

Figure 2: The framework of the proposed PP-OCR. The model size in the figure is about Chinese and English characters recognition. For alphanumeric symbols recognition, the model size of text recognition is from 1.6M to 0.9M. The rest of the models are the same size.



considering the cost. In particular, the OCR system need to be run on embedded devices in many scenarios, such as cell phones, which makes it necessary to consider the model size. Trade off model size and performance is difficult but of great value. In this paper, we propose a practical ultra lightweight OCR system, named as PP-OCR, which consists of three parts, text detection, detected boxes rectification and text. recognition as shown in Figure2

Text Detection The purpose of text detection is to locate the text area in the image. In PP-OCR, we use Differentiable Binarization (DB) (Liao et al. 2020) as text detector which is based on a simple segmentation network. The simple post processing of DB makes it very efficient. In order to further improve its effectiveness and efficiency, the following six strategies are used: light backbone, light head, remove SE module, cosine learning rate decay, learning rate warm-up, and FPGM pruner. Finally, the model size of the text detector is reduced to 1.4M.

is reduced to 1.4M. Detection Boxes Rectify Before recognizing the detected text, the text box needs to be transformed into a horizon- tal rectangle box for subsequent text recognition, which is

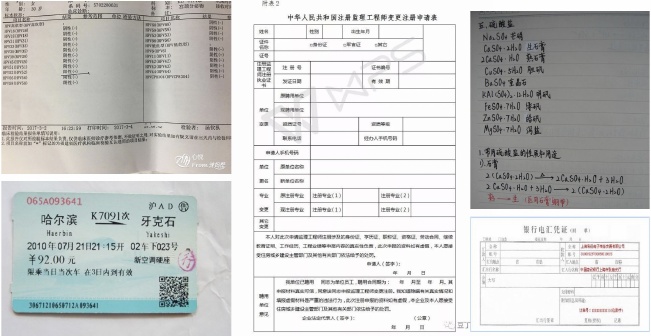


Figure 4: Some images contained document text..

easy to be achieved by geometric transformation as the de- tection frame is composed of four points. However, the rec- tified boxes may be reversed. Thus, a classifier is needed to determine the text direction. If a box is determined re- versed, further flipping is required. Training a text direction classifier is a simple image classification task. We adopt the following four strategies to enhance the model ability and reduce the model size: light backbone, data augmentation, input resolution and PACT quantization. Finally, the model. size of the text direction classifier is 50oKB.

Text Recognition In PP-OCR, we use CRNN (Shi, Bai, and Yao 2016) as text recognizer, which is widely used and practical for text recognition. CRNN integrates feature ex- traction and sequence modeling. It adopts the Connection-. ist Temporal Classification(CTC) loss to avoid the inconsis. tency between prediction and label. To enhance the model ability and reduce the model size of a text recognizer, the. following nine strategies are used: light backbone, data aug- mentation, cosine learning rate decay, feature map resolu-.