Visual Defect Obfuscation based Self-Supervised Anomaly Detection

YeongHyeon Park[1,2] Sungho Kang[1] Myung Jin Kim[2]
Yeonho Lee[1] Hyeong Seok Kim[2] Juneho Yi[1]
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University
[2]SK Planet Co., Ltd.

SCIENTIFIC REPORTS
natureresearch

YeongHyeon Park

Introduction

Reconstruction-by-inpainting based methods with an effective masking strategy of suspected defective regions enhance the UAD performance but there still remain issues to overcome.

1. Time-consuming inference due to multiple masking
2. Output inconsistency by random masking
3. Inaccurate reconstruction of normal patterns by large masks

This study proposes a novel reconstruction-by-inpainting method, dubbed *Excision And Recovery* (EAR).

- Pre-trained attention of DINO-ViT [1] effectively cuts out suspected defective regions and resolves issues 1 and 2
- Hint-providing proves to enhance the performance than emptying those regions by binary masking, thereby overcomes issue 3.

Design components of EAR

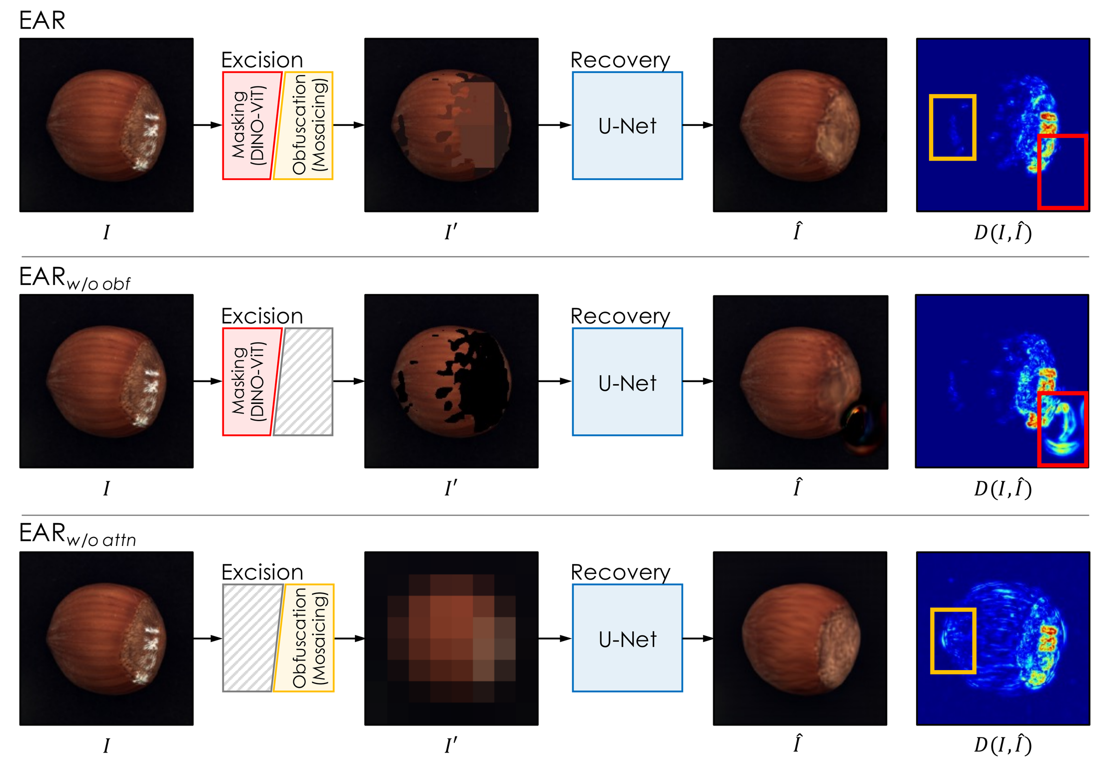

Figure 1: Visual comparison of the results when disabling each design component of EAR: visual obfuscation by mosaicing and saliency masking.

Mosaic scale prediction

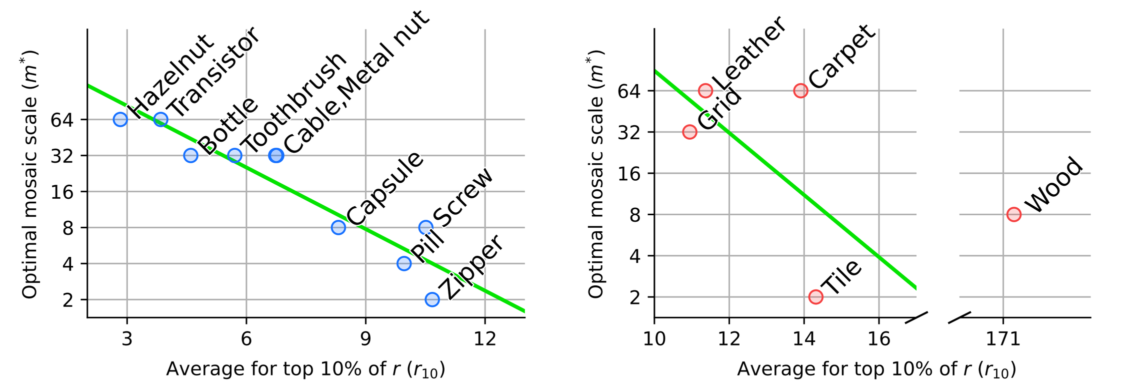

Figure 2: Linear regression model between r_{10} and m^*. m^* found by grid search is denoted by blue and red circles for r_{10}, and their correlation coefficients are -0.939 and -0.497 for 10 object subsets and 5 texture subsets, respectively.

Visual Defect Obfuscation

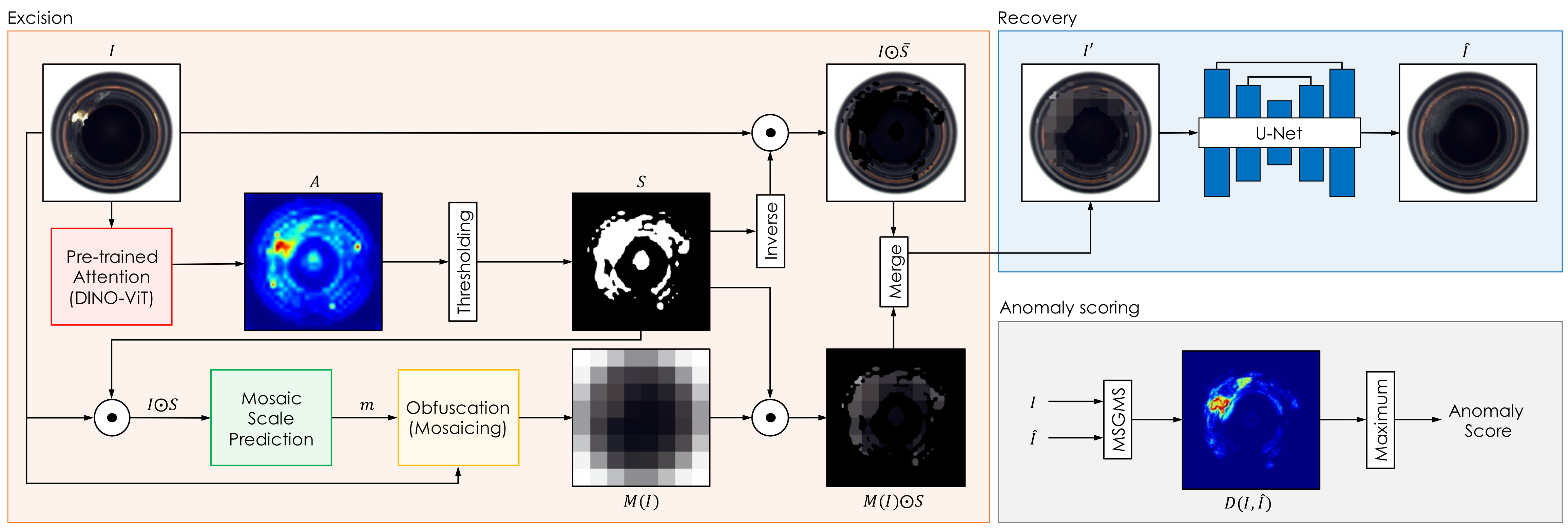

Figure 3: An overview of EAR. EAR takes the reconstruction-by-inpainting approach and is characterized by single deterministic masking and visual obfuscation of masked regions for hint-providing.

Reconstruction results

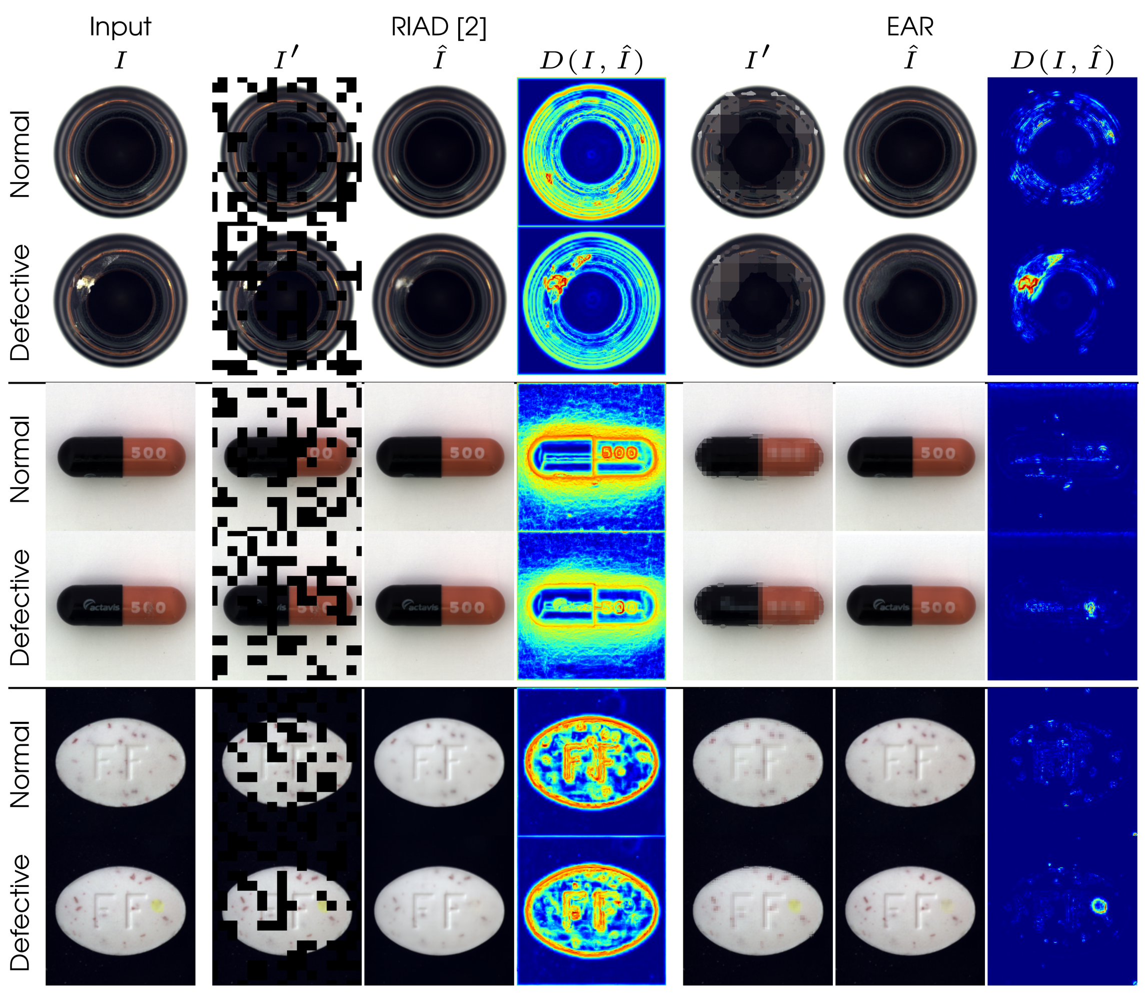

Figure 4: Visual comparison of RIAD [2] and EAR.

Results for industrial dataset

Table 1: Summary of the AUROC for the MVTec AD dataset [3]. For EAR, AUROCs are shown for two cases of $\hat{m}$ and m^*, in $\hat{m}$ (m^*) form. Abbreviations of attention module, discriminator, and memory module are 'Att', 'Dis', and 'Mem' respectively.

Model	MS-CAM	GANomaly	SCADN	MemAE	U-Net	DAAD	RIAD [2]	EAR (proposed)
Backbone	AE	AE	AE	AE	U-Net	U-Net	U-Net	U-Net
Additional Module	Att	Dis	Dis	Mem	-	Dis & Mem	-	-
Bottle	0.940	0.892	0.957	0.930	0.863	0.976	**0.999**	0.997 (0.997)
Cable	**0.880**	0.732	0.856	0.785	0.636	0.844	0.819	0.853 (0.871)
Capsule	0.850	0.708	0.765	0.735	0.673	0.767	**0.884**	0.870 (0.870)
Carpet	**0.910**	0.842	0.504	0.386	0.774	0.866	0.842	0.850 (0.899)
Grid	0.940	0.743	0.983	0.805	0.857	0.957	**0.996**	0.952 (0.959)
Hazelnut	0.950	0.794	0.833	0.769	0.996	0.921	0.833	**0.997 (0.997)**
Leather	0.950	0.792	0.659	0.423	0.870	0.862	**1.000**	**1.000 (1.000)**
Metal nut	0.690	0.745	0.624	0.654	0.676	0.758	**0.885**	0.856 (0.876)
Pill	0.890	0.757	0.814	0.717	0.781	0.900	0.838	**0.922 (0.922)**
Screw	**1.000**	0.699	0.831	0.257	**1.000**	0.987	0.845	0.779 (0.886)
Tile	0.800	0.785	0.792	0.718	0.964	0.882	**0.987**	0.918 (0.965)
Toothbrush	**1.000**	0.700	0.891	0.967	0.811	0.992	**1.000**	**1.000 (1.000)**
Transistor	0.880	0.746	0.863	0.791	0.674	0.876	0.909	**0.947 (0.947)**
Wood	0.940	0.653	0.968	0.954	0.958	0.982	0.930	**0.946 (0.985)**
Zipper	0.910	0.834	0.846	0.710	0.750	0.859	**0.981**	0.949 (0.955)
Average	0.902	0.761	0.812	0.707	0.819	0.895	0.917	**0.922 (0.942)**

This study proposes a strategy to **maximize the UAD performance without changing the NN structure.** Thus, the performance is compared with recent studies that use NNs of the same or similar scale.

- Best performance in hazelnut, pill, and transistor: The common characteristic of defective samples in these subtasks is surface damage which can be recovered into normal form by EAR.
- Relatively low performance in cases of capsules, screws, and zippers: The detailed pattern alignment of screw thread or the zipper teeth by reconstruction may be slightly missed due to visual defect obfuscation.

Computational efficiency

Table 2: Processing time for each training and inference.

Model	Training (sec)	Inference (msec)
RIAD [2]	35,478	366
$EAR_{w/o\ obf}$	3,084	156
$EAR_{w/o\ attn}$	**3,078**	**37**
EAR	3,109	197

Ablation study

Table 3: Summary of the ablation study.

Model	RIAD [2]	Ablations			EAR (proposed)
Masking	✓(multi)	✓		✓	✓
Hint			✓	✓	✓
KD				✓	
Bottle	0.999	0.995	**1.000**	0.994 (0.995)	0.997 (0.997)
Cable	0.819	0.795	**0.888**	0.851 (0.855)	0.853 (0.871)
Capsule	0.884	0.784	**0.918**	0.869 (0.869)	0.870 (0.870)
Carpet	0.842	0.848	0.718	0.846 (0.880)	**0.850 (0.899)**
Grid	**0.996**	0.969	0.963	0.976 (0.976)	0.952 (0.959)
Hazelnut	0.833	0.986	0.996	0.992 (0.996)	**0.997 (0.997)**
Leather	**1.000**	**1.000**	**1.000**	**1.000 (1.000)**	**1.000 (1.000)**
Metal nut	**0.885**	0.832	0.841	0.868 (0.868)	0.856 (0.876)
Pill	0.838	0.738	0.867	0.870 (0.873)	**0.922 (0.922)**
Screw	**0.845**	0.800	0.825	0.776 (0.854)	0.779 (0.886)
Tile	**0.987**	0.928	0.939	0.956 (0.956)	0.918 (0.965)
Toothbrush	**1.000**	0.994	**1.000**	**1.000 (1.000)**	**1.000 (1.000)**
Transistor	0.909	0.891	0.943	0.895 (0.933)	**0.947 (0.947)**
Wood	0.930	0.904	0.945	**0.986 (0.995)**	0.946 (0.985)
Zipper	**0.981**	0.900	0.963	0.951 (0.961)	0.949 (0.955)
Average	0.917	0.891	0.920	0.922 (0.934)	**0.922 (0.942)**

Conclusions

- The proposed **pre-trained spatial attention-based single deterministic masking method** has advanced the state-of-the-art methods in the reconstruction-by-inpainting approach for UAD, securing both higher throughput and output reliability.
- The proposed **hint-providing strategy by visual obfuscation on masked regions** further enhances the UAD performance with the proposed mosaic scale estimation method.

References

[1] Caron, M., *et al.* "Emerging properties in self-supervised vision transformers." ICCV 2021

[2] Zavrtanik, V., *et al.* "Reconstruction by inpainting for visual anomaly detection." *Pattern Recognition.* 2021

[3] Bergmann, P., *et al.* "MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection." CVPR 2019