

# Supplementary

## 1. Additional Temporal Super-Resolution Results

In the Fig. 1, 2, and 3, additional results of temporal super-resolution are presented. There are three rows for each sequence – the: (i) baseline results, (ii) results of the proposed method, and (iii) ground truth. Alpha masks are also shown for each sequence in the right half of the figure.

## 2. Selected Videos

We enriched our submission by incorporating 7 videos that illustrate the results achieved by both the baseline and our proposed method, while also encompassing the ground truth sequences and input images. Each video was created from 8-frame-long sequences, with every sequence reconstructed from a single input image.

Since the direction of time in each reconstructed 8-frame segment is ambiguous, we had to determine it manually, which was a very time consuming task. Therefore, we present only 7 videos from the test datasets.

In each video, there are four columns or rows, depending on the orientation of the input image. Each column or row is labeled by one of the following labels: “input image”, “baseline”, “proposed”, “high-speed CAM GT”. We believe that the labels are self-explanatory.

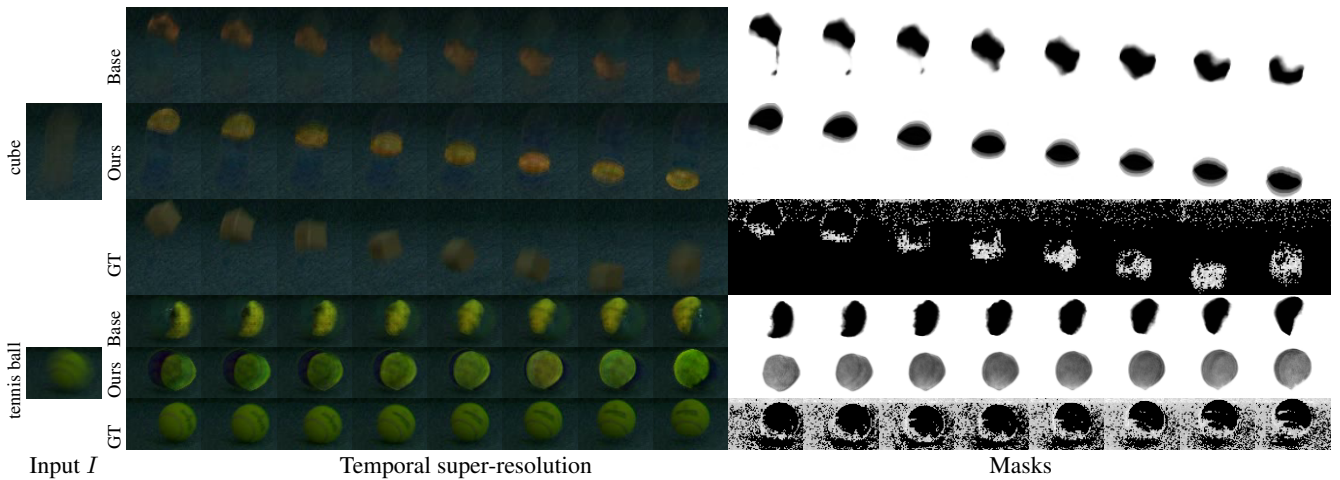


Figure 1. **Temporal super-resolution, by a factor of 8, on selected sequences from “TbD” test dataset.** The SI-DeFMO baseline method (Sec. 3) compared with the ground truth of high-speed footage (GT). Ground-truth masks were computed as a difference image between the GT sub-frames and the background. The proposed SI-DDPM-FMO and baseline method generates all outputs just from a single input image  $I$  on the left.

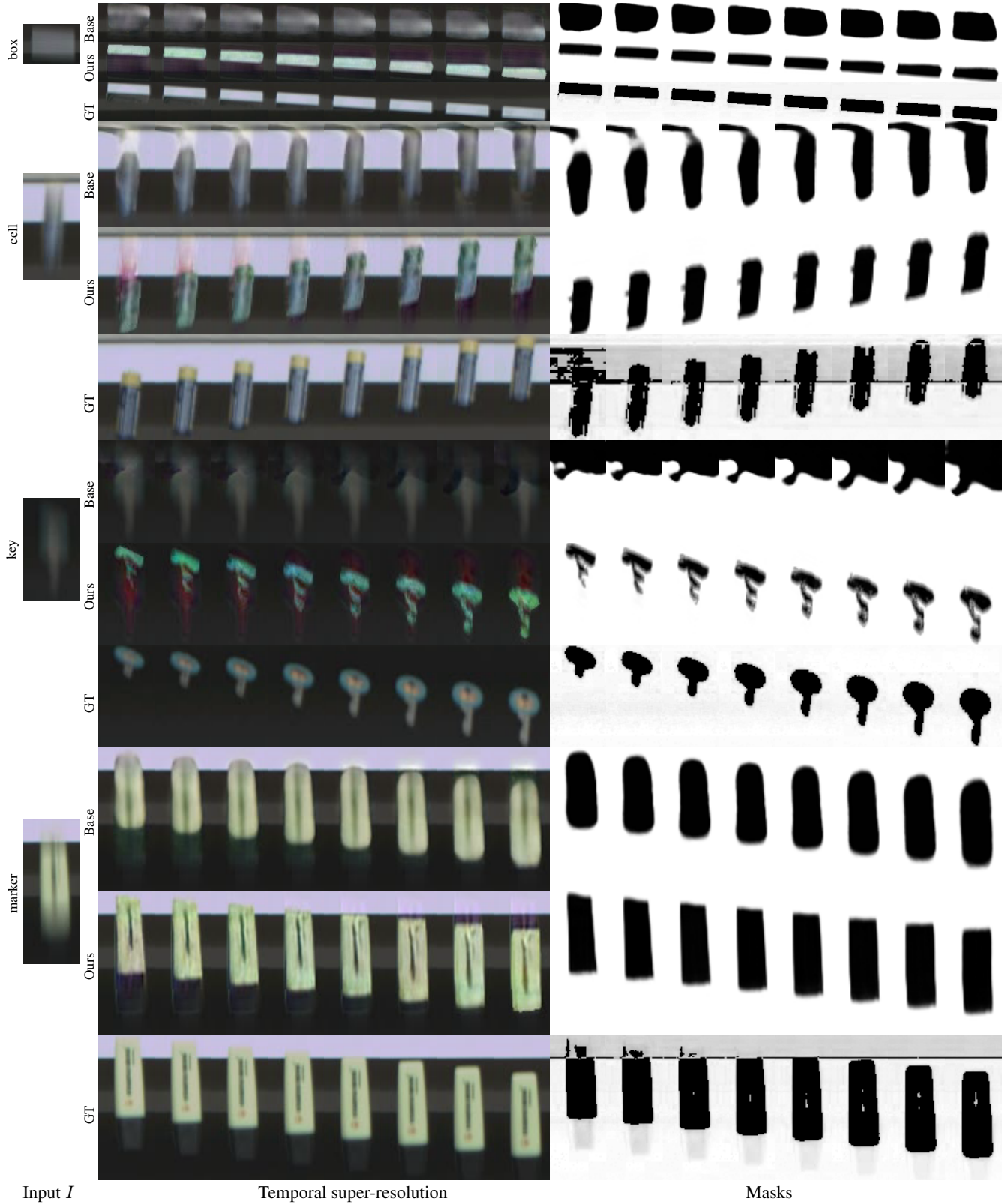


Figure 2. **Temporal super-resolution**, by a factor of 8, on selected sequences from “falling objects” test dataset. The SI-DeFMO baseline method (Sec. 3) compared with the ground truth of high-speed footage (GT). Ground-truth masks were computed as a difference image between the GT sub-frames and the background. The proposed SI-DDPM-FMO and baseline method generates all outputs just from a single input image  $I$  on the left.

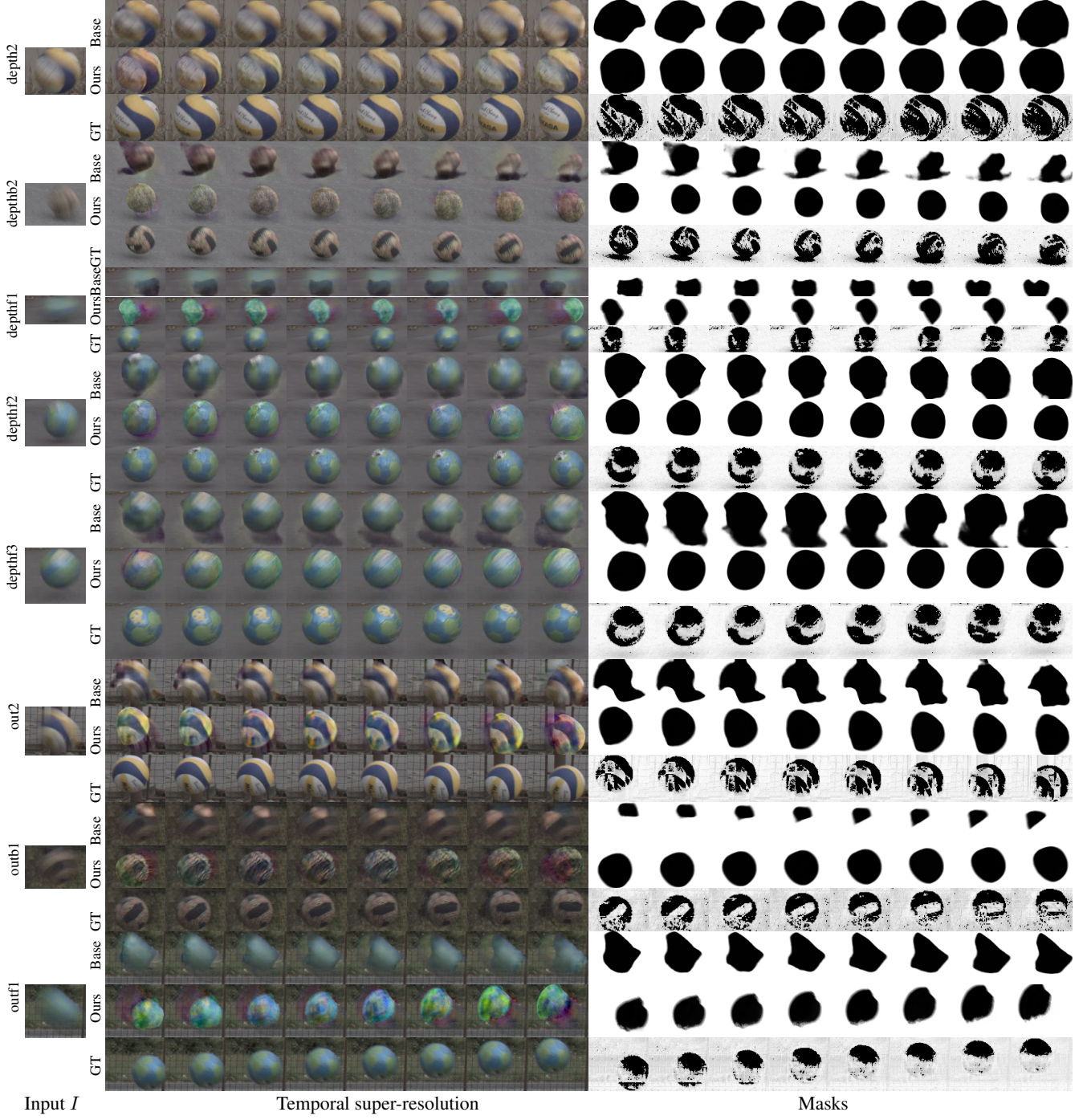


Figure 3. **Temporal super-resolution, by a factor of 8, on selected sequences from “TbD-3D” test dataset.** The SI-DeFMO baseline method (Sec. 3) compared with the ground truth of high-speed footage (GT). Ground-truth masks were computed as a difference image between the GT sub-frames and the background. The proposed SI-DDPM-FMO and baseline method generates all outputs just from a single input image  $I$  on the left.