Supplementary Material

Abstract

In the supplementary material, we provide more details of our instance-segmentation model, and more visual results of our system. We first show more examples of the training data we used to train our text instance segmentation model in Sec. 1. We then show some examples about the ability of IA-CNN to handle imperfect text line inputs in Sec. 2. With simple post processing, we can recover the final text instance. We show more results for both our word instance segmentation model and end to end results with instance masks, instead of bounding box. This gives readers more intuition of our system.

1   Training Samples

We show several training examples in Fig. 1 and Fig. 2. Random scaling and blurry is conducted for each training sample on the fly. The green bounding boxes for each text instance is shown for each text block training sample. The ground truth of training IA-CNN and TL-CNN is conducted based on these ground truth. In Fig. we show several examples of negative training samples added to TL-CNN training and IA-CNN training. Note that for IA-CNN training, negative samples are also added by creating lines in the background region of a positive text block as mentioned in the paper.

2   Error Handling and Post Processing

Some error handling results are shown in Fig. 4. Note that the text lines extracted are not correct. But with simple post-processing to merge instance regions, we will be able to recover the final bounding box.

Figure 1: Some training samples of the instance segmentation model.
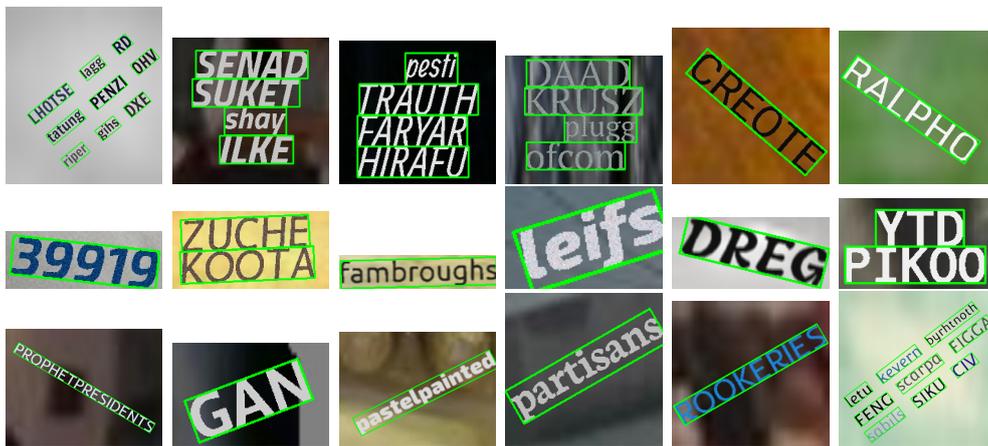
Figure 2: Some training samples of the instance segmentation model.
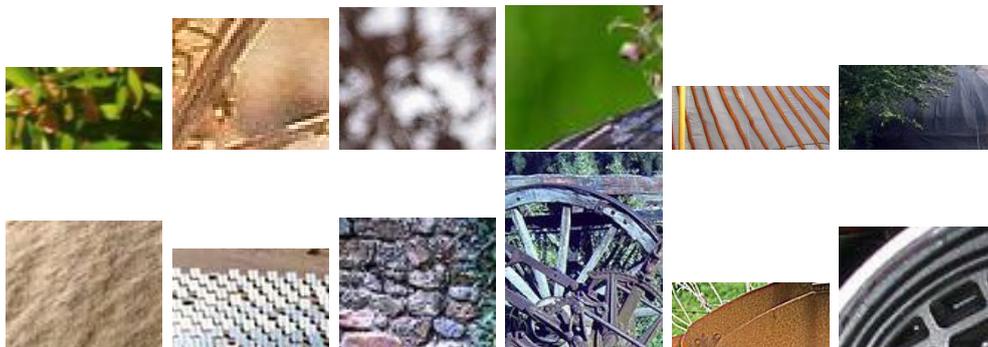
Figure 3: Some training negative samples of the instance segmentation model.
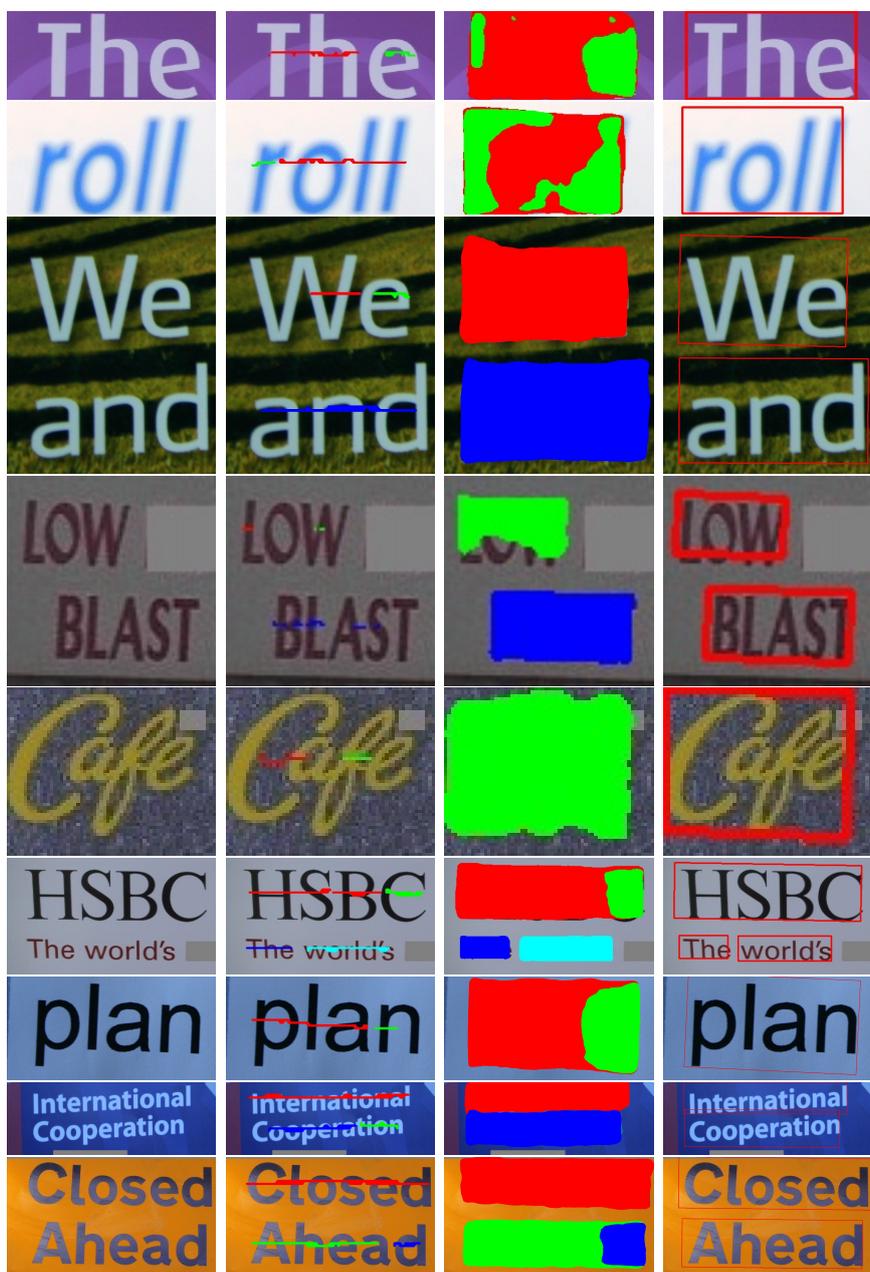
Figure 4: Some examples showing the error handling of IA-CNN step. From left to right: input text block image, extracted text line instance, raw instance segmentation result, final extracted instance box. Note that, the extracted text instances are not correct, but we can finally recover the results. In the raw instance segmentation results, each pixel might belong to two instance masks, since the operation of IA-CNN is done separately for each extracted text line. The instance mask draw later will cover the previous one for the raw instance results shown here.

Figure 5: More results of the word or text line instance segmentation

Figure 6: More results of the word or text line instance segmentation

Figure 7: More results of the final segmentation map. Better to look when zooming. Left: Input image. Right: Output Segmentation map
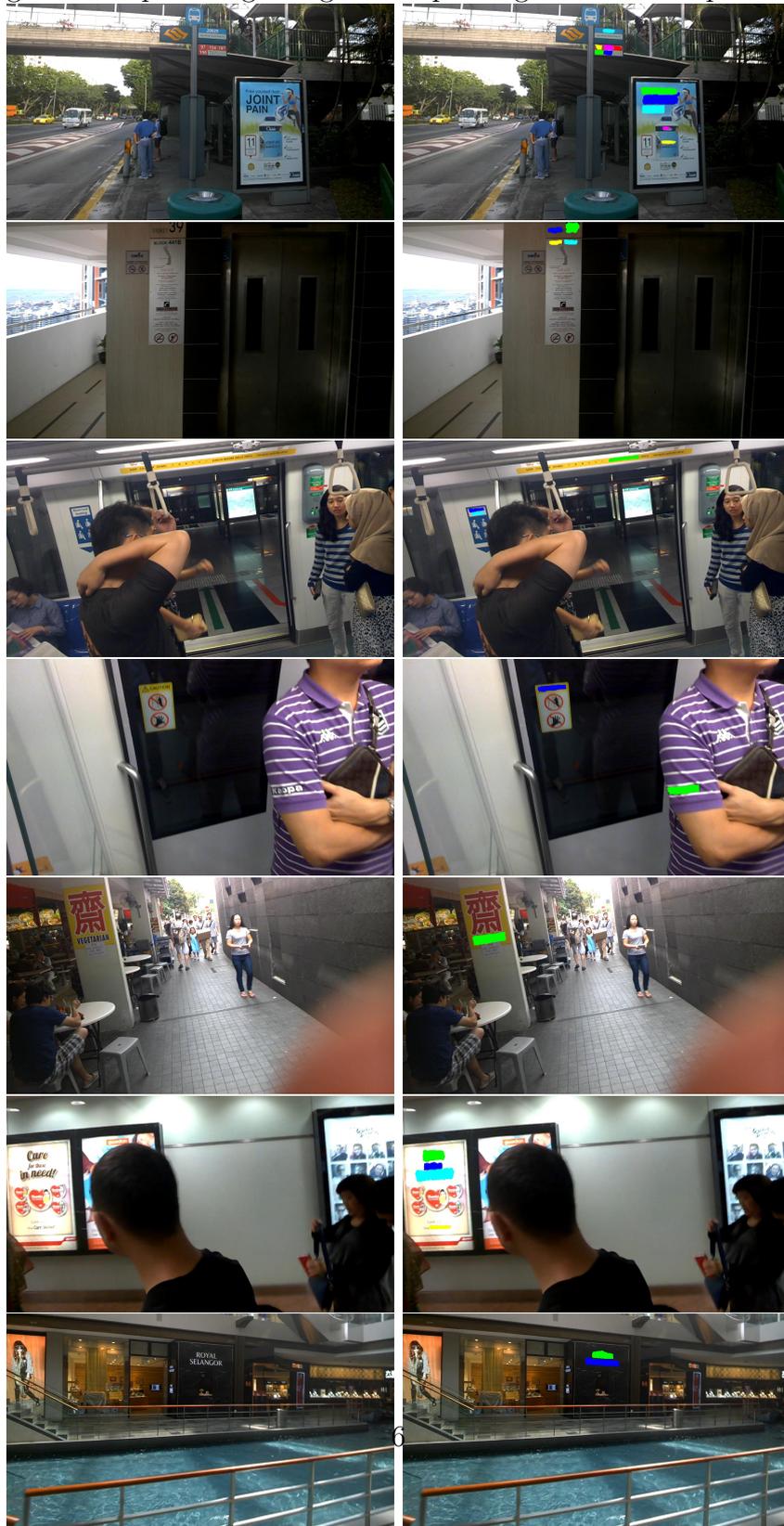
Figure 8: More results of the final segmentation map. Better to look when zooming. Left: Input image. Right: Output Segmentation map
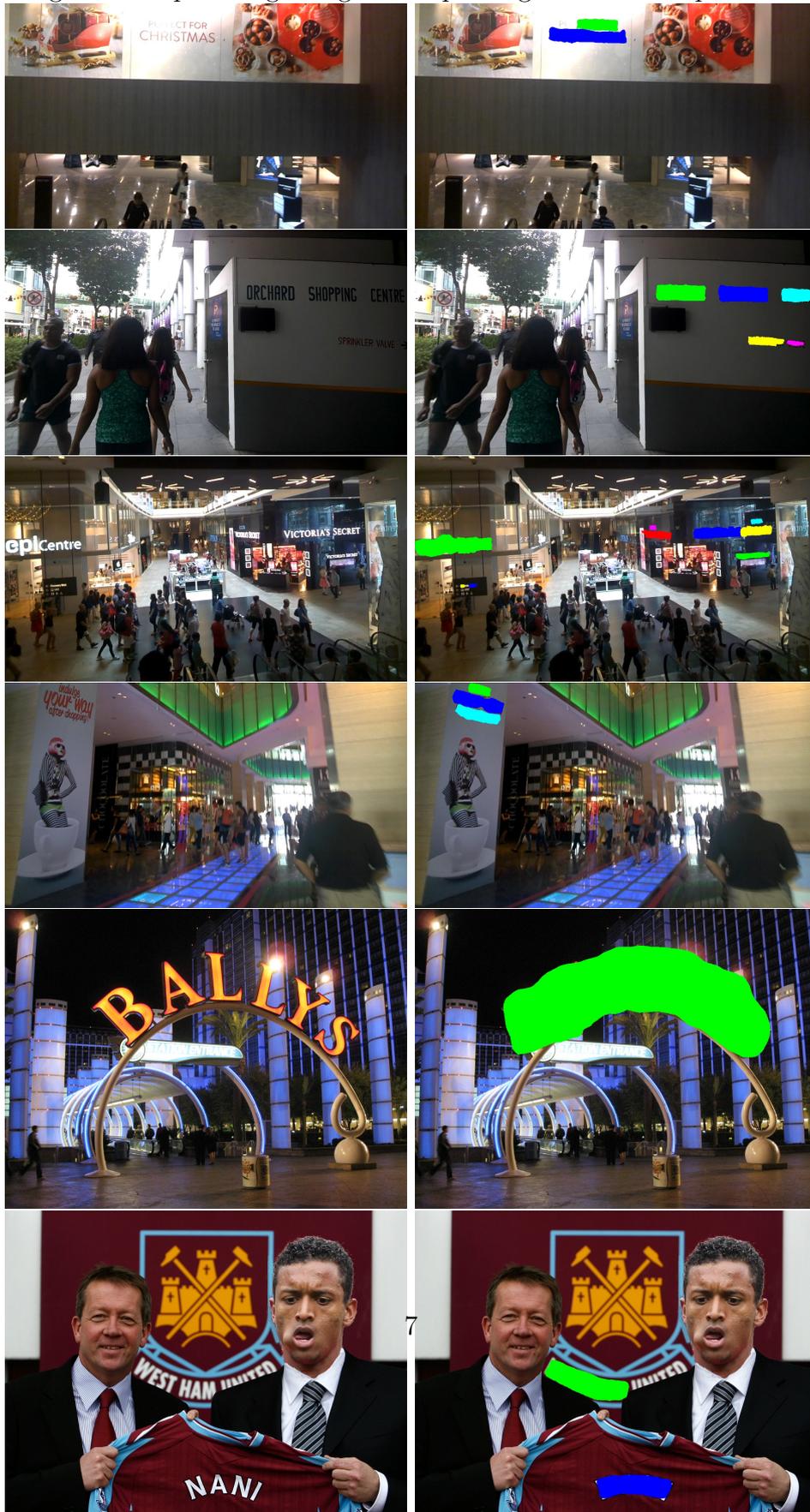
Figure 9: More results of the final segmentation map. Better to look when zooming. Left: Input image. Right: Output Segmentation map