-calibrate channel-wise feature responses adaptively. Be- cause SE blocks can improve the accuracy of the vision tasks obviously, the search space of MobileNetV3 contains them and numerous of SE blocks are in MobileNetV3 architec- ture. However, when the input resolution is large, such as 640 640, it is hard to estimate the channel-wise feature responses with the SE block. The accuracy improvement is limited, but the time cost is very high. When the SE blocks are removed from the backbone, the model size is reduced from 4.1M to 2.5M, but the accuracy has no effect.

from 4.1M to 2.5M, but the accuracy has no effect. Cosine Learning Rate Decay The learning rate is the hyperparameter to control the learning speed. The lower the learning rate, the slower the change of the loss value Though using a low learning rate can ensure that you will not miss any local minimum, but it also means that the con- vergence speed is slow. In the early stage of training, the weights are in random initialization state, so we can set a relatively large learning rate for faster convergence. In the late stage of training, the weights are close to the optimal values, so a relatively smaller learning rate should be used. Cosine learning rate decay has become the preferred learn-

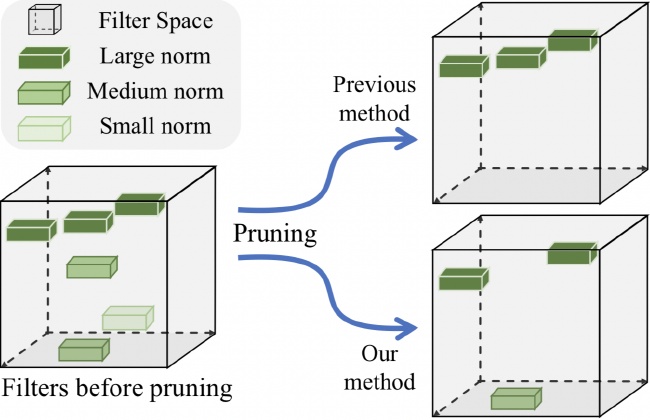


Figure 9: Illustration of FPGM Pruner. This figure comes from the paper (He et al. 2019b)

ing rate reduction strategy for improving model accuracy During the entire training process, cosine learning rate de cay keeps a relatively large learning rate, so its convergence is slower, but the final convergence accuracy is better. Figure 8compares the different ways of learning rate decay.

8compares the different ways of learning rate decay. Learning Rate Warm-up The paper (He et al.2019a) shows that using learning rate warm-up operation can help to improve the accuracy in the image classification. At the beginning of the training process, using a too large learning rate may result in numerical instability, a small learning rate is recommended to be used. When the training process is sta- ble, the initial learning rate is to be used. For text detection, the experiments show that this strategy also is effective.

the experiments show that this strategy also is effective. FPGM Pruner Pruning is another method to improve the inference efficiency of neural network model. In order to avoid the model performance degradation caused by the model pruning, we use FPGM (He et al. 2019b) to find the unimportant sub-network in original models. FPGM uses geometric median as the criterion and the each filter in a con- volution layer is considered as a point in Euclidean space. Then calculate the geometric median of these points and re. move the filters with the similar values, as shown in Figure The compress ratio of each layer is also important for prun- ing a model. Pruning every layer uniformly usually leads to significant performance degradation. In PP-OCR, the prun- ing sensitivity of each layer is calculated according to the method in (Li et al. 2016) and then used to evaluate the re- dundancy of each layer.

# 2.2Direction Classification

In this section, the details of four strategies for enhancing the model ability or reducing the model size of a direction classifier will be introduced.

Light Backbone We also adopt MobileNetV3 as the backbone of the direction classifier which is the same as the text detector. Because this task is relatively simple, we use MobileNetV3\_small\_x0.35 to balance accuracy and effi- ciency empirically. When using larger backbones, the accu- racy doesn't improve more.