SketchyScene Supplementary Material

1 Examples of Augmented Data

As explained in Section 3 of our main paper, we augmented the dataset by replacing the object sketch with the rest components from the same category based on a given scene sketch template. As shown in Fig. 1, given the scene sketch templates in the second column, we further generated another 20 sketches and selected 3 among them. Column 3 to 5 show the newly synthesised scene sketches.
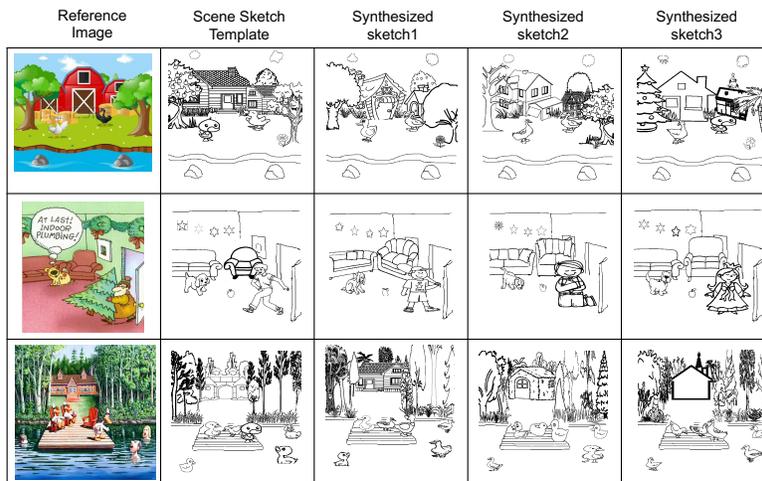
Fig. 1. Examples of augmented scene sketches based on the scene sketch templates.

2 Segmentation Results

Figure 2 shows some good examples of our segmentation results. And Fig. 3 shows some failure examples. As explained in Section 4.2 of the main paper, the failure cases are mainly caused by two reasons: (1) the intra-class variation is large while sketch itself is significantly deformed (Fig. 2(a)) and (2) occlusions between different object instances or instances being spatially too close (Fig. 2(b)). The problem region(s) are highlighted by the red box.

Fig. 2. Examples of good segmentation results.

3 Effect of Background

As discussed in the main paper (Section 4), the large blank areas in scene sketches should be treated differently. We propose to ignore the background class during model training. In this section, we provide more visualisations to show how the background affect the segmentation in scene sketches (as shown in Fig. 4). Specifically, based on the Figure 8 of the main paper, we add **withBG(train)_w/o CRF** (the 2nd column) to show the segmentation results without applying dense-CRF while considering the background class during model training. In the visualisations of this column, it is clear to see that the contours of some objects are discontinuous, like the "dog" in the first example and the "chicken" in the second example, this is because some pixels are misclassified as the "background".
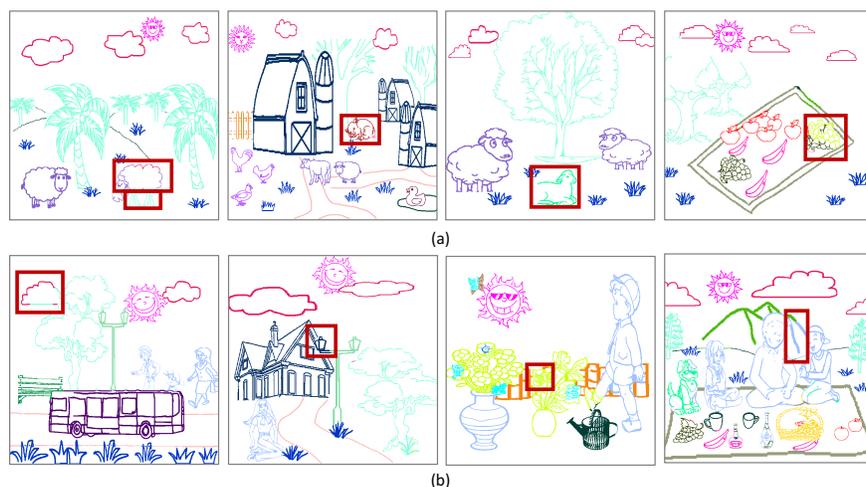
Fig. 3. Examples of failure cases. Top row: failure examples result from the problem of large intra-class variation. Bottom row: failure examples are caused by the occlusions. We highlight the problem region(s) using the red box. (a) 1st: part of the "sheep" body is classified as "cloud" while its legs are recognised as the "tree"; 2nd: the "pig" is classified as a "rabbit"; 3rd: the "sheep" is recognised as a "dog"; 4th: "grapes" are classified as "flowers". (b) The "cloud", "street lamp", "butterfly" and "mountain" are segmented as part of the objects they are spatially close to, i.e., "tree", "house", "flower" and "people".
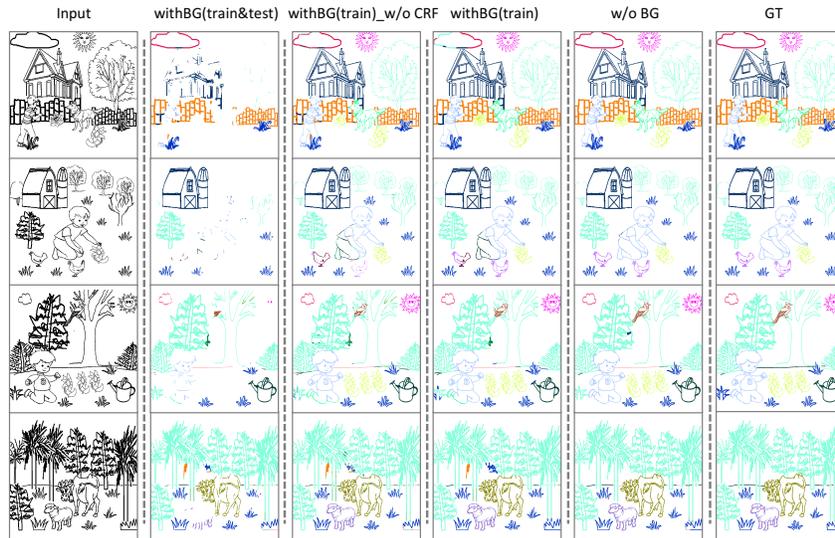
Fig. 4. Effect of the background. **withBG (train&test)**: considering the background during training the deep model and applying denseCRF for refinement; **withBG(train)_w/o CRF**: considering the background during training but without applying denseCRF for refinement; **withBG (train)**: only consider the background during training but ignore this class for refinement; **w/o BG**: our proposed method that ignore the background class during model training.