Glancing at the Patch: Anomaly Localization with Global and Local Feature Comparison

210723

김찬영

Glancing at the Patch: Anomaly Localization with Global and Local Feature Comparison

- CVPR 2021에 공개된 Anomaly Detection 논문
- Local Information과 Global Information을 동시에 고려하여 이상 탐지한 모델

Glancing at the Patch: Anomaly Localization with Global and Local Feature Comparison

Shenzhi Wang[1] Liwei Wu[1] Lei Cui[1] Yujun Shen[2]
[1] SenseTime Research [2] The Chinese University of Hong Kong
{wangshenzhi, wuliwei, cuilei}@sensetime.com sy116@ie.cuhk.edu.hk

Abstract

Anomaly localization, with the purpose to segment the anomalous regions within images, is challenging due to the large variety of anomaly types. Existing methods typically train deep models by treating the entire image as a whole yet put little effort into learning the local distribution, which is vital for this pixel-precise task. In this work, we propose an unsupervised patch-based approach that gives due consideration to both the global and local information. More concretely, we employ a Local-Net and Global-Net to extract features from any individual patch and its surrounding respectively. Global-Net is trained with the purpose to mimic the local feature such that we can easily detect an abnormal patch when its feature mismatches that from the context. We further introduce an Inconsistency Anomaly Detection (IAD) head and a Distortion Anomaly Detection (DAD) head to sufficiently spot the discrepancy between global and local features. A scoring function derived from the multi-head design facilitates high-precision anomaly localization. Extensive experiments on a couple of real-world datasets suggest that our approach outperforms state-of-the-art competitors by a sufficiently large margin.

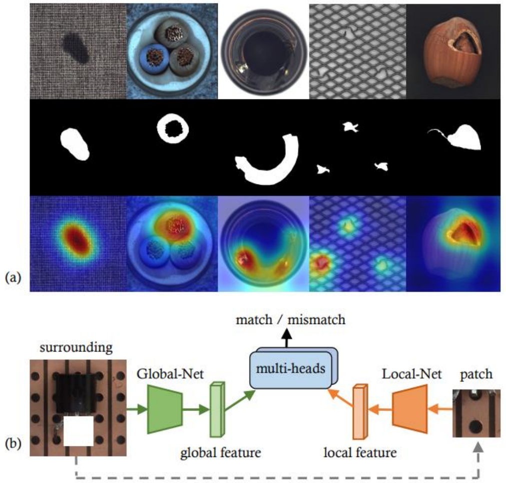


Figure 1. **(a) Anomaly localization results** where our approach can precisely segment the anomalous regions. From top to bottom: abnormal samples, ground-truth, and anomaly score maps produced by our algorithm. **(b) Concept diagram of global and local feature comparison.** Local-Net and Global-Net are employed to extract features from a patch and its surrounding

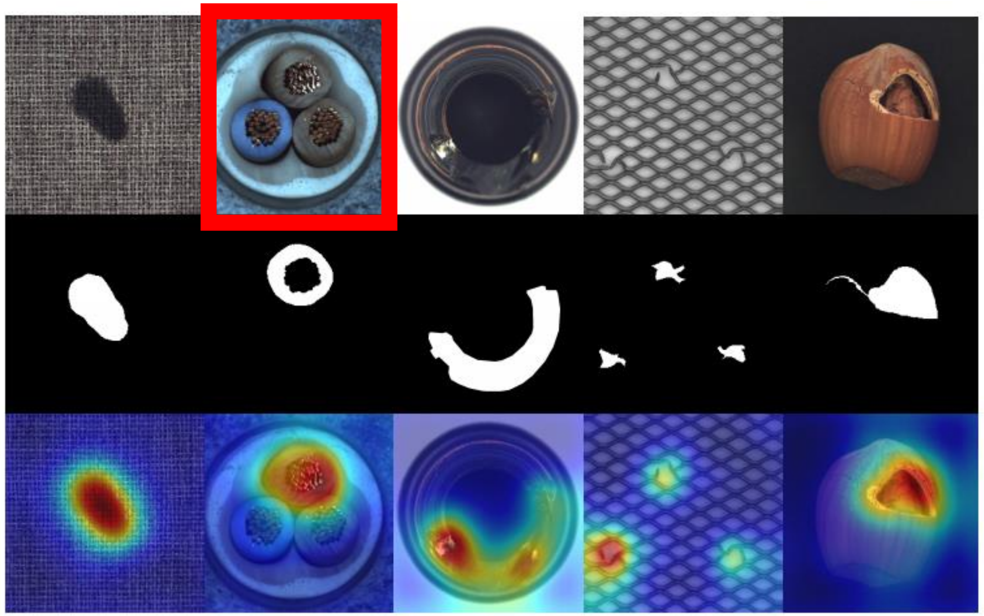
Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Glancing at the Patch: Anomaly Localization with Global and Local Feature Comparison

- 기존의 AutoEncoder과 GAN은 Local Information은 생략하고 전체 이미지에 대한 정보만 다룸

- 이미지에서 모든 패치에 대한 검사를 진행하는 방향은 패치와 주변 환경 간의 상관 관계를 고려하지 않음

- 그 결과, 오른쪽 빨간색 케이블에서 초록색이어야 하는 영역을 검출하지 못함



Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Contribution of Glancing at the Patch

- 본 논문에서는 Global과 Local Information을 동시에 고려함
- Patch와 Surrounding을 각각 Local-Net과 Global-Net으로 특징을 추출하는 방식을 채택함
- IAD(Inconsistency Anomaly Detection) Head와 DAD(Distortion Anomaly Detection) Head를 도입해 global feature과 local feature 간의 유사성을 더 잘 측정하기 위한 융합 메트릭을 채택함

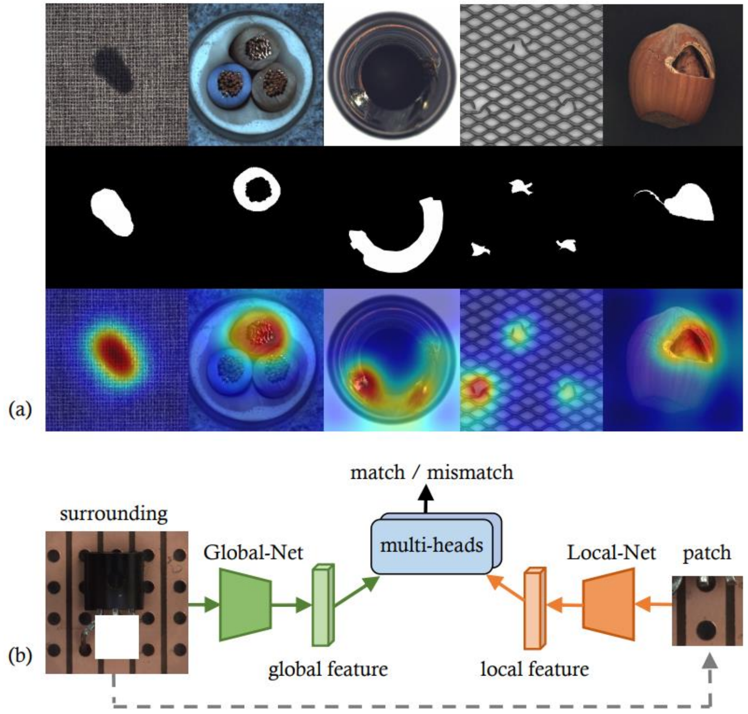


Figure 1. **(a) Anomaly localization results** where our approach can precisely segment the anomalous regions. From top to bottom: abnormal samples, ground-truth, and anomaly score maps produced by our algorithm. **(b) Concept diagram of global and local feature comparison.** Local-Net and Global-Net are employed to extract features from a patch and its surrounding respectively. Multiple anomaly detection heads are designed to determine whether the global and local features match or not.

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Local-Net

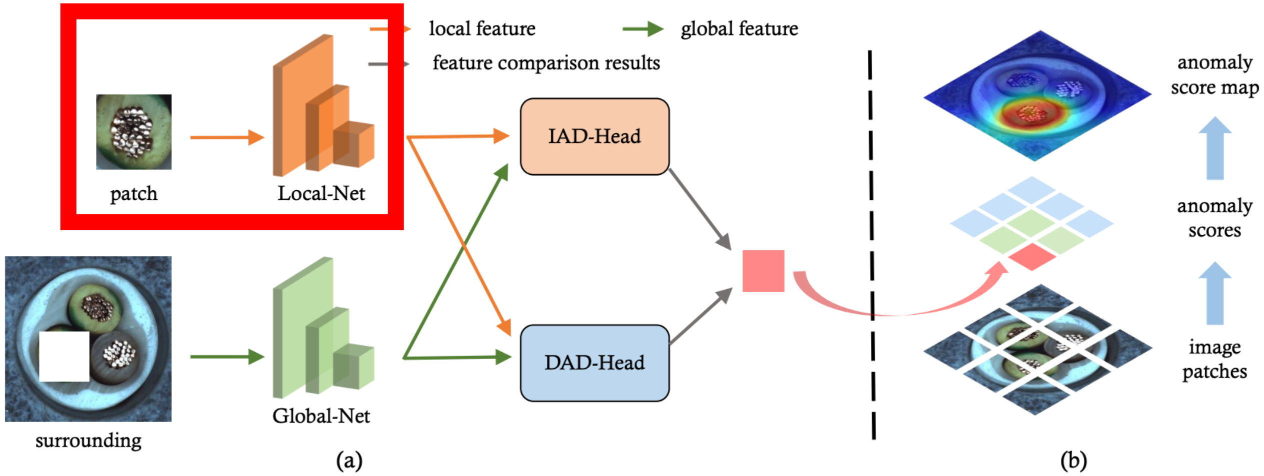


Figure 2. **Anomaly Localization Framework.** (a) At the training stage, a Local-Net and Global-Net are employed to extract features from an image patch and its surrounding respectively. The Global-Net is jointly learned with an Inconsistency Anomaly Detection (IAD) head and a Distortion Anomaly Detection (DAD) head to mimic the output from the Local-Net. (b) At the inference stage, a scoring function is developed based on the feature comparison results produced by the IAD-head and DAD-head. Anomaly scores corresponding to different patches are aggregated together into an anomaly score map for anomaly localization.

Local-Net

- 경량 신경망인 Pretrained ResNet-18으로부터 Knowledge Distillation 후 Fine-tuning

$$l_k = ||\mathcal{D}(\mathcal{L}(\mathbf{p})) - \mathcal{R}(\mathbf{p})||_2^2,$$

The compactness loss is formulated as

p : *image patch*

$||\cdot||_2$: L_2 *norm*

$\mathcal{L}(\cdot)$: Local-Net

$\mathcal{R}(\cdot)$: Teacher Model

$\mathcal{D}$: Decoder to ensure $\mathcal{L}$ & $\mathcal{R}$ to have same output dimension

c_{ij}: correlation matrix over the local net outputs

- Optimize witl $l_{local} = \lambda_k l_k + \lambda_c l_c,$
- Distillation과 fine-tuning 후 Local Feature은

$$\mathbf{Z}_l = \mathcal{L}(\mathbf{p})$$

다음과 같이 표현

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Global-Net

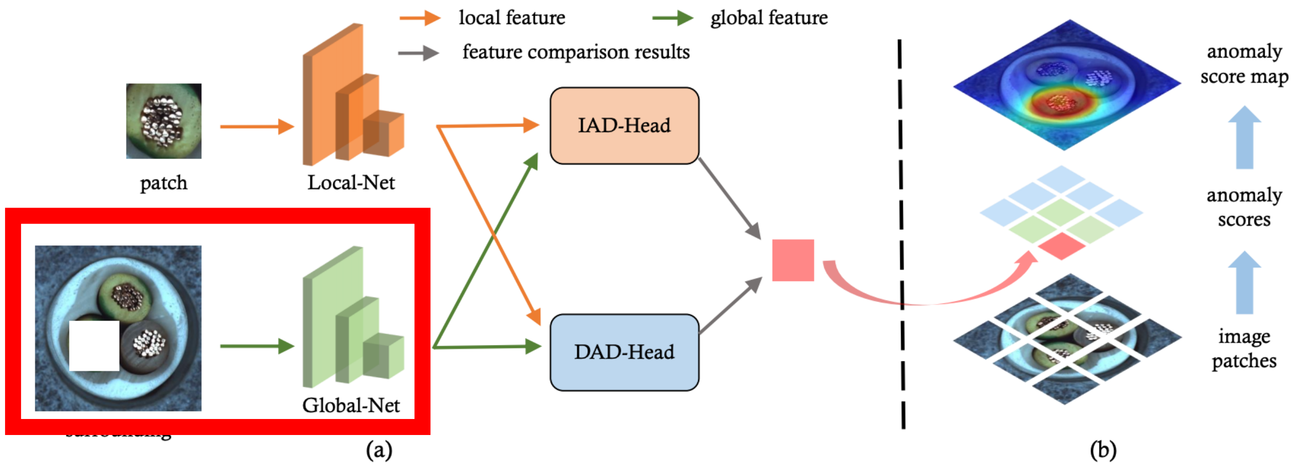


Figure 2. **Anomaly Localization Framework.** (a) At the training stage, a Local-Net and Global-Net are employed to extract features from an image patch and its surrounding respectively. The Global-Net is jointly learned with an Inconsistency Anomaly Detection (IAD) head and a Distortion Anomaly Detection (DAD) head to mimic the output from the Local-Net. (b) At the inference stage, a scoring function is developed based on the feature comparison results produced by the IAD-head and DAD-head. Anomaly scores corresponding to different patches are aggregated together into an anomaly score map for anomaly localization.

Global-Net

- Global-Net이라는 또 다른 심층 모델을 사용하여 패 치 주변에서 global feature를 추출
- Local Feature가 Global Feature를 방해하는 것을 막기 위해 partial convolution을 Global-Net에 적용

$$x' = \begin{cases} \mathbf{W}^{\mathbf{T}}(\mathbf{X} \odot \mathbf{M}) \frac{\text{sum}(\mathbf{1})}{\text{sum}(\mathbf{M})} + b, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

$\odot$: $element - wise\ product$

X : input feature map

M : binary mask in current layer

- Global Feature를 다음과 같이 표현

$$\mathbf{Z}_g = \mathcal{G}(\mathbf{I}, \mathbf{M_0}).$$

M_0 : binary matrix where the patch's pixels are zero and the others are one

$\mathcal{G}(\cdot)$: Global-Net

Z_g : global feature

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Heads

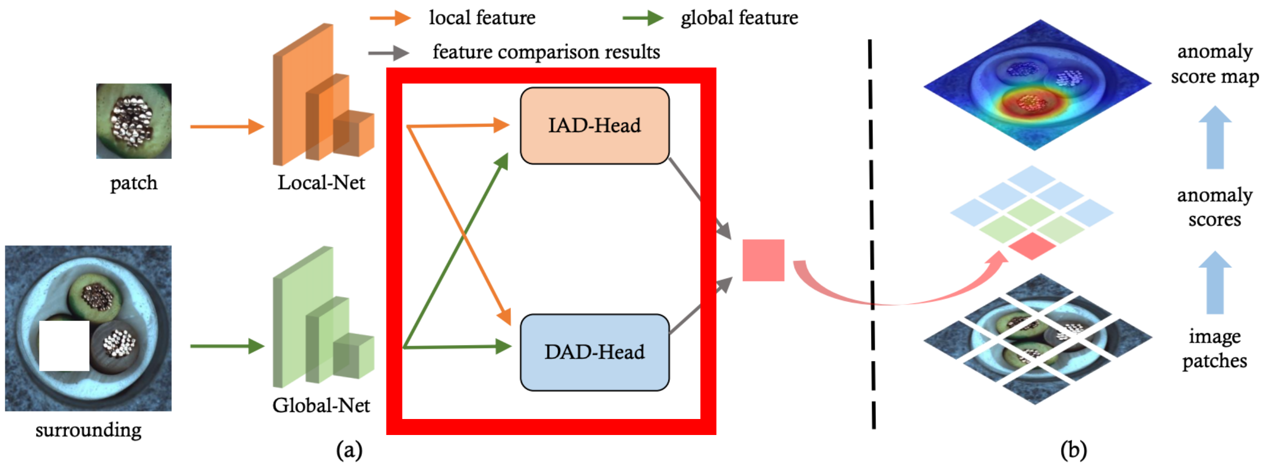


Figure 2. **Anomaly Localization Framework.** (a) At the training stage, a Local-Net and Global-Net are employed to extract features from an image patch and its surrounding respectively. The Global-Net is jointly learned with an Inconsistency Anomaly Detection (IAD) head and a Distortion Anomaly Detection (DAD) head to mimic the output from the Local-Net. (b) At the inference stage, a scoring function is developed based on the feature comparison results produced by the IAD-head and DAD-head. Anomaly scores corresponding to different patches are aggregated together into an anomaly score map for anomaly localization.

Heads

- IAD-Head & DAD-Head는 local feature과 global feature를 추출해 두 feature를 비교함
- 구부러진 격자 및 절단 카펫. 패치와 그 주변의 불일치에 초점을 맞춘 IAD 헤드에 비해 DAD 헤드는 패치에 국한된 작은 결함을 찾아

Inconsistency Anomaly Detection Head

- local feature과 global feature간의 차이를 검출

$$l_{\text{IAD}} = \frac{1}{n}||\mathbf{Z}_l - \mathbf{Z}_g||_2^2,$$

Z_l : local feature

Z_g : global feature

Distortion Anomaly Detection Head

- 이미지의 결함을 감지하는 것을 목표로 하는 훈련 가능한 Head

$$p = \mathcal{C}(\mathbf{Z}^*, \mathbf{Z}_g),$$

$\mathcal{C}$: is the classification network in the DAD-head

Z^*: can be either local feature or negative local feature

$$l_{\text{DAD}} = -(y\log(p) + (1-y)\log(1-p)),$$

y : target output of the classifier

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Scoring Function

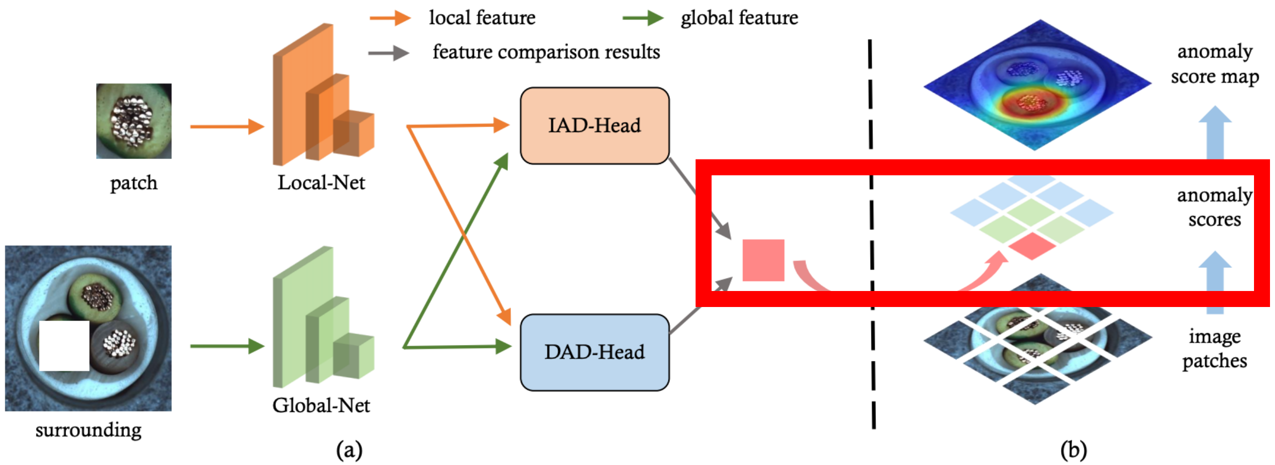


Figure 2. **Anomaly Localization Framework.** (a) At the training stage, a Local-Net and Global-Net are employed to extract features from an image patch and its surrounding respectively. The Global-Net is jointly learned with an Inconsistency Anomaly Detection (IAD) head and a Distortion Anomaly Detection (DAD) head to mimic the output from the Local-Net. (b) At the inference stage, a scoring function is developed based on the feature comparison results produced by the IAD-head and DAD-head. Anomaly scores corresponding to different patches are aggregated together into an anomaly score map for anomaly localization.

Scoring Function

$$s_{\mathrm{IAD}} = \frac{1}{n}||\mathbf{Z}_l - \mathbf{Z}_g||_2^2.$$

$$s_{\mathrm{DAD}} = 1 - \mathcal{C}(\mathbf{Z}_l, \mathbf{Z}_g).$$

$$s = \lambda_s s_{\mathrm{IAD}} + (1 - \lambda_s) s_{\mathrm{DAD}},$$

λ_s : hyper-parameter to balance the inconsistency anomaly score

본 논문에서는 0.8로 설정 후 실험함

Anomaly Score Map

- 특정 패치에 이상 점수를 할당하는 스코어링 기능을 사 용하여 다른 패치에 대한 이상 점수를 이상 점수 맵으로 집계하는 파이프라인

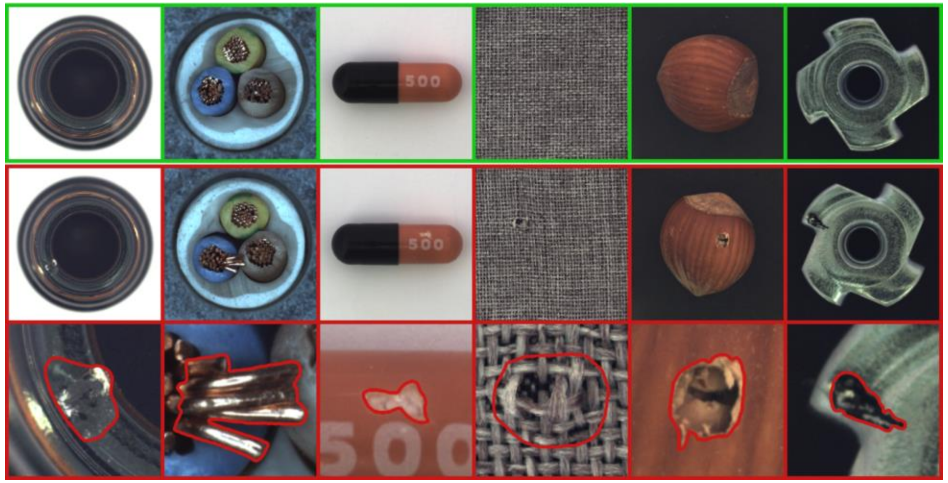
Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Experiment

Experiment Dataset

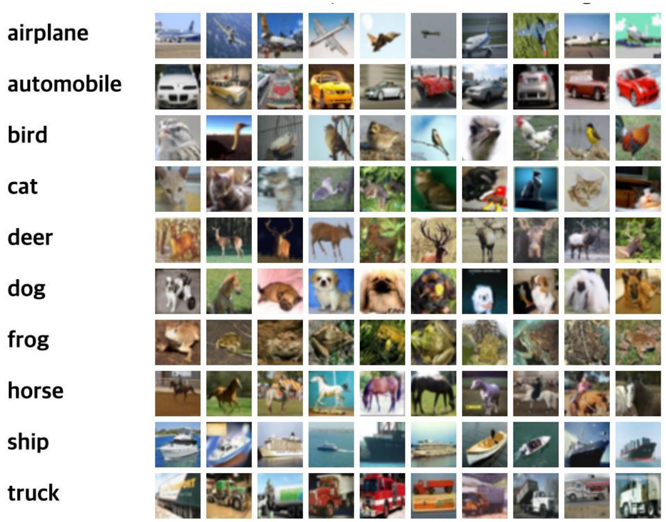
MVTec AD

- 15개 카테고리의 5354개 고해상도 이미지가 포함된 실제 산업 이미지 이상 감지 데이터셋
- Train Data : 3629장, Test Data : 1725장



CIFAR-10

- 10개의 클래스가 있는 60000개의 작은 이미지로 구성된 데이터셋
- Train Data : 50000장, Test Data : 10000장



Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Experiment

Pixel-level Anomaly Localization

Table 1. Comparison results among different anomaly detection methods in the **pixel-level anomaly localization task on MVTec AD dataset** [6]. Competitors include 1-NN [3], OC-SVM [43], K-Means [29], l_2-AE [16], VAE [4], SSIM-AE [8], AnoGAN [42], CNN-FD [32] and TS [7]. The results of baselines are borrowed from [6, 7]. **Per-region-overlap (PRO)** [7] is used as the evaluation metric.

	Category	1-NN	OC-SVM	K-Means	l_2-AE	VAE	SSIM-AE	AnoGAN	CNN-FD	TS	Ours
Texture	Carpet	0.512	0.355	0.253	0.456	0.501	0.647	0.204	0.469	0.879	**0.977**
	Grid	0.228	0.125	0.107	0.582	0.224	0.849	0.226	0.183	**0.952**	0.932
	Leather	0.446	0.306	0.308	0.819	0.635	0.561	0.378	0.641	**0.945**	0.909
	Tile	0.822	0.722	0.779	0.897	0.870	0.175	0.177	0.797	**0.946**	0.883
	Wood	0.502	0.336	0.411	0.727	0.628	0.605	0.386	0.621	0.911	**0.941**
Object	Bottle	0.898	0.850	0.495	0.910	0.897	0.834	0.620	0.742	0.931	**0.968**
	Cable	0.806	0.431	0.513	0.825	0.654	0.478	0.383	0.558	0.818	**0.980**
	Capsule	0.631	0.554	0.387	0.862	0.526	0.860	0.306	0.306	**0.968**	0.960
	Hazelnut	0.861	0.616	0.698	0.917	0.878	0.916	0.698	0.844	**0.965**	0.962
	Metal Nut	0.705	0.319	0.351	0.830	0.576	0.603	0.320	0.358	0.942	**0.967**
	Pill	0.725	0.544	0.514	0.893	0.769	0.830	0.776	0.460	0.961	**0.978**
	Screw	0.604	0.644	0.550	0.754	0.559	0.887	0.466	0.277	0.942	**1.000**
	Toothbrush	0.675	0.538	0.337	0.822	0.693	0.784	0.749	0.151	0.933	**0.961**
	Transistor	0.680	0.496	0.399	0.728	0.626	0.725	0.549	0.628	0.666	**0.999**
	Zipper	0.512	0.355	0.253	0.839	0.549	0.665	0.467	0.703	0.951	**0.992**
	Mean	0.640	0.479	0.423	0.790	0.639	0.694	0.443	0.515	0.914	**0.961**

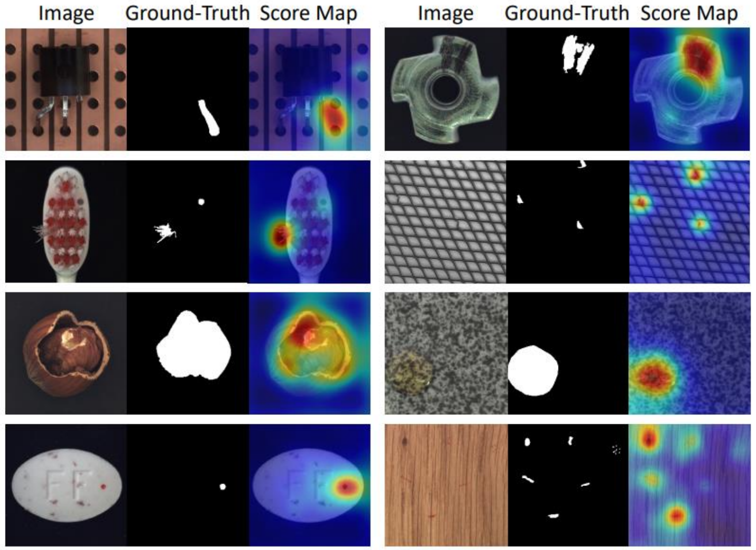


Figure 3. Qualitative anomaly localization results on MVTec AD dataset [6]. For each example, the images from left to right are the defective image, the ground-truth, and the anomaly score map produced by our algorithm. Zoom in for details.

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.

Experiment

Image-level Anomaly Detection

Table 3. Comparison results among different one-class classification methods in the **image-level anomaly detection task on CIFAR-10** [21]. Competitors include OC-SVM [43], KDE [9], l_2-AE [16], VAE [4], PixelCNN [45], LSA [1], AnoGAN [42], DSVDD [38], OCGAN [37] and GradCon [22]. The results of baselines are borrowed from [22, 37]. **Image-level AUROC** is utilized as the evaluation metric.

Normal Class	OC-SVM	KDE	l_2-AE	VAE	PixelCNN	LSA	AnoGAN	DSVDD	OCGAN	GradCon	Ours
Airplane	0.630	0.658	0.411	0.634	0.788	0.735	0.671	0.617	0.757	0.760	**0.791**
Automobile	0.440	0.520	0.478	0.442	0.428	0.580	0.547	0.659	0.531	0.598	**0.703**
Bird	0.649	0.657	0.616	0.640	0.617	0.690	0.529	0.508	0.640	0.648	**0.675**
Cat	0.487	0.497	0.562	0.497	0.574	0.542	0.545	0.591	**0.620**	0.586	0.561
Deer	0.735	0.727	0.728	0.743	0.511	**0.761**	0.651	0.609	0.723	0.733	0.739
Dog	0.500	0.496	0.513	0.515	0.571	0.546	0.603	**0.657**	0.620	0.603	0.638
Frog	0.725	**0.758**	0.688	0.745	0.422	0.751	0.585	0.677	0.723	0.684	0.732
Horse	0.533	0.564	0.497	0.527	0.454	0.535	0.625	0.673	0.575	0.567	**0.674**
Ship	0.649	0.680	0.487	0.674	0.715	0.717	0.758	0.759	**0.820**	0.784	0.814
Truck	0.508	0.540	0.378	0.416	0.426	0.548	0.665	0.731	0.554	0.678	**0.722**
Mean	0.586	0.610	0.536	0.583	0.551	0.641	0.618	0.648	0.657	0.664	**0.705**

Wang, Shenzhi, et al. "**Glancing at the Patch: Anomaly Localization With Global and Local Feature Comparison**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2021.