MEASURING DISTORTION STRENGTHS WITH DEWARPING DIFFUSION MODEL IN ANOMALY DETECTION
SUPPLEMENTARY MATERIAL

1. ANOCLIP DATASET

In this section, we presented the details of the AnoClip dataset used for the experiments in section 3.3. The AnoClip dataset consists of 200 training images and 94 test images. The training set includes only defect-free images, while the test set contains both deformed defective images and defect-free images. Each image contains only a single clip. The camera height is kept fixed, while the positions of clips are roughly aligned by hand, but are not exact. Defects are manually generated. The deformed defective images in the test set have annotations for the maximum deformation. The maximum deformation is defined as the distance to the clip edge belonging to the defective region from its starting position. We apply a feature point matching technique with outlier rejection to more precisely align the objects before measuring the maximum deformation. Note that in this analysis, microscopic effects of plasticity in the physical deformation process (i.e., non-trivial change in molecular patterns) are ignored for simplicity because they are outside the scope of this study. Instead, the "Distortion strength" is defined by the Euclidean distance in the image plane between the corresponding points given by the warping process. Additionally, in our method, we pre-process to extract the foreground [1] from the AnoClip dataset and fill in the background with black color.

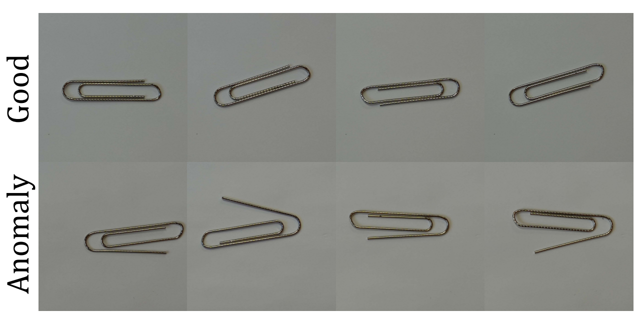

Fig. 7. Examples of the AnoClip data.

2. DETAIL OF THE WARPING OPERATION

The w_{inv} used in Eq. 8 is used to generate x_{t-1} from x_t and $F_{t\to t-1}$. This process is the inverse of the warping process defined by w, which generates x_t from x_0 and $F_{t\to 0}$. To achieve this, we utilize linear splatting based on the warping operation [2]. As for the background regions that were hidden by moving an object part, those regions are filled with the corresponding background part of the estimated normal image $\hat{x}_0$. Here, we assume that these new regions can be represented by $\hat{x}_0$, as is adopted in the concept of Video Frame Interpolation [3],

3. DESIGN OF DEFORMATION F FOR THE ANOCLIP DATASETS

In this section we describe how to create a deformation map $F_{t\to 0}$ for the AnoClip dataset. For the precise way of processing, readers are referred to our code. Below, we sketch the overview of the process. We utilize 2D rotational matrices to express $F_{t\to 0}$, specifically for the AnoClip dataset. We first extract an object mask that traces a clip wire using a conventional method. Then we extract a connected subpart (segment) of the object mask in a random manner, and we define such a subpart as the pre-deformed mask W_0. The position of the end point of W_0 is defined as (a, b). By using a 2D rotational matrix $R(\theta_T)$,

$$R(\theta_T) = \begin{bmatrix} \cos\theta_T & -\sin\theta_T \\ \sin\theta_T & \cos\theta_T \end{bmatrix}, \tag{11}$$

then an image coordinate (x, y) is transformed to:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R(\theta_T) \begin{bmatrix} x-a \\ y-b \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix}. \tag{12}$$

The optical flow at point (x, y) is given by $[x'-x, y'-y]^\top$. Let us denote the image coordinate grids X and Y as vectors containing corresponding x and y coordinates of the image. Then, the global optical flow field is defined as

$$f_{T\to 0} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} (X-a)(\cos\theta_T - 1) - (Y-b)\sin\theta_T \\ (X-a)\sin\theta_T + (Y-b)(\cos\theta_T - 1) \end{bmatrix}. \tag{13}$$

Flow field at an intermediate time t is given by a linear schedule with respect to the rotational angle θ as

$$f_{t\to 0} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} (X-a)(\cos\frac{t}{T}\theta_T - 1) - (Y-b)\sin\frac{t}{T}\theta_T \\ (X-a)\sin\frac{t}{T}\theta_T + (Y-b)(\cos\frac{t}{T}\theta_T - 1) \end{bmatrix}. \tag{14}$$

The object mask at time t is then given by the warping of the initial object mask W_0 with the flow $f_{t\to 0}$ as

$$W_t = w(W_0, f_{t\to 0}). \tag{15}$$

In the experiments with the AnoClip dataset, 2D rotation and the locally smooth deformation are combined within the object mask at time t as

$$F_{t\to 0} = W_t \odot \left(f_{t\to 0} + \frac{t}{T} f_s \right), \tag{16}$$

where f_s is the smooth deformation strength maps defined by the simplex noise in Eq. (3). The newly exposed regions after this warping are filled with a black background color.

4. MORE QUALITATIVE RESULTS.

In this section, we provide more qualitative results for the experiments in section 3. Examples are shown in Fig. 8. The first column shows the anomaly test samples and the second column shows the corresponding ground truth, while the third and fourth columns respectively show the reconstructed images and anomaly-score heat maps produced by the proposed method. DiffuDewarp reproduces only the defective areas and can accurately detect the defects most of the times. Furthermore, the simple modifications explained in the main text show that DiffuDewarp-c can identify not only deformation defects, but also discoloration defects.

5. REFERENCES

[1] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool, "Highly accurate dichotomous image segmentation," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 38–56.

[2] Simon Niklaus and Feng Liu, "Softmax splatting for video frame interpolation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5437–5446.

[3] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz, "Super slomo: High quality estimation of multiple intermediate frames for video interpolation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9000–9008.

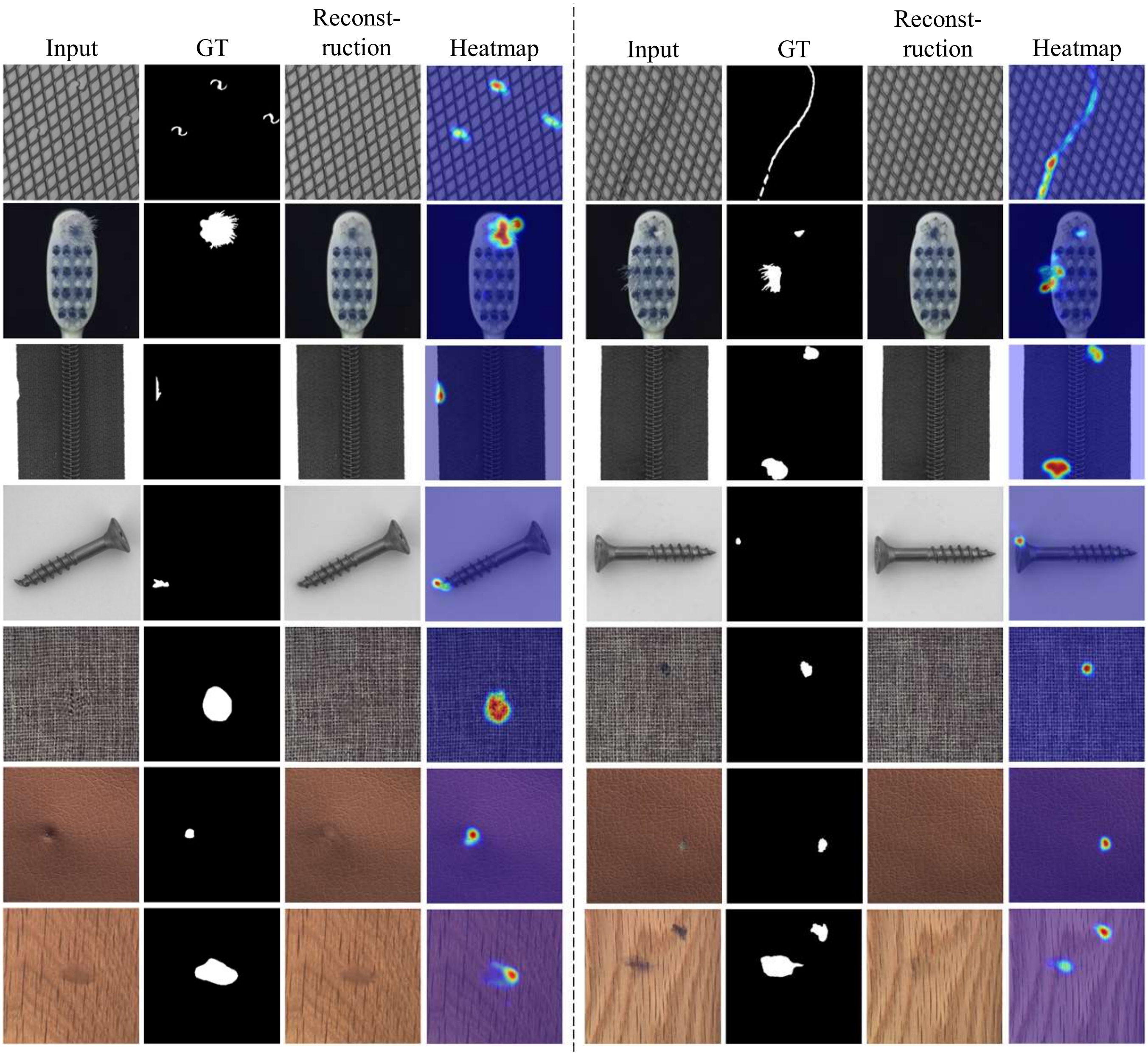

Fig. 8. More qualitative examples produced by the proposed method. The top three rows show the results from Gr. 1 with DiffuDewarp alone, and the bottom three show the results from Gr. 2 with DiffuDewarp-c, the extended method that uses both distortion criterion and reconstruction-error criterion.