PSVL

RE • WORK
DEEP LEARNING HYBRID SUMMIT

Normalizing Flows for Real-Time Unsupervised Anomaly Detection

Denis Gudovskiy
Panasonic AI Lab, Mountain View, CA
Email: denis.gudovskiy@us.panasonic.com

Panasonic AI

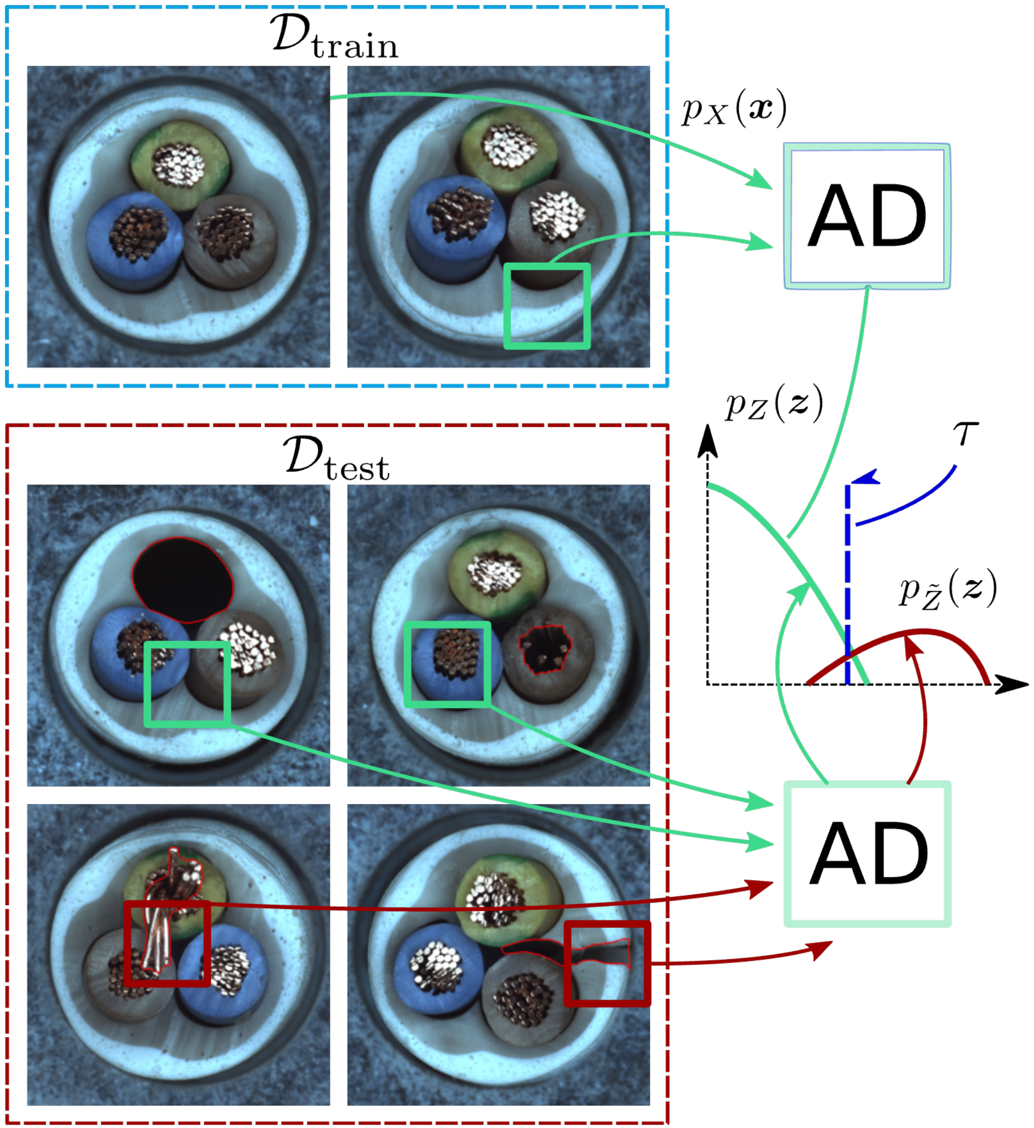
Background: Panasonic Business

Deployments outside of a geo-fenced areas are not scalable without robust perception

https://news.panasonic.com/global/press/data/2020/12/en201214-1/en201214-1.html
https://www.youtube.com/watch?v=G4BnC4NZ7vE

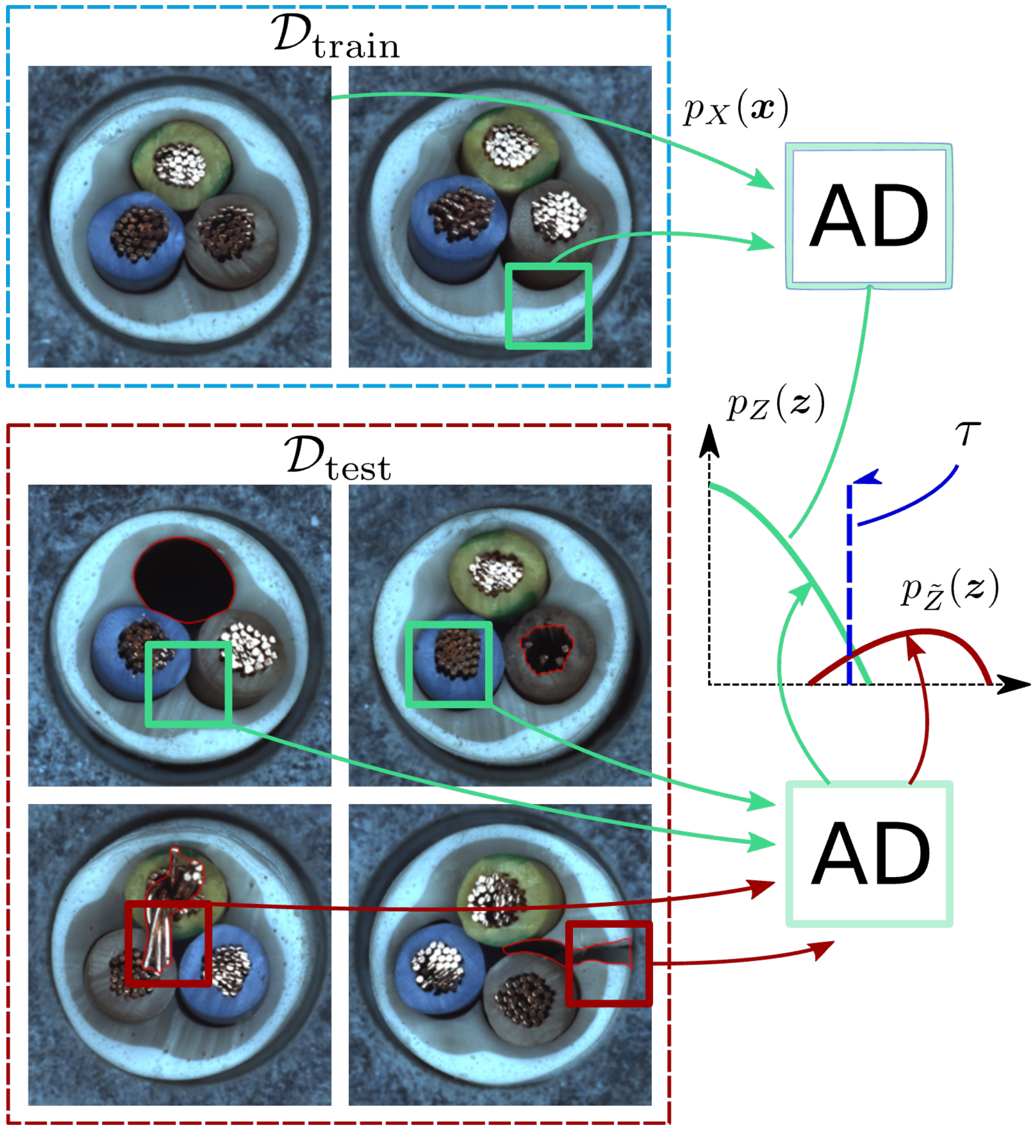
Motivation: Unsupervised Anomaly Detection as OOD

○ Anomaly detection (AD) can be used in many practical applications: robust AV perception systems, industrial inspection, road traffic monitoring, medical diagnostics *and many others*



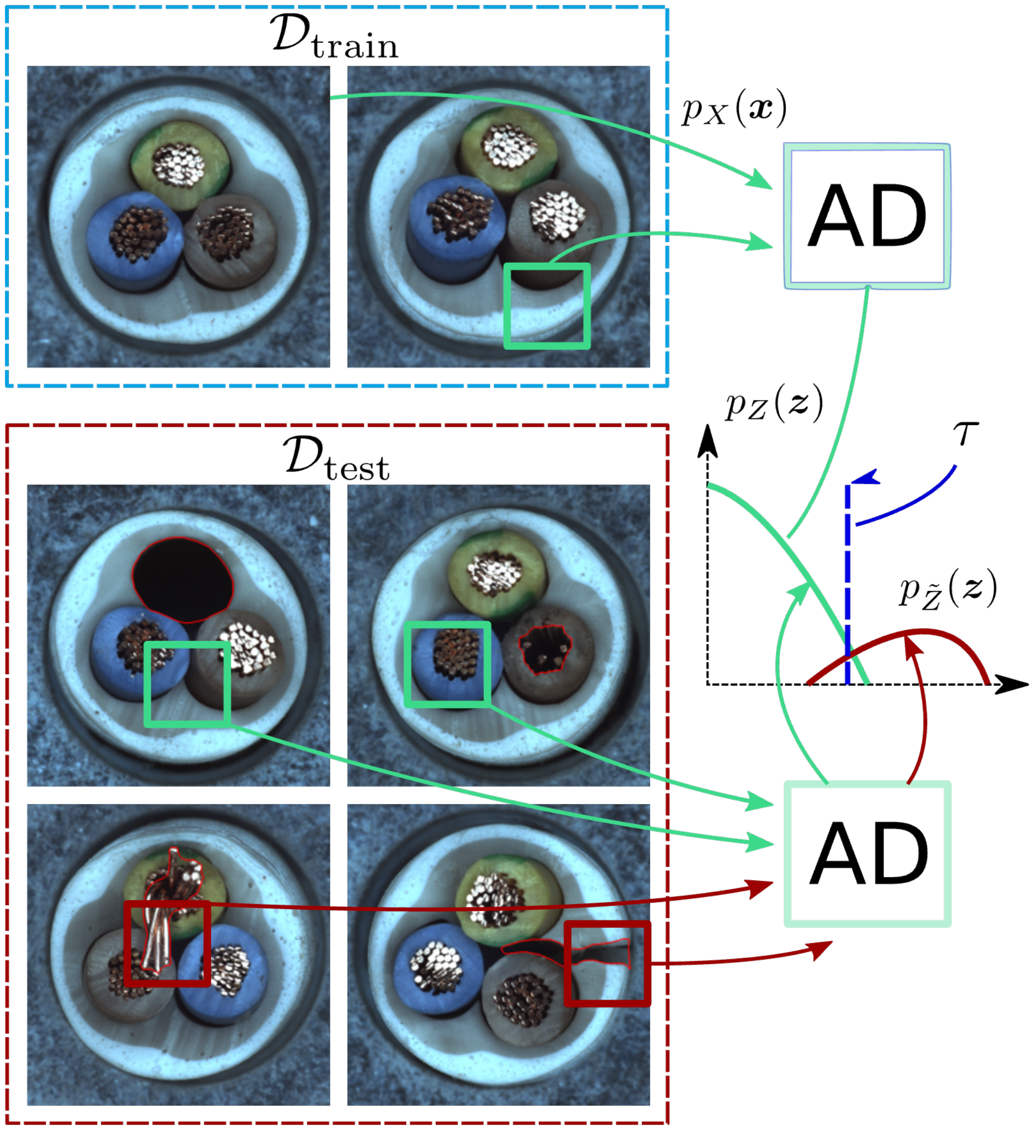
Motivation: Unsupervised Anomaly Detection as OOD

- Anomaly detection (AD) can be used in many practical applications: robust AV perception systems, industrial inspection, road traffic monitoring, medical diagnostics *and many others*
- However, the supervised AD requires costly annotations and, in some cases, is not applicable



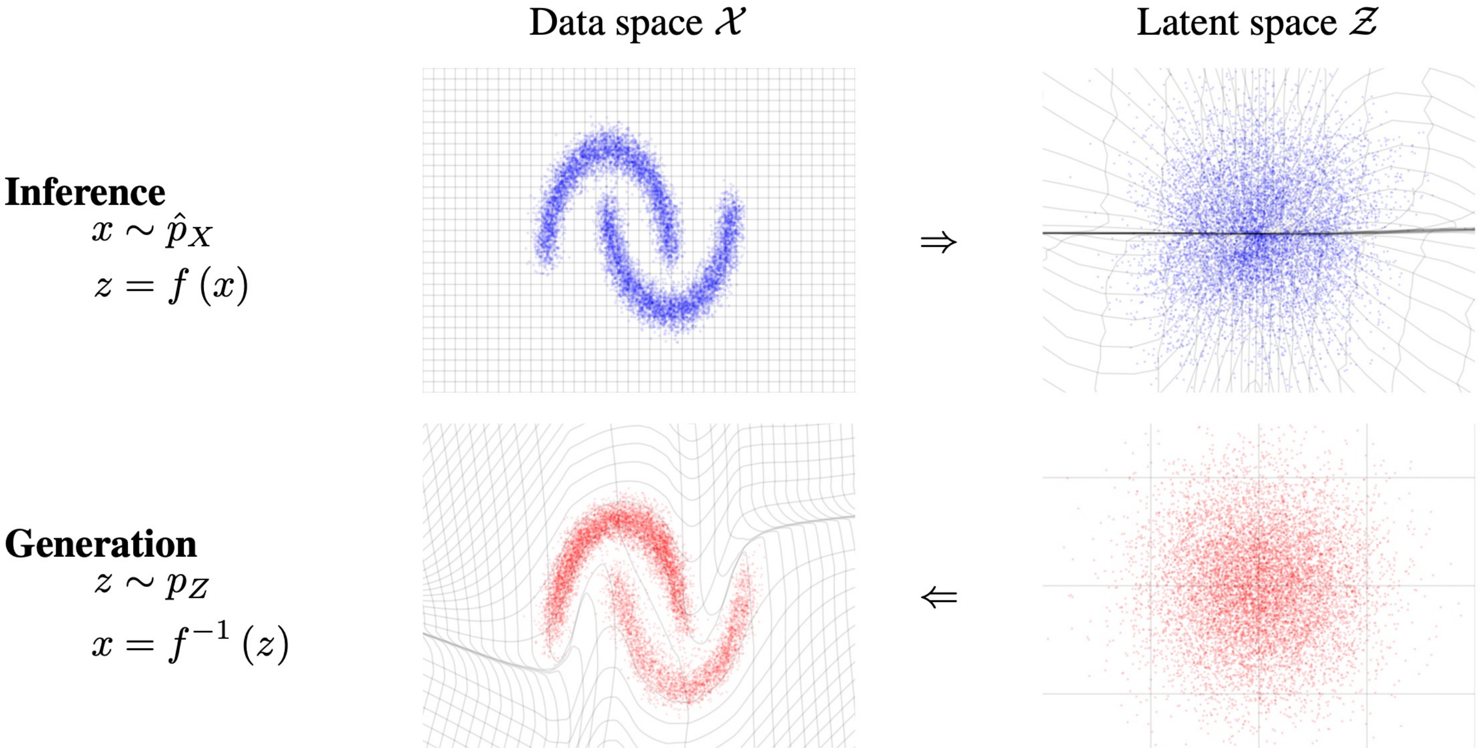
Motivation: Unsupervised Anomaly Detection as OOD

- Anomaly detection (AD) can be used in many practical applications: robust AV perception systems, industrial inspection, road traffic monitoring, medical diagnostics *and many others*
- However, the supervised AD requires costly annotations and, in some cases, is not applicable
- A more appealing approach is to collect only unlabeled anomaly-free images for a train dataset
- Then, any deviation from distribution of anomaly-free images can be classified as an anomaly
- Hence, the AD task can be reformulated as a task of out-of-distribution detection (OOD) with the objective to estimate data likelihoods



Solution: Normalizing Flows for OOD

- Unlike others, normalizing flows can estimate the exact data likelihoods $\hat{p}_X(x, \boldsymbol{\theta}) \cong p_X(\boldsymbol{x})$
- A set of invertible layers convert an arbitrary density $p_X(\boldsymbol{x})$ to a base distribution $p_Z(\boldsymbol{z})$
- Then, the $\log \hat{p}_X(\boldsymbol{x}, \boldsymbol{\theta}) = \log p_Z(\boldsymbol{z}) + \sum_l \log|\det \boldsymbol{J}_l|$, where a sample $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$ and a Jacobian determinants can be efficiently computed for certain layer architectures[1]



References:

[1] Laurent Dinh, Jascha Sohl-Dickstein, Samy Bengio. Density estimation using Real NVP. In ICLR, 2017

Our Work: Conditional Normalizing Flows for OOD

- In our recent CFLOW-AD paper[1], we propose to extend conventional flow models by incorporating a conditional vector $\boldsymbol{c}$ to encode spatial information into the model:
  - We are interested in anomaly segmentation task for perception systems
  - Our conditional vector contains sin/cos harmonics from a positional encoding[2]
  - It is concatenated with the intermediate outputs inside each flow's coupling layer
  - We efficiently share flow parameters $\boldsymbol{\theta}$ between feature map's spatial dimensions

- Finally, we train our conditional flow model $\hat{p}(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{\theta})$ for AD (CFLOW-AD) using conventional maximum likelihood objective as:

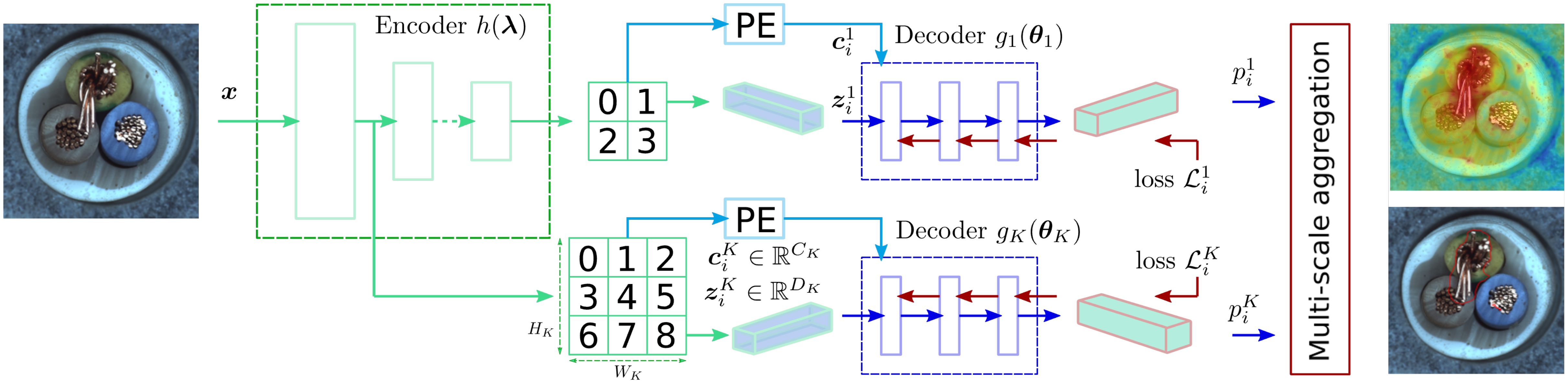
$$L(\boldsymbol{\theta}) = D_{KL}[p_X(\boldsymbol{x}) \parallel \hat{p}_X(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{\theta})] \approx \frac{1}{N} \sum_{i=1}^{N} [\|\boldsymbol{z}_i\|_2^2/2 - \log|\det \boldsymbol{J_i}|] + \text{const}$$

References:

[1] Denis Gudovskiy et al. CFLOW-AD: Real-Time Unsupervised AD with Localization via Conditional Flows. In WACV, 2022

[2] Ashish Vaswani et al. Attention Is All You Need. In Advances in Neural Information Processing Systems, 2017

Our Work: CFLOW-AD Architecture



- Encoder is a conventional CNN/transformer feature extractor pretrained on natural images
- Multi-scale pyramid pooling extracts local and global features in the latent space
- Decoders are the flow models with the positional encoding (PE) conditional inputs
- We estimate a final anomaly score map by aggregating multi-scale likelihoods

Experiments: MVTec and STC

- MVTec and STC are the datasets with factory defects and surveillance camera videos
- AUROC and AUPRO are popular threshold-agnostic metrics for AD

Table 1: Average AUROC and AUPRO on the MVTec dataset, %. Both the best detection and localization metrics are presented, if available. CFLOW-AD is with WideResNet-50 encoder.

Metric	AUROC		AUPRO
Model	Detection	Localization	
SVDD [5]	92.1	95.7	-
SPADE [1]	85.5	96.0	91.7
CutPaste [3]	97.1	96.0	-
PaDiM [2]	97.9	97.5	92.1
CFLOW-AD (ours)	**98.26**	**98.62**	**94.60**

Table 2: Average AUROC on the STC dataset, %. Both the best available detection and localization metrics are showed. CFLOW-AD is with WideResNet-50 encoder.

Metric	AUROC	
Model	Detection	Localization
CAVGA [4]	-	85.0
SPADE [1]	71.9	89.9
PaDiM [2]	-	91.2
CFLOW-AD (ours)	**72.63**	**94.48**

References

[1] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv:2005.02357v3*, 2021.

[2] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In *ICPR Workshops*, 2021.

[3] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. CutPaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, 2021.

[4] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *ECCV*, 2020.

[5] Jihun Yi and Sungroh Yoon. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation. In *ACCV*, 2020.

CFLOW-AD is a Real-Time Model

- Previous methods have high complexity:
  - ✓ Pretrained encoder is fully-convolutional and fast
  - ❖ Post-processing is slow due to high memory consumption
- CFLOW-AD has significantly lower complexity:
  - ✓ Encoder and decoders are fully-convolutional
  - ✓ Memory requirements are a factor of 10× lower
  - ✓ Hence, inference speed is much higher on a GPU
  - ✓ Lightweight MobileNetV3L encoder has a minor drop in performance compared to WideResNet-50
- CFLOW-AD is suitable for AD on edge devices:
  - ✓ Less than 25MB model for MobileNetV3L
  - ✓ Inference speed is ~35 fps on 1080 GPU

Table 6. Complexity comparison in terms of inference speed (fps) and model size (MB). Inference speed for CFLOW-AD models from Table 3 is measured for (256×256) / (512×512) inputs.

Complexity metric and Model	Inference speed, fps	Model size, MB STC	Model size, MB MVTec
R18 encoder only	80 / 62	45	
PaDiM-R18 [2]	4.4	210	170
CFLOW-AD-R18	34 / 12	96	
WR50 encoder only	62 / 30	268	
SPADE-WR50 [1]	0.1	37,000	1,400
PaDiM-WR50 [2]	1.1	5,200	3,800
CFLOW-AD-WR50	27 / 9	947	
MNetV3 encoder only	82 / 61	12	
CFLOW-AD-MNetV3	35 / 12	25	

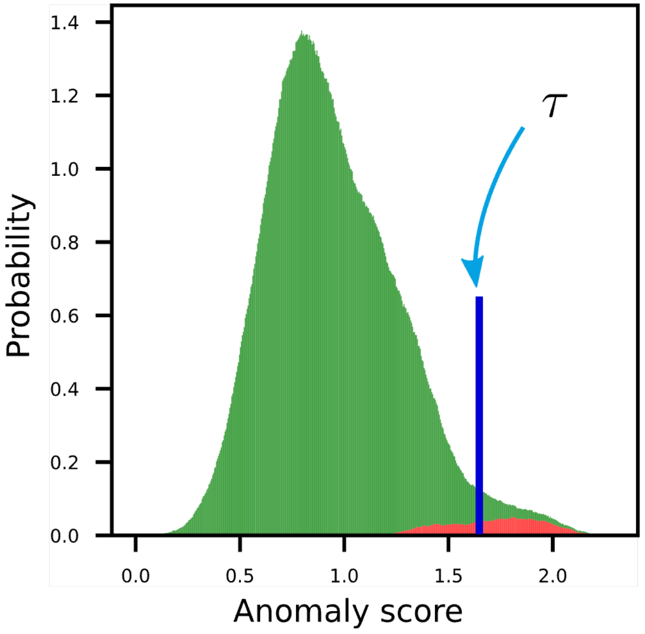
References

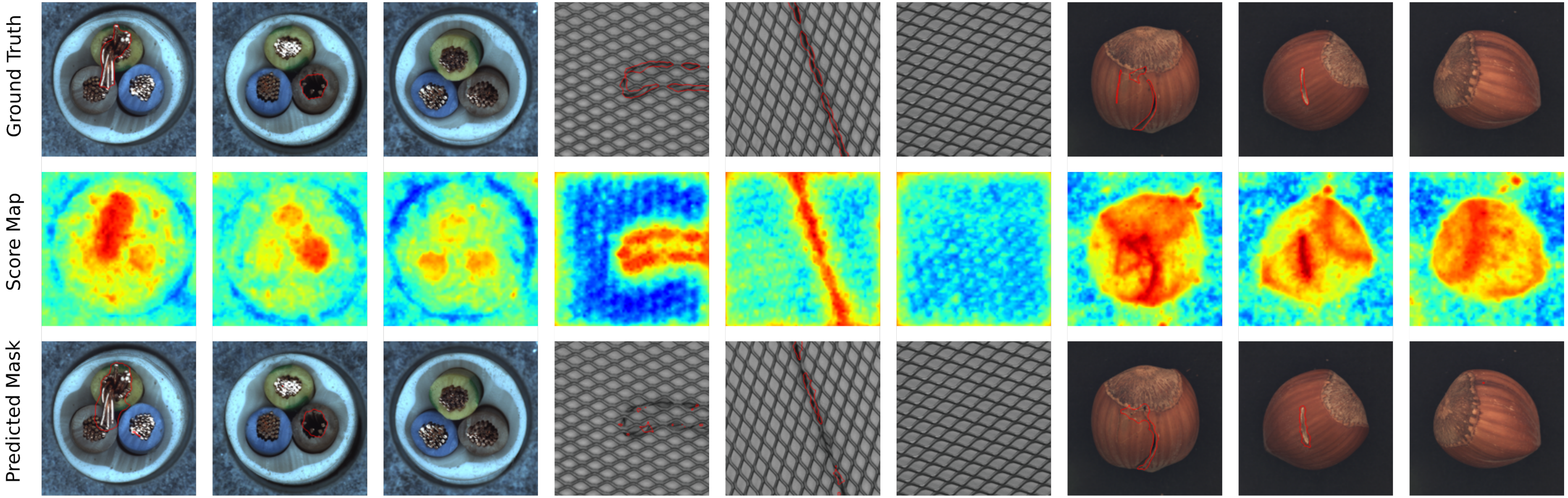
[1] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv:2005.02357v3*, 2021.

[2] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: a patch distribution modeling framework for anomaly detection and localization. In *ICPR Workshops*, 2021.

CFLOW-AD Qualitative Results

- Anomaly score distribution (right) proves successful OOD
- Ground truth masks (below) are from the MVTec test dataset
- Score maps are the aggregated CFLOW-AD anomaly scores
- Predicted masks are selected using the F_1-maximized threshold
- Positives and negatives are successfully detected by CFLOW-AD





Conclusions

code to reproduce experiments:
github.com/gudovskiy/cflow-ad

- ✓ Normalizing flow models work well for unsupervised AD
- ✓ Small tweaks such as in CFLOW-AD allow real-time processing while being SOTA
- ✓ Within a year, the MVTEC-AD leaderboard is switched to FLOW-based models

Rank	Model	Detection AUROC	**Segmentation AUROC**	Overall AUC	Extra Training Data	Paper	Code	Result	Year	Tags
1	**CFLOW-AD**	98.26	98.62		✓	CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows			2021	
2	**Fastflow**	99.4	98.5		✓	FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows			2021	Transformer ResNet

Thank you for Attention!
Questions?

Panasonic AI