Supplementary Material

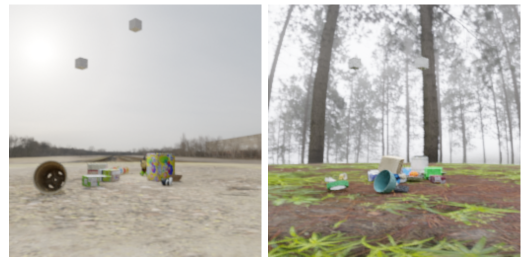
1. Overview

In this supplementary material we include a few more details that we omitted in the original paper.

- In Sec. 2 we talk about the pipeline used to curate the Kubric Change dataset along with some visualisations showing the change in camera positions for "before" and "after" images.
- In Sec. 3 we show our method's predictions on projective transformations.
- In Sec. 4 we show more qualitative examples for each of the test datasets.

2. More on Kubric-Change

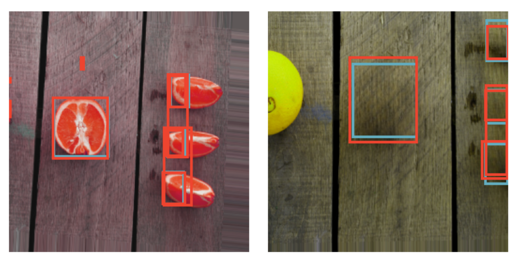
In this section we briefly describe the pipeline used to acquire the Kubric-Change dataset. We build upon the movi_d script provided by Kubric authors, which selects $n \in [10, 20]$ random objects out of $1000+$ assets and spawns them into a random scene at random locations bounded by [(-7, -7, 0), (7, 7, 10)]. It then runs a physics simulator for 100 frames for the objects to fall and settle. We then spawn the camera randomly in a cuboid bounded by [(-5,-5,12), (5,5,18)] and take a picture of the scene. We then remove the "most visible" object and re-spawn the camera randomly in the cuboid at maximum distance of 7 and maximum rotation of $\frac{\pi}{6}$ from its previous pose. We then take another picture. This process is repeated multiple times to collect a large enough dataset. Fig. 1 shows two examples of before and after camera configurations to help put things into perspective.



(a) (b)

Figure 1: The cubes "floating" in air represent the before and after camera positions.

In addition, we make sure that the changed annotations are only over the regions visible in both the images. Fig. 2 illustrates this. Even though the orange slices have disappeared, the ground truth annotations (blue boxes) are only over the regions which are visible in both the images.

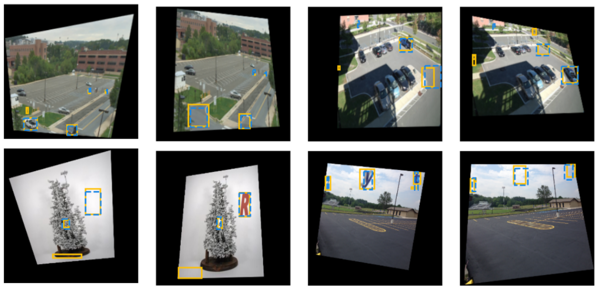


(a) (b)

Figure 2: Ground truth annotations in blue. The right half of the three (vertically arranged) orange slices in (a) is not visible in (b) so we cannot be sure whether it has changed or not. Consequently, only the left half is annotated in the ground truth.

3. Affine to Projective Generalisation

While it was not the focus of this work, we show that our model is also able to generalise to projective transformations while having only been trained using affine transformations. We have no doubt that an explicit training on projective transformations will result in even better predictions.



(a)

Figure 3: Predictions of our model on images related by a projective transformation.

4. More dataset examples

In this section we show more examples for the four test sets presented in this paper. The first two columns show Image 1 and Image 2 that are fed into the model. The last two columns show the top-5 predicted bounding boxes (in yellow, solid), suppressing the ones with significant overlap, and the ground truth (in blue, dashed). Please find the figures below.

Figure 4: COCO-Inpainted

Figure 5: Kubric-Change

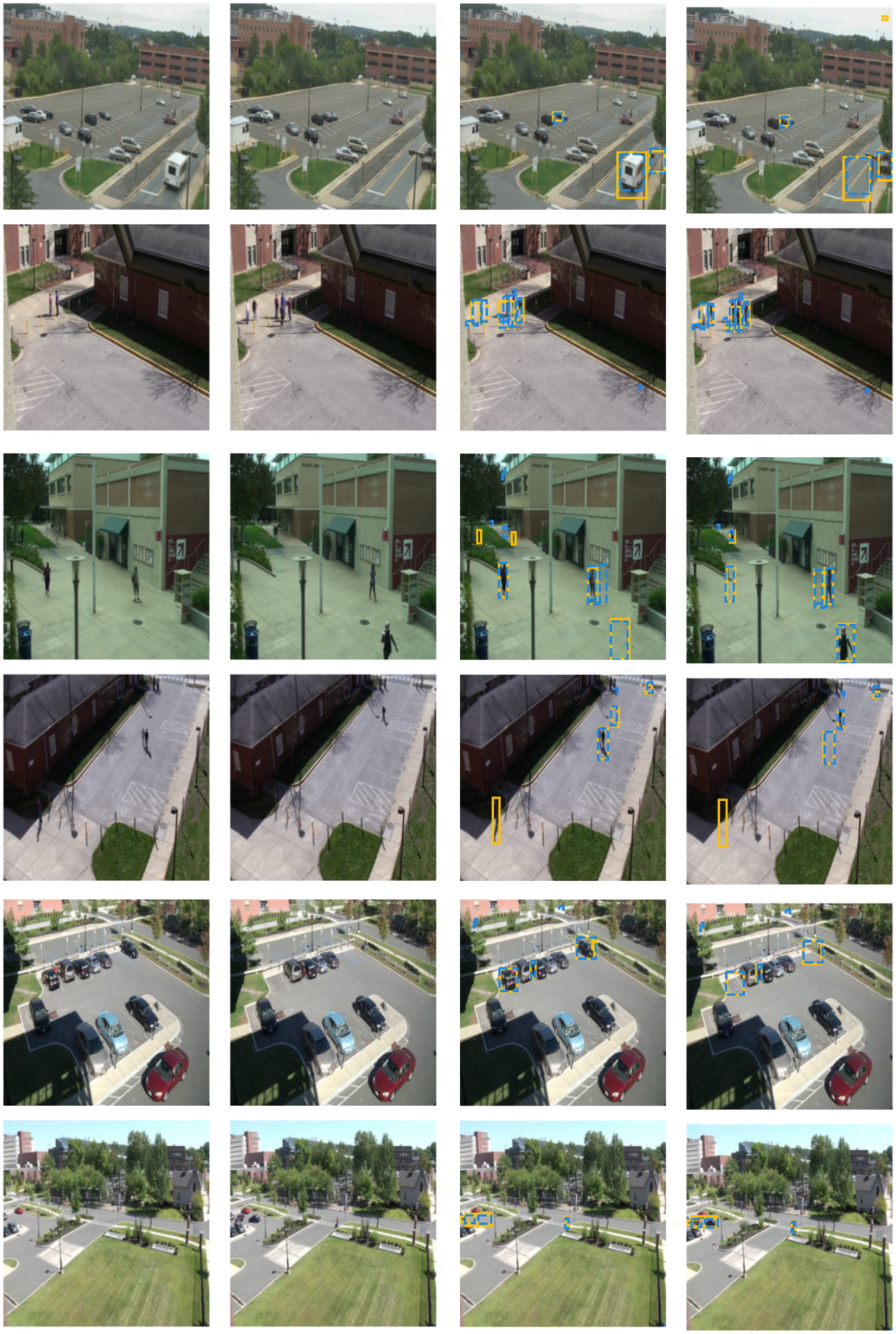


Figure 6: VIRAT-STD

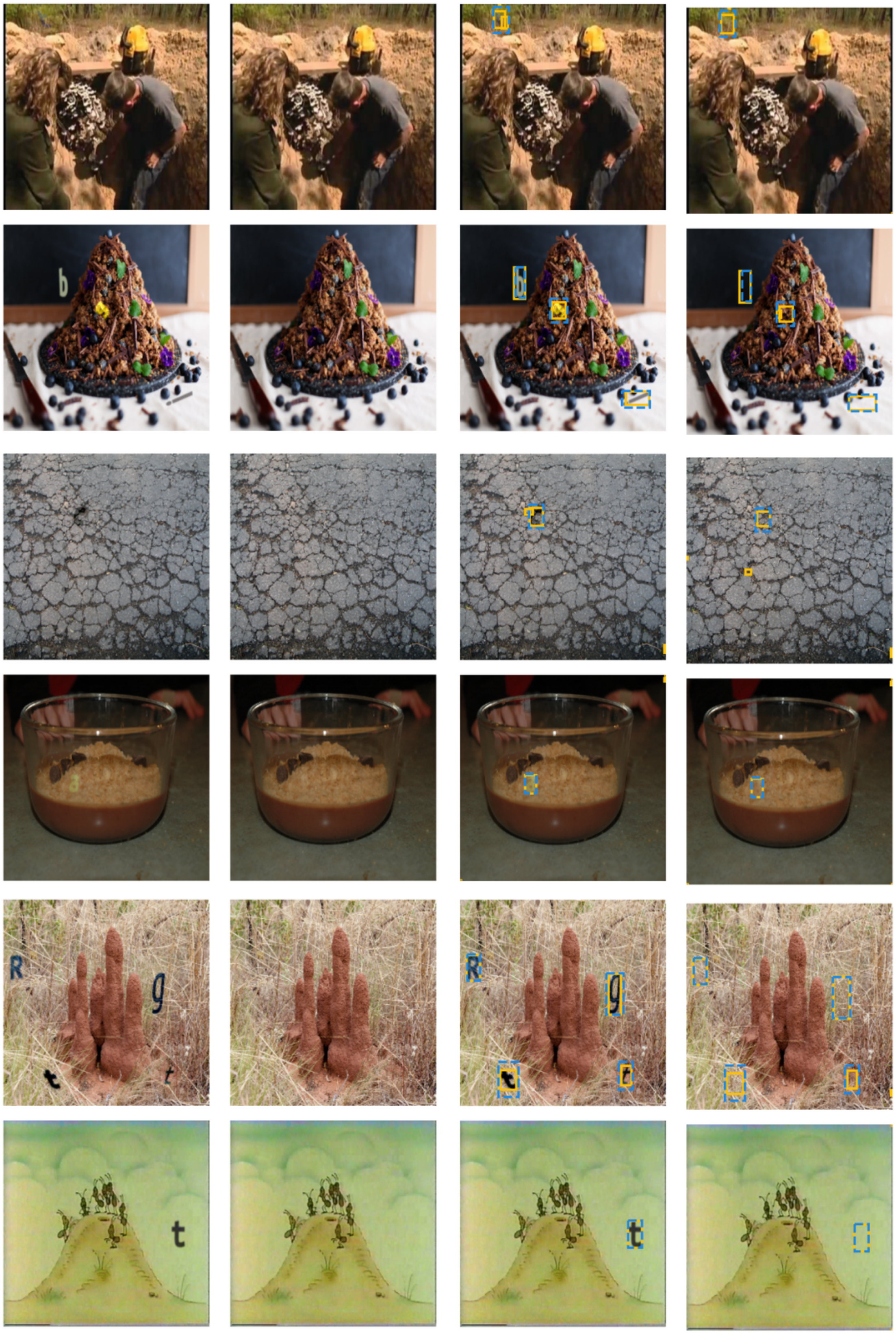


Figure 7: Synthtext-Change