MEASURING DISTORTION STRENGTH WITH DEWARPING DIFFUSION MODELS IN ANOMALY DETECTION

Akira Uchida[1], Satoshi Ikehata[1,3], Yuichi Yoshida[2], Ikuro Sato[1,2]

[1]Institute of Science Tokyo, Tokyo, Japan
[2]Denso IT Laboratory, Tokyo, Japan
[3]National Institute of Informatics, Tokyo, Japan

ABSTRACT

Surface anomaly detection aims to localize abnormal regions in images. A representative approach is the reconstruction-based method, which detect defects via reconstruction errors using generative models trained on normal images. However, these methods cannot directly estimate local distortion levels, which are commonly used for products made of metal or resin. To address this issue, we propose DiffuDewarp, a novel method that directly estimates local distortions. Our approach defines a pseudo-deformation defect generation process as a new diffusion process based on localized warping. Experiments on the MVTec dataset demonstrate that our method outperforms state-of-the-art techniques in categories where local deformations are the main cause of defects. We will release our code upon acceptance.

***Index Terms*—** Anomaly Detection, Visual Inspection, Diffusion model

1. INTRODUCTION

Surface anomaly detection identifies unusual regions in product images. Due to the rarity of defects, unsupervised methods that do not require anormalous data are widely used [1, 2, 3, 4]. These methods fall into two main categories: embedding-based [3, 4, 5, 6] and reconstruction-based [1, 2, 7, 8]. Embedding-based approaches use pre-trained networks to extract image features, effectively separating anomalous from normal patterns. Reconstruction-based approaches detect anomalies by comparing pixel-level errors between test images and reconstructions generated by autoencoders [8, 9], GANs [8, 10, 11], or diffusion models [12, 13, 14, 15, 16]. While embedding-based methods are efficient for image-level detection, they often miss fine-grained details at the pixel level. Conversely, reconstruction-based methods excel at pixel-level localization at the expense of more computational cost. However, reconstruction-based approaches usually do not explicitly evaluate the severity of detected anomalies while they can locate anomalies. This gap is especially critical for distortion-based defects in metallic, resinous, or fiber components, where local deformations must be assessed against criteria like flatness or straightness [18, 19]. Reconstruction errors alone are insufficient to quantify the geometric or structural deviations necessary to determine whether a distortion exceeds acceptable tolerances. Addressing this limitation requires more advanced unsupervised techniques that both detect and quantify anomalies to meet industrial standards.

[1]Please refer to supplementary material for more details: https://sigport.org/documents/measuring-distortion-strength-dewarping-diffusion-models-anomaly-detection

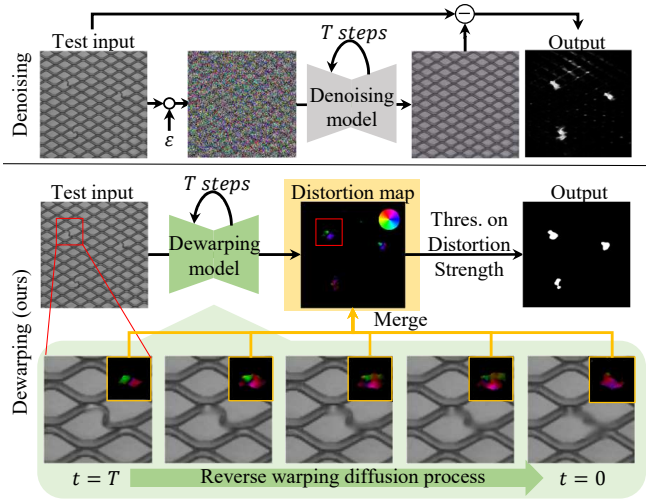


Fig. 1: Comparison ofanomaly detection methods based on denoising (*e.g.*, [12]) and dewarping (ours) diffusion models. **Top**: denoising-based methods (*e.g.*, [17]); **Bottom**: our DiffuDewarp. Denoising-based methods reconstruct normal images via the denoising process. In DiffuDewarp, a sequence of dewarping operations is applied to test images to compute the deformation fields, which are accumulated into a cumulative deformation field. Its strength (*i.e.*, distortion strength map) is used to identify anomalous pixels. The color of the distortion map represents the direction of deformation.

In this paper, we introduce *DiffuDewarp*, a reconstruction-based unsupervised anomaly detection framework that simultaneously predicts the degree of spatial distortion relative to normal images. Unlike conventional methods [12, 13, 14, 15, 16] that rely on denoising diffusion models [20] (see Fig. 1-(top)), our approach redefines the diffusion process by modeling image degradation as a series of localized warping operations. Specifically, our method employs a forward diffusion process that progressively applies spatial deformations to normal images—effectively mimicking defect formation—and a reverse diffusion process in which a dedicated network simultaneously predicts both the cumulative deformation field and the normal image from an anomalous image at a specific timestep, thereby generating interpretable pixel-level distortion strength maps that are available for localized anomaly detection. Our method is further enhanced in an extended variant, *DiffuDewarp-c*, which integrates an anomaly simulator [1] with our forward warping process for generating training data to handle non-deformative anomalies such as dis-

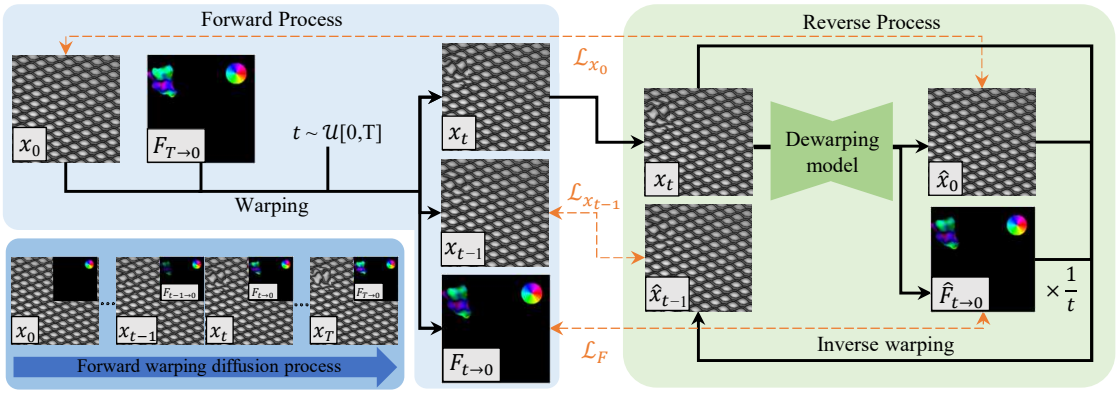


Fig. 2: The overview of DiffuDewarp. **left**: Given a normal image x_0, the forward process progressively deforms x_0 over T steps using a deformation field $F_{t\to0}$, where each intermediate image x_t is generated via the warping function $w(x_0, F_{t\to0})$. **right**: In the reverse process, the dewarping model M_θ takes an anomalous image x_t and time t as input and predicts both the cumulative deformation field $\hat{F}_{t\to0}$ and the reconstructed normal image $\hat{x}_0$. Reconstruction losses are computed between x_0 and $\hat{x}_0$, as well as between $F_{t\to0}$ and $\hat{F}_{t\to0}$. Additionally, an inverse warping is applied to iteratively reconstruct $\hat{x}_{t-1}$ using the predicted deformation field and the loss is computed between x_{t-1} and $\hat{x}_{t-1}$, enabling a progressive reversal of the deformation process. Anomaly detection is performed by thresholding the intensity of the reconstructed $F^*_{T\to0}$ obtained from recursively predicted $\hat{F}_{t\to t-1}$ $(t = T, \ldots, 1)$ in Eq. 7. See Section 2 for details.

coloration. As illustrated in Fig. 1-(bottom), DiffuDewarp recovers a near-normal image solely through warping operations.

Our experiments on MVTec-AD [21] demonstrate that DiffuDewarp outperforms state-of-the-art methods—especially for defect categories driven primarily by local distortions (e.g., Grid, Toothbrush, and Zipper)—while also achieving competitive performance in cases where defects arise from both local distortions and corruptions of textual or color patterns (e.g., Carpet, Leather, Screw, and Wood). In addition, we introduce the AnoClip datasets, which is constructed to evaluate metal distortions in industrial components by providing precise ground-truth distortion values. Our results on AnoClip show that the anomaly scores generated by DiffuDewarp closely align with these values.

2. METHOD

Unsupervised anomaly detection aims to determine whether a given image or pixel deviates from the learned patterns of normal images. The training dataset consists exclusively of normal images: $\mathcal{D} = \{x_0^p\}_{p=1}^N$, $x_0^p \in \mathbb{R}^{h\times w\times 3}$ where N is the number of samples, and h and w denote the height and width, respectively. A test image $x_T \in \mathbb{R}^{h\times w\times 3}$ is then evaluated to determine whether it is normal or anomalous on a pixel-wise basis. We train separate models to detect anomalies for different object classes, similar to [17, 22, 23].

Our method, *DiffuDewarp*, builds upon a reconstruction-based approach that leverages diffusion models [12, 13, 14, 15, 16]. However, it fundamentally differs in how the diffusion process is defined. Instead of employing conventional denoising diffusion models [20], we formulate diffusion as a local warping process, where the image undergoes continuous degradation through localized deformations. This novel formulation offers two key advantages: (1) it constrains the degradation process spatially at each time step, ensuring localized control over distortions, and (2) it enables simultaneous prediction of local distortion strengh, effectively approximating the anomaly formation process by analyzing the temporal evolution of warping magnitudes. These properties make DiffuDewarp uniquely suited for precise and interpretable anomaly detection especially for distortion-based defect in metallic, resinous, fiber and so on.

Forward warping diffusion process: We define a forward process that progressively deforms a normal image x_0 (at $t = 0$) into a distorted image x_T over T steps. At each timestep t, x_0 is warped via a warping function w using a deformation field $F_{t\to0} \in \mathbb{R}^{h\times w\times 2}$.[2]

$$x_t = w(x_0, F_{t\to0}). \tag{1}$$

The output y of $w(x, F)$ is defined as $y_{u,v} = x_{u+F(u,v),\, v+F(u,v)}$.

The deformation field at time T is constructed as follows. First, a binary mask W_T is generated from Perlin noise P [24] with threshold γ:

$$W_T = \begin{cases} 1, & \text{if } P - \gamma > 0, \\ 0, & \text{if } P - \gamma \le 0. \end{cases} \tag{2}$$

Next, two-dimensional smooth deformation strength maps for u and v are sampled independently using simplex noise [25]:

$$f_T = \begin{bmatrix} \epsilon_u \\ \epsilon_v \end{bmatrix}, \quad \epsilon_u,\ \epsilon_v \sim \text{Simplex}. \tag{3}$$

With a maximum deformation strength in pixel α, $F_{T\to0}$ is defined as

$$F_{T\to0} = W_T \odot (\alpha f_T), \tag{4}$$

where $\odot$ denotes element-wise multiplication.

For intermediate steps $t = 1, \ldots, T-1$, the deformation field is interpolated linearly so that x_0 is smoothly deformed during the forward process:

$$F_{t\to0} = \frac{t}{T} F_{T\to0}. \tag{5}$$

Thus, the backward warping function systematically applies the accumulated local deformations to transform x_0 into the sequence $\{x_t\}_{t=0}^T$. An example of this process is shown in Fig. 2-(left). Note that the definition of $F_{T\to0}$ here is effective for most deformative anomalies in the MVTech [21], as will be demonstrated in Sec. 3.

[2]Image warping is implemented with `grid_sample()` in PyTorch.

On the other hand, by changing the definition of $F_{T\to 0}$, it is possible to define a forward process based on a different deformation patterns. In Sec. 3, the effectiveness of models trained using different deformation fields is evaluated using our AnoClip datasets.

Reverse warping diffusion process: In our reverse process, the dewarping model $M_\theta(x_t, t)$ directly predicts the cumulative deformation field $\hat{F}_{t\to 0}$ and the corresponding reconstruction of the normal image $\hat{x}_0$. We then invert the forward warping process—analogous to how denoising diffusion models predict and scale the noise at time t—to generate the dewarped image $\hat{x}_{t-1}$ at time $t-1$ as follows:

$$\hat{x}_0, \hat{F}_{t\to 0} = M_\theta(x_t, t), \tag{6}$$

$$\hat{F}_{t\to t-1} = \frac{1}{t}\,\hat{F}_{t\to 0}, \tag{7}$$

$$\hat{x}_{t-1} = w_{\text{inv}}(x_t, \hat{x}_0, \hat{F}_{t\to t-1}), \tag{8}$$

where $w_{\text{inv}}(\cdot)$ represents the inverse warping operation detailed in Section 2 of the supplementary material[1]. This dewarping step, inspired by the denoising process, progressively reconstructs the normal image by inverting the forward warping transformation at each timestep until x_0, as illustrated in Fig. 2-(right). Unlike conventional denoising models, $M_\theta(x_t, t)$ predicts x_0 to (i) inpaint undegraded holes arising during dewarping and (ii) improve training via explicit x_0 supervision [20].

During training, $M_\theta(x_t, t)$ is optimized by minimizing a summation of three ℓ_1-based losses: the deformation field reconstruction loss (*i.e.*, $\mathcal{L}_F \triangleq \|\hat{F}_{t\to 0} - F_{t\to 0}\|_1$), the previous-time image reconstruction loss (*i.e.*, $\mathcal{L}_{x_{t-1}} \triangleq \|\hat{x}_{t-1} - x_{t-1}\|_1$), and the x_0 prediction loss (*i.e.*, $\mathcal{L}_{x_0} \triangleq \|\hat{x}_0 - x_0\|_1$) over a batch of input images $x_0^p \in \mathcal{D}$ and their corresponding x_t^p at random timesteps $t \sim \mathcal{U}(1, T)$, generated according to Eq. (1). More details are provided in Sec. 3.

Anomaly detection via DiffuDewarp: In most reconstruction-based methods [12, 13, 14, 15, 16], anomaly pixels are identified by comparing an anomalous test image with its reconstructed normal counterpart—often using a segmentation network to reduce artifacts. In contrast, we use the predicted deformation fields, $\hat{F}_{t\to t-1|t=T,\ldots,0}$, to approximate $F_{T\to 0}$ and detect anomalies based on pixelwise deformation strength. Recall that $F_{T\to 0}$ represents the net deformation between time T and time 0, not merely the sum of one-step deformations, $F_{t\to t-1}$. For example, suppose $T = 4$ and each one-step deformation shifts a pixel at $(0, 0)$ by +1 pixel in the x-direction. In that case, the pixel's final position would be $(4, 0)$, so the cumulative deformation field should indicate a correction of $(-4, 0)$ to restore the pixel to its original location, rather than a constant 1-pixel shift between $(0, 0)$ and $(0, 4)$. Reflecting this idea, given $\hat{F}_{t\to t-1|t=T,\ldots,0}$ at test time, we recursively compute the approximation of $F_{T\to 0}$ (*i.e.*, $F^*_{T\to 0}$), starting with $F^*_{1\to 0}$, then $F^*_{2\to 0}$, and so on, until $\hat{F}^*_{T\to 0}$ is obtained as follows:

$$\begin{aligned} F^*_{t\to 0} \leftarrow \hat{F}_{t\to t-1} + w\big(F^*_{t-1\to 0}, \hat{F}_{t\to t-1}\big) \odot W_{t\to t-1}, \\ \forall t = 1, 2, \ldots, T. \end{aligned} \tag{9}$$

Here, $F^*_{t-1\to 0|t=1}$ is set to zero. $W_{t\to t-1}$ is a binary map that stores one where the norm of $\hat{F}_{t\to t-1}$ is positive, restricting the warping of $\hat{F}_{t-1\to 0}$ to regions with nonzero flow in $\hat{F}_{t\to t-1}$. We then compute the magnitude of $F^*_{T\to 0}$ to create a distortion strength map, which is thresholded to detect anomalies. Thanks to the localized reconstruction framework, our method performs well without a post-hoc segmentation network, unlike existing approaches (e.g., [12]).

Handling non-distortive anomalies: DiffuDewarp primarily addresses anomalies caused by spatial distortion. However, many anomalies arise from other factors (*e.g.*, discoloration). While cases without distortion are out of scope, some images may exhibit both distortion and other anomalies (*e.g.*, chemical discoloration). To improve robustness in such cases, we propose extending our training data by synthesizing distortions using the forward process on *simulated anomalous images*, following DRÆM [1], instead of normal images—without modifying the training objectives. In addition, to detect anomalies that do not involve distortion, we sum the x_0 reconstruction error (*i.e.*, $\mathcal{L}_{x_0}$) and the distortion strength for the anomaly score. This variant, termed *DiffuWarping-c*, is evaluated separately from the original DiffuDewarp.

3. RESULTS

We conducted extensive experiments to evaluate DiffuDewarp and DiffuDewarp-C on the MVTec [21] dataset and the AnoClip dataset, which we will create to numerically evaluate the performance of distortion strength prediction.

Inplementation Details: DiffuDewarp and DiffuDewarp-C are trained on MVTec [21] training set with synthesized deformations (and simulated anomalies in the case of DiffuDewarp-C) based on our forward warping diffusion process. The model is trained with a batch size of 16 for up to 4000 epochs using a single NVIDIA H100 SXM5 94GB HBM2e GPU. The number of diffusion steps is set to $T = 32$. All images in the training dataset are resized to 256×256. The dewarping model (M_θ) employs a U-Net architecture identical to the one used as the denoising U-Net in DiffusionAD [17]. Since deformation fields are assumed to be smooth, a Gaussian kernel with $\sigma = 5$ is applied to reconstructed $F^*_{T\to 0}$ before the computation of distortion strength map. The optimization is performed using the Adam optimizer with an initial learning rate of 10^{-4}. In Eq. 4, α is set to either 10 or 0 with a 50% probability for each. The choice of $\alpha = 0$ is made to prevent the network from over-evaluating the deformation when a normal image is input at test time similar to [17].

Evaluation Dataset: A) We examined Grid, Toothbrush, and Zipper classes (referred to as Gr. 1) in MVTec [21]. In this group, spatial distortion is mostly the main cause of defects; e.g., excessive bending of grid, distortion of woven patterns, and shifts of toothbrush bristles are often observed. B) We examined Screw, Carpet, Leather, and Wood classes (referred to as Gr. 2) in MVTec [21]. Defects in this group typically include discoloration and contamination with other materials on top of local spatial distortion. In the evaluation of this group, we equip DiffuDewarp with a simple extension (similar to [1]) so that it can handle reconstruction error map as well as spatial deformation field for anomaly identification. C) We examined AnoClip datasets, constructed by us, which contains images of clips with corresponding ground-truth distortion information. With this new dataset, we evaluated the accuracy of distortion strengths. Details of the AnoClip datasets is given in Section 1 of supplementary material[1]. Furthermore, the new definition of the deformation field represents an extension to other defects that cannot be represented by the simplex noise-based deformation field. Details of the new definition of the deformation field is given in Section 3 of supplementary material[1].

3.1. Gr. 1: Grid, Toothbrush, and Zipper Classes

To evaluate the effectiveness of anomaly detection based on distortion strength in classes where local distortions are the main cause of

Table 1: Anomaly localization performance with Pixel-AUROC / Image-AUROC (%) on the Group 1 and 2 in MVTec-AD [21]. In each row, First and second are highlighted. "DiffuDewarp" uses distortion strength only, while "DiffuDewarp-c" incorporates both distortion strength and pixel-wise color difference. TransFusion [22]'s Pixel-AUROC is from our reproduction.

	Using Segmentation Networks			Not Using Segmentation Networks				
	DiffAD [23]	TransFusion [22]	DiffusionAD [17]	AnomDiff [13]	DDAD [14]	MDPS [15]	DiffuDewarp (ours)	DiffuDewarp-c (ours)
Grid	99.7/100	99.4/100	99.7/100	99.1/99.7	99.4/100	99.4/100	99.4/100	99.5/100
Toothbrush	99.2/100	97.6/100	98.8/100	98.9/96.9	98.7/100	98.8/100	99.3/100	99.1/100
Zipper	99.0/100	98.5/100	99.2/100	97.6/85.5	98.2/100	98.5/100	99.4/100	99.2/100
Gr. 1 Avg	99.3/100	98.5/100	99.2/100	98.5/94.0	98.8/100	98.9/100	99.4/100	99.3/100
Carpet	98.1/98.3	98.3/99.2	99.0/99.3	98.9/99.9	98.7/99.3	94.4/100	98.4/98.0	99.3/99.5
Leather	99.1/100	98.9/100	99.8/100	99.5/100	99.4/100	99.5/100	99.6/100	99.7/99.9
Screw	99.0/97.2	97.0/97.2	99.0/99.4	99.0/89.9	99.3/99.0	98.9/96.7	91.3/99.4	99.7/98.8
Wood	96.7/100	94.6/99.4	96.7/99.9	94.7/98.1	95.0/100	95.7/99.1	95.0/97.8	97.8/99.6
Gr. 2 Avg	98.2/98.9	97.2/99.0	98.6/99.7	98.0/97.0	98.1/99.6	97.1/99.0	96.1/98.8	99.1/99.5
ALL Avg	98.7/99.4	97.8/99.4	98.9/99.8	98.2/95.7	98.4/99.8	97.9/99.4	97.5/99.3	99.2/99.7

Table 2: AUROC (%) of anomality classification on the AnoClip datasets. DiffusionAD [17] used the sum of reconstruction errors, while our method used the largest magnitude of the distortion map as anomaly scores.

Threshold (pix)	0	20	40	60
DiffuionAD [17]	97.9	99.1	95.0	86.8
DiffuDewarp (ours)	100.0	99.7	100.0	99.4

defects in most cases, we conducted experiments on Gr. 1. The results of anomaly detection and anomaly localization are summarized in Table 1. Our method achieved the highest pixel-level AUROC (99.4%) in terms of average accuracy within Gr. 1. Furthermore, among methods that do not use segmentation networks, our approach demonstrated the highest accuracy across all classes in Gr. 1. It is remarkable that approximate distortion strengths can identify defects at such a high accuracy, while all baseline methods compared here are based on measured reconstruction error. localization results and reconstruction results are depicted in Fig. 3 and Fig. 4. In some cases, the reconstructed images by the diffusion model in DiffusionAD [17], which uses a noise-based diffusion model, contain subtle changes in the normal areas. These can be the potential cause of false positives. To suppress such parts, subsequent segmentation network is applied in some existing methods [17, 22, 23]. Note again that our method does not require a segmentation network. The main difference is that our method does not add artificial noise at test time, unlike other methods, which superimpose noise that can distort input signals and potentially affect subtle changes in normal areas. A qualitative example of the MVTec data set is shown in Fig. 5.

3.2. Gr. 2: Screw, Carpet, Leather, and Wood Classes

This group contains defects caused by spatial distortion and discoloration. With this group, we evaluate DiffuDewarp-c, which is the extended version of DiffuDewarp by adding color shift measure as well as distortion strength in anomaly detection. As shown in Table 1, DiffuDewarp-c achieved comparable pixel-AUROC in average for Group 1, Group 2, and all classes combined, as well as the highest pixel-AUROC for all classes without segmentation. Fig. 3 and Fig. 4 show that both deformation- and color-oriented defects are well transformed to normal ones in our method.

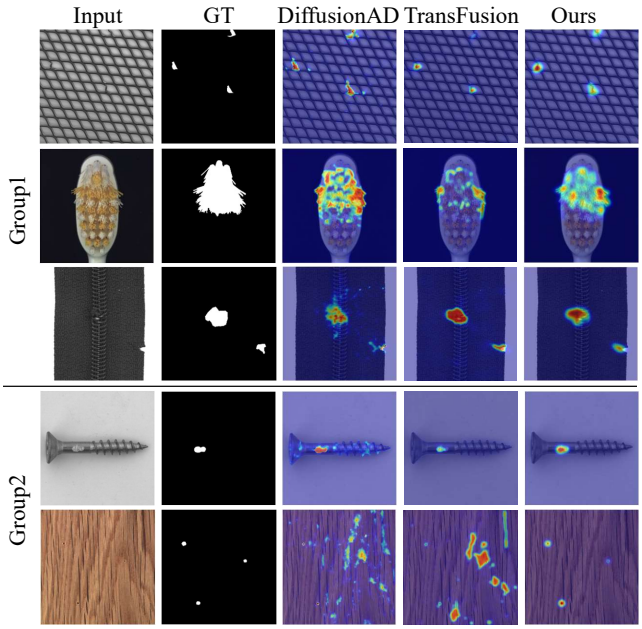


Fig. 3: Visualization of anomaly detection results on MVTec [21]. The third and forth columns show reconstruction error maps produced by the baselines. The last column shows distortion strength maps produced by our DiffuDewarp.

3.3. Evaluating Distortion Strength Prediction on AnoClip

Evaluation on MVTec does not verify whether the predicted distortion strength map reflects actual anomaly severity. To address this, we evaluated DiffuDewarp on the AnoCLIP dataset, which includes ground truth anomaly intensity annotations (see Sec.1 in Supplementary[1]). In AnoCLIP, an external force is applied to a metal clip's wire, inducing a localized deformation around the viewpoint. The distance between the pre- and post-deformation positions at the wire's tip (*i.e.*, maximum deformation) is used as the ground truth. Furthermore, this evaluation demonstrates the applicability of different deformation models. While the MVTec training model used simplex noise in Eq. (3), for AnoCLIP we propose an alternative model that better captures clip anomaly characteristics by defining the distortion strength as a composite of simplex noise and smooth rotation around a point in the foreground object region. Details of

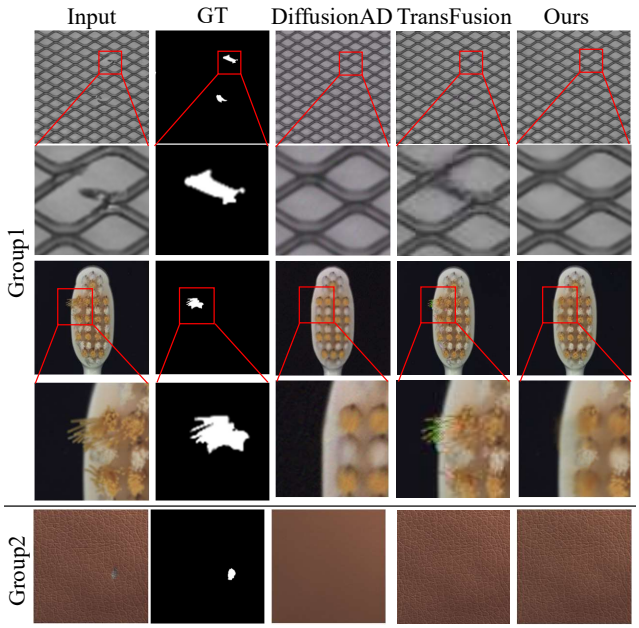


Fig. 4: Visualization of reconstructed images. DiffusionAD [17] tends to generate slightly different patterns in the normal regions. TransFusion [22] leaves anomalous parts but DiffuDewarp can transform abnormal to normal, while leaving normal regions unchanged.

the alternative model are provided in Sec. 3 of the supplementary material[1]. The results of anomaly detection and the predicted distortion map are shown in Fig. 5-(right) and Table 2. This evaluation examines whether classification based on distortion strength is feasible using the anomaly score. The conventional method uses the sum of reconstruction errors, while the proposed method employs the maximum value of the estimated deformation map. Note that the threshold (*e.g.*, 20 px) serves as a criterion for classifying anomalies with a maximum true deformation of 20 pixels or more. As shown in Table 2, DiffuDewarp achieves a 100% Image-AUROC by detecting anomalies via zero thresholding of the distortion strength map, whereas DiffusionAD [17] reaches 97.9% AUROC at 0 pixels and performs poorly at other thresholds. Our approach maintains over 99% accuracy at all evaluated thresholds, underscoring the robustness of our distortion strength-based classification. Fig. 6 shows the correlation between the anomaly score and the annotated anomaly intensity. We compare annotated distortion strengths with both the sum of reconstruction errors from DiffusionAD and the computed distortion strengths from DiffuDewarp on the AnoCLIP dataset. DiffuDewarp's predictions align closely with the ground truths, with a linear fit yielding a slope of 1.01 and an intercept of 2.11 pixels—remarkably close to the ideal values of 1.00 and 0.00—whereas DiffusionAD's reconstruction errors do not align as well and exhibit high variance for normal samples, resulting in a non-negligible false positive rate.

4. CONCLUSION

In this paper, we propose DiffuDewarp, a diffusion-based anomaly detection method that leverages localized warping instead of conventional denoising diffusion models. By utilizing a distortion strength map predicted in the reverse diffusion process, we can not only detect anomalies but also estimate their severity. Through extensive experiments, we demonstrate that DiffuDewarp achieves state-of-the-art performance on three MVTec classes and effectively evaluates the degree of deformation in AnoCLIP.

However, our method currently relies on a heuristically designed local deformation field, limiting its capability to handle global displacement anomalies, such as transistor misplacements. Additionally, defining a general deformation model capable of handling all distortive anomalies remains challenging. Addressing these limitations is a key objective for future work, where we aim to incorporate few-shot learning techniques to improve generalization and robustness across various defect types.

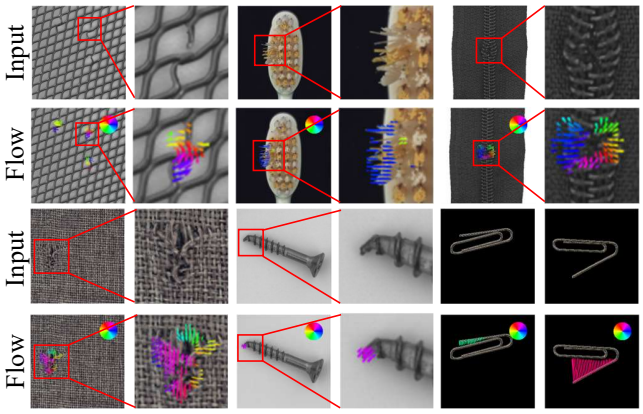
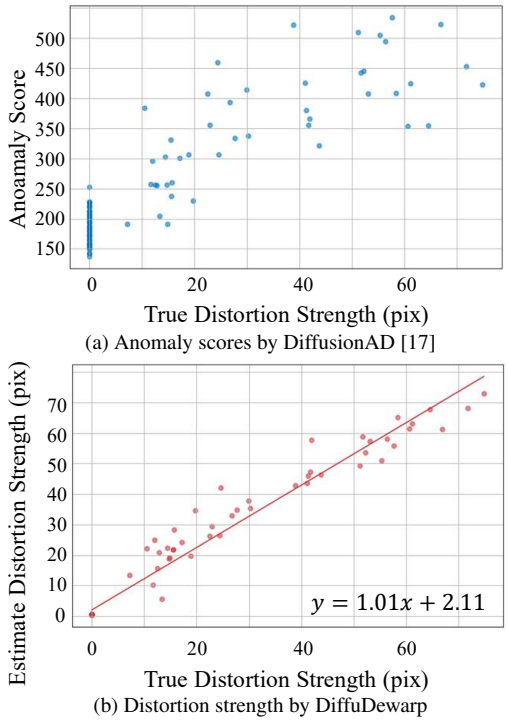


Fig. 5: Visualization of the distortion map by DiffuDewarp. Note that we modeled the forward process with different distortion models for MVTec and AnoCLIP.



(a) Anomaly scores by DiffusionAD [17]

(b) Distortion strength by DiffuDewarp

Fig. 6: Relationships of anomalities on the AnoClip.

5. REFERENCES

[1] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj, "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 8330–8339.

[2] Jinhui Zhao, Hongxia Gao, and Tongtong Liu, "Surface anomaly detection with anomalous feature restriction and difference-aware enhancement," in *2024 IEEE International Conference on Image Processing (ICIP)*, 2024, pp. 1377–1383.

[3] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 14318–14328.

[4] Hanqiu Deng and Xingyu Li, "Anomaly detection via reverse distillation from one-class embedding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 9737–9746.

[5] Niv Cohen and Yedid Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," *arXiv preprint arXiv:2005.02357*, 2020.

[6] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *International Conference on Pattern Recognition (PR)*. Springer, 2021, pp. 475–489.

[7] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, pp. 107706, 2021.

[8] Fabio Carrara, Giuseppe Amato, Luca Brombin, Fabrizio Falchi, and Claudio Gennaro, "Combining gans and autoencoders for efficient anomaly detection," in *2020 25th International Conference on Pattern Recognition (PR)*, 2021, pp. 3939–3946.

[9] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," *arXiv preprint arXiv:1807.02011*, 2018.

[10] Yong Shi, Jie Yang, and Zhiquan Qi, "Unsupervised anomaly segmentation via deep feature reconstruction," *Neurocomputing*, vol. 424, pp. 9–22, 2021.

[11] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.

[12] Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks, "Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 650–656.

[13] Fanbin Lu, Xufeng Yao, Chi-Wing Fu, and Jiaya Jia, "Removing anomalies as noises for industrial defect localization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 16166–16175.

[14] Arian Mousakhan, Thomas Brox, and Jawad Tayyub, "Anomaly detection with conditioned denoising diffusion models," *arXiv preprint arXiv:2305.15956*, 2023.

[15] Di Wu, Shicai Fan, Xue Zhou, Li Yu, Yuzhong Deng, Jianxiao Zou, and Baihong Lin, "Unsupervised anomaly detection via masked diffusion posterior sampling," *arXiv preprint arXiv:2404.17900*, 2024.

[16] Cosmin I Bercea, Michael Neumayr, Daniel Rueckert, and Julia A Schnabel, "Mask, stitch, and re-sample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models," *arXiv preprint arXiv:2305.19643*, 2023.

[17] Hui Zhang, Zheng Wang, Zuxuan Wu, and Yu-Gang Jiang, "Diffusionad: Norm-guided one-step denoising diffusion for anomaly detection," *arXiv preprint arXiv:2303.08730*, 2023.

[18] International Organization for Standardization, Geneva, *Geometrical product specifications (GPS) — Geometrical tolerancing — Tolerances of form, orientation, location and run-out*, 4nd edition, 2017, ISO 1101:2017.

[19] American Society of Mechanical Engineers (ASME), New York, *Dimensioning and Tolerancing*, 2018, ASME Standard Y14.5-2018.

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[21] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9592–9600.

[22] Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj, "Transfusion–a transparency-based diffusion model for anomaly detection," in *European Conference on Computer Vision (ECCV)*. Springer, 2025, pp. 91–108.

[23] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia, "Unsupervised surface anomaly detection with diffusion probabilistic model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 6782–6791.

[24] Ken Perlin, "An image synthesizer," *ACM Siggraph Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985.

[25] Ken Perlin, "Improving noise," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 681–682.