Neural Network Training Strategy to Enhance Anomaly Detection Performance: A Perspective on Reconstruction Loss Amplification

YeongHyeon Park[1,2] Sungho Kang[1] Myung Jin Kim[2] Hyeonho Jeong[3]
Hyunkyu Park[1] Hyeong Seok Kim[2] Juneho Yi[1]

[1]Department of Electrical and Computer Engineering, Sungkyunkwan University
[2]SK Planet Co., Ltd. [3]College of Computing, Sungkyunkwan University

YeongHyeon Park

ICASSP 2024 KOREA

Introduction

- Unsupervised anomaly detection (UAD) is a widely adopted approach in industry due to rare anomaly occurrences and data imbalance.
- A desirable characteristic of an UAD model is *contained generalization ability*.
  - Excels in the reconstruction of seen normal patterns
  - Struggles to reconstruct unseen anomaly patterns
- Reconstruction loss amplification is a simple way to achieve the contained generalization ability of an UAD model without altering the structure of the NNs or training strategy.

Loss landscape for contained generalization

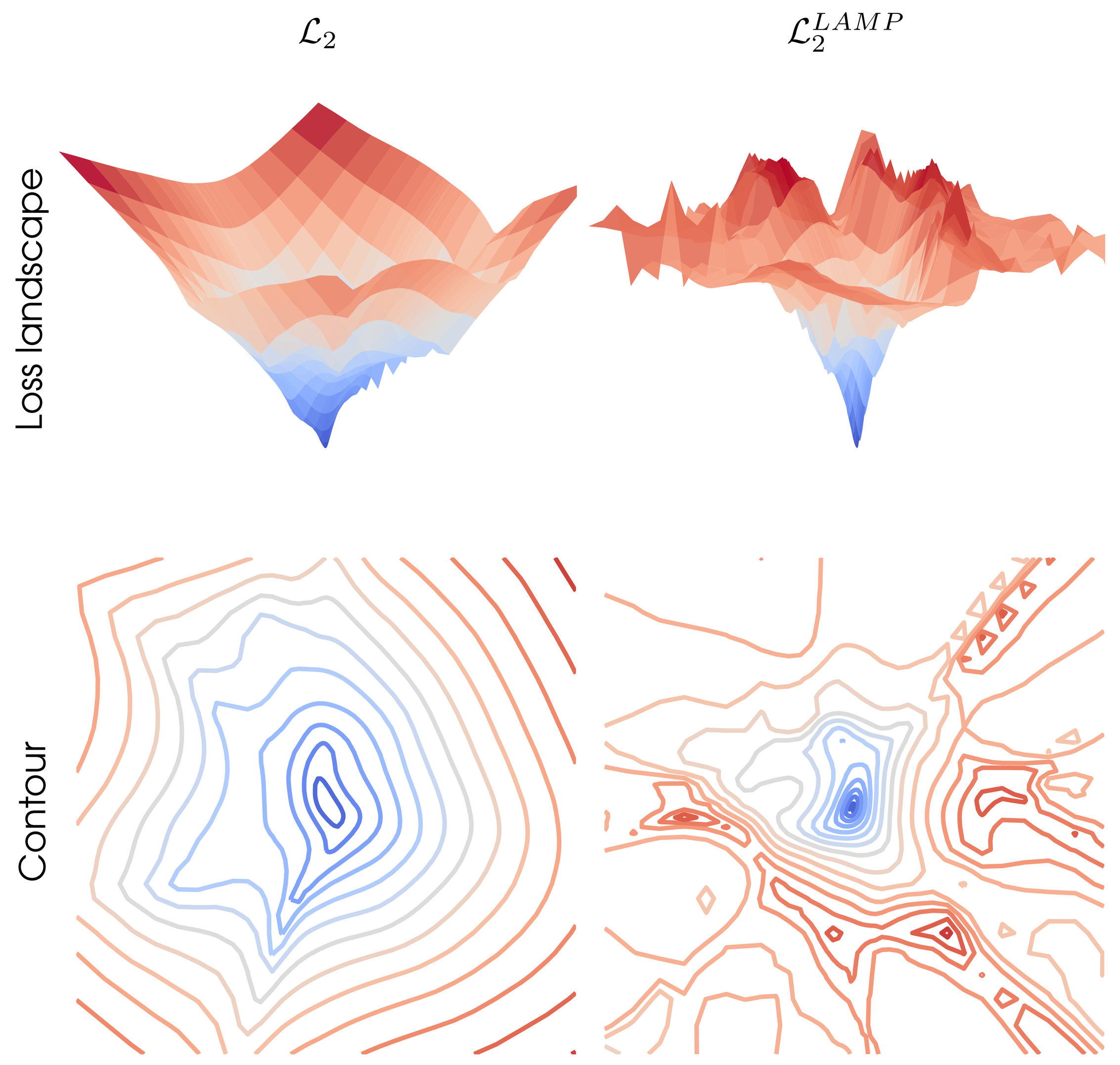

Figure 1: The loss landscapes and their contour projections for $\mathcal{L}_2$ and $\mathcal{L}_2^{LAMP}$. The loss landscape for an UAD model should be shaped with a sharp form in order to contain the reconstruction generalization ability.

- When the loss landscape is smooth, a reconstruction model has high generalization ability [1].
- *Loss AMPlification* (LAMP) can be easily and safely applied across any reconstruction error metrics because an UAD model is only trained using anomaly-free samples.
- Loss landscape sharpening method, LAMP, improves anomaly detection performance without any change of the NN architecture.

$$\mathcal{L}_{base}^{LAMP}(y,\hat{y}) = \sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} -\log\Big(1-\mathcal{L}_{base}(y_{h,w,c},\hat{y}_{h,w,c})\Big), \quad (1)$$

$$w.r.t.\ \ y \in \mathbb{R}^{H\times W\times C}$$

Changes in loss curves by LAMP

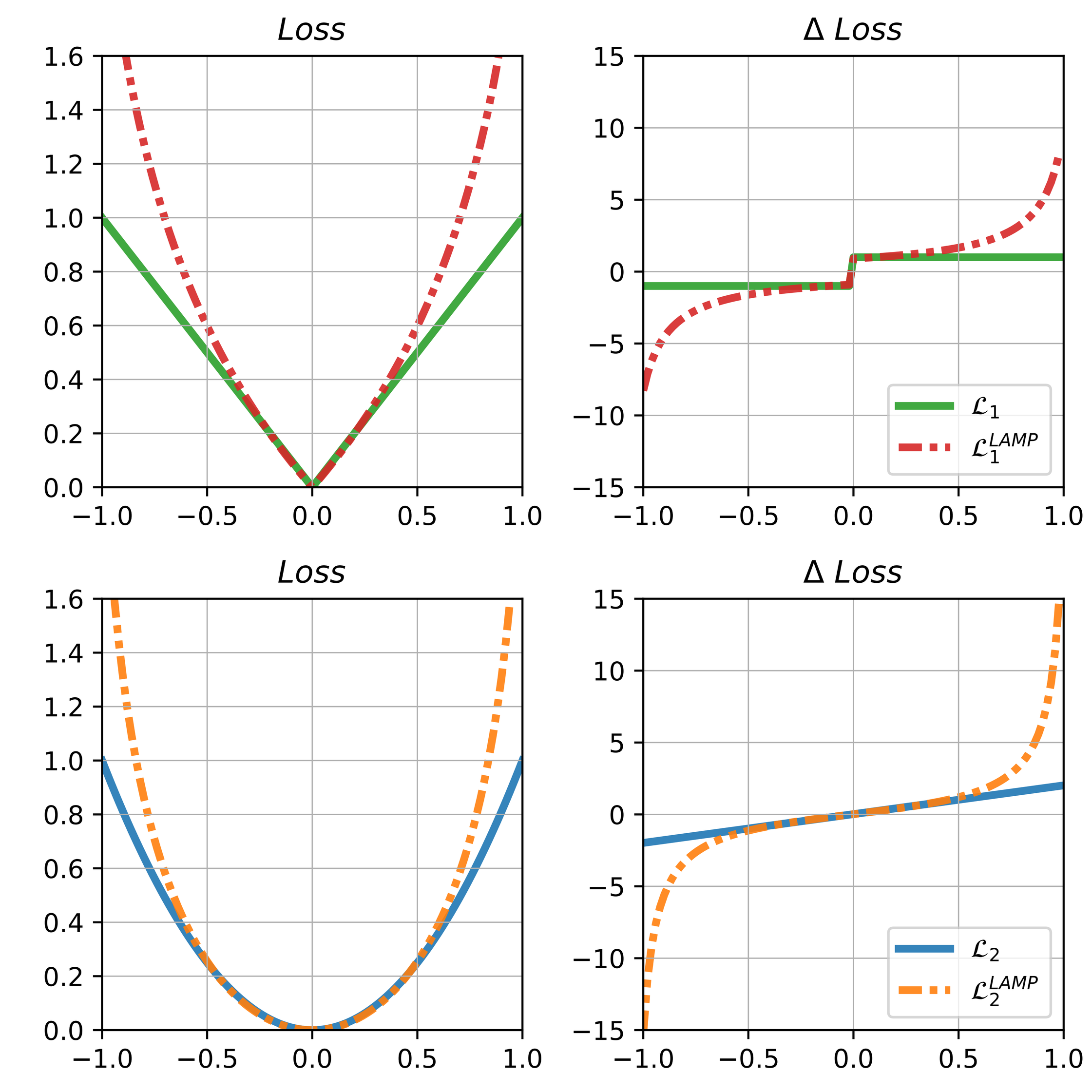

Figure 2: Loss curves for LAMP-applied $\mathcal{L}_1$ and $\mathcal{L}_2$ cases. LAMP makes gradients steeper than the base loss function, accelerating loss convergence, and transforms the loss landscape shape of an UAD model into a sharp form.

Comparison of loss landscapes

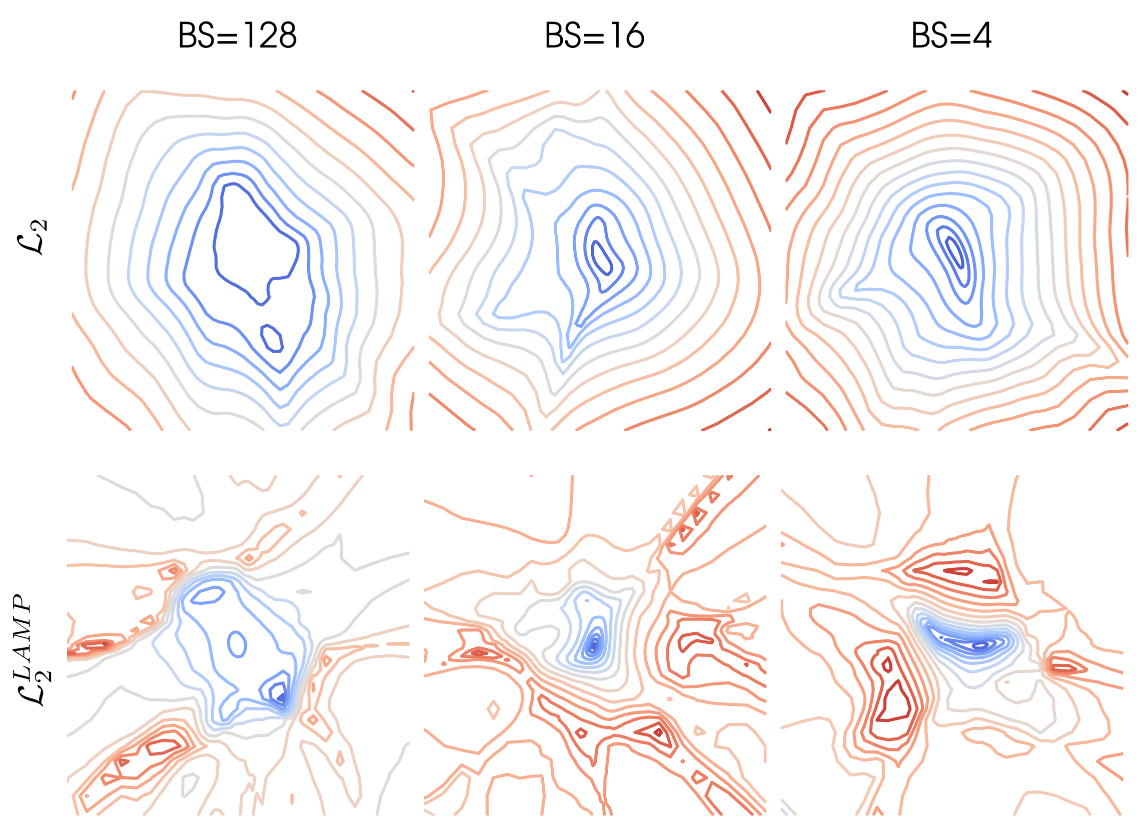

Figure 3: The contour of the loss landscapes with three batch size (BS) conditions. It is known that the generalization ability is contained when the BS is small [1].

Table 1: Summary of the AUROC for ten AD tasks using the MNIST dataset [2].

Loss	Batch size					
	1024	128	32	16	4	1
$\mathcal{L}_2$	0.658	0.919	0.921	0.926	0.931	0.919
$\mathcal{L}_2^{LAMP}$	**0.712**	**0.925**	**0.929**	**0.929**	**0.932**	**0.927**

Results for industrial dataset

Table 2: Summary of the AUROC for the MVTec AD dataset [3].

Training	$\mathcal{L}_2 \rightarrow \mathcal{L}_2^{LAMP}$			$\mathcal{L}_1 \rightarrow \mathcal{L}_1^{LAMP}$			$\mathcal{L}_{SSIM} \rightarrow \mathcal{L}_{SSIM}^{LAMP}$			Best
Optimizer	**SGD**	**RMSprop**	**Adam**	**SGD**	**RMSprop**	**Adam**	**SGD**	**RMSprop**	**Adam**	$\mathcal{L}_{base} \rightarrow \mathcal{L}_{base}^{LAMP}$
Bottle	0.987 → 0.983	0.990 → 0.990	0.993 → 0.991	0.989 → 0.992	0.993 → 0.994	0.994 → 0.992	0.983 → 0.980	0.994 → 0.994	0.994 → 0.993	**0.994** → **0.994**
Cable	0.806 → 0.813	0.832 → 0.830	0.817 → 0.812	0.823 → 0.790	0.832 → 0.830	0.835 → 0.823	0.728 → 0.755	0.798 → 0.781	0.811 → 0.792	**0.835** → 0.830
Capsule	0.816 → 0.791	0.782 → 0.800	0.810 → 0.775	0.764 → 0.799	0.757 → 0.811	0.801 → 0.816	0.801 → 0.801	0.798 → 0.786	0.793 → 0.825	0.816 → **0.825**
Hazelnut	0.980 → 0.981	0.974 → 0.993	0.965 → 0.974	0.982 → 0.981	0.984 → 0.988	0.972 → 0.983	0.894 → 0.938	0.956 → 0.959	0.947 → 0.951	0.984 → **0.993**
Metal nut	0.637 → 0.665	0.762 → 0.691	0.785 → 0.694	0.711 → 0.684	0.685 → 0.677	0.718 → 0.708	0.728 → 0.709	0.776 → 0.782	0.715 → 0.819	0.785 → **0.819**
Pill	0.810 → 0.803	0.864 → 0.864	0.860 → 0.885	0.856 → 0.845	0.867 → 0.874	0.834 → 0.836	0.824 → 0.827	0.857 → 0.832	0.837 → 0.830	0.867 → **0.885**
Screw	0.817 → 0.827	0.826 → 0.826	0.831 → 0.804	0.774 → 0.827	0.826 → 0.826	0.724 → 0.831	0.752 → 0.712	0.827 → 0.832	0.789 → 0.788	0.831 → **0.832**
Toothbrush	0.969 → 0.950	0.956 → 0.969	0.981 → 0.978	0.956 → 0.964	0.919 → 0.964	0.983 → 0.986	0.850 → 0.844	0.958 → 0.972	0.972 → 0.958	0.983 → **0.986**
Transistor	0.866 → 0.885	0.889 → 0.901	0.906 → 0.932	0.882 → 0.899	0.894 → 0.881	0.902 → 0.902	0.825 → 0.847	0.879 → 0.888	0.895 → 0.888	0.906 → **0.932**
Zipper	0.860 → 0.893	0.864 → 0.867	0.918 → 0.859	0.876 → 0.887	0.839 → 0.855	0.914 → 0.907	0.829 → 0.809	0.924 → 0.923	0.929 → 0.938	0.929 → **0.938**
Carpet	0.709 → 0.721	0.872 → 0.856	0.677 → 0.657	0.640 → 0.702	0.921 → 0.806	0.652 → 0.671	0.654 → 0.669	0.610 → 0.621	0.643 → 0.641	**0.921** → 0.856
Grid	0.791 → 0.787	0.868 → 0.888	0.920 → 0.894	0.758 → 0.722	0.859 → 0.868	0.869 → 0.904	0.652 → 0.651	0.895 → 0.825	0.880 → 0.833	**0.920** → 0.904
Leather	0.988 → 0.983	0.967 → 0.978	0.997 → 0.993	0.986 → 0.984	0.994 → 0.992	0.993 → 0.993	0.869 → 0.834	0.996 → 0.964	0.992 → 0.978	**0.997** → 0.993
Tile	0.562 → 0.697	0.836 → 0.911	0.658 → 0.670	0.576 → 0.651	0.811 → 0.802	0.712 → 0.620	0.601 → 0.609	0.847 → 0.785	0.744 → 0.714	0.847 → **0.911**
Wood	1.000 → 0.994	0.995 → 1.000	1.000 → 0.997	0.988 → 0.999	0.994 → 0.992	0.991 → 0.995	0.987 → 0.999	0.996 → 0.999	0.999 → 0.997	**1.000** → **1.000**
Average	0.840 → 0.851	0.885 → 0.891	0.874 → 0.861	0.837 → 0.848	0.878 → 0.877	0.860 → 0.864	0.798 → 0.799	0.874 → 0.863	0.863 → 0.863	0.908 → **0.913**

- The average AD performance is equal or greater when LAMP is applied in 5 out of 9 experimental settings.
  - Three base loss functions: $\mathcal{L}_2$, $\mathcal{L}_1$, and $\mathcal{L}_{SSIM}$
  - Three optimizers: SGD, RMSprop, and Adam
- The last column shows the best performance for each subtask and $\mathcal{L}_{base}^{LAMP}$ attains better AUROC than $\mathcal{L}_{base}$.

Reconstruction results

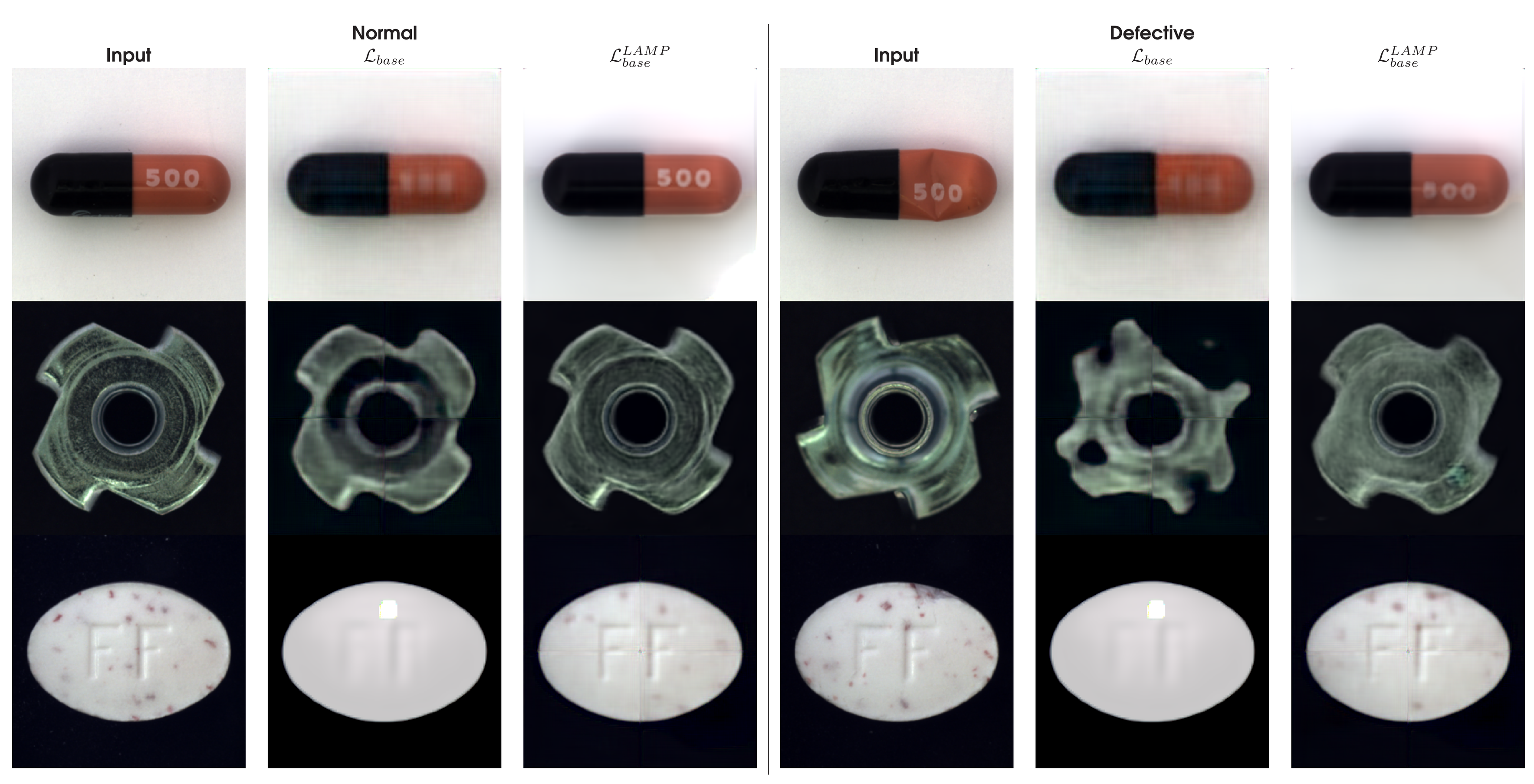

Figure 4: The $\mathcal{L}_{base}^{LAMP}$ case demonstrates improved reconstructions.

- $\mathcal{L}_{base}$ produces blurry results for normal products in capsule, metal nut, and pill cases.
- In contrast, $\mathcal{L}_{base}^{LAMP}$ case demonstrates accurate reconstructions for normal samples. Note the clear visibility of the number '500' on the normal capsule.

Conclusions

- We enhance the AD performance in an UAD setting from the perspective of reconstruction loss amplification by noting that contained generalization ability is highly related to sharp-shaped loss landscapes.
- Extensive experiments with MNIST and MVTec AD datasets demonstrate quantitative and qualitative performance enhancement of an UAD model by LAMP under various conditions.
- LAMP can be safely applied to any reconstruction error metrics in an UAD setup where a reconstruction model is trained with anomaly-free samples only.

References

[1] Hao Li, et al. "Visualizing the loss landscape of neural nets." NeurIPS. 2018

[2] Yann LeCun, et al, "Gradient-based learning applied to document recognition." *Proc. of the IEEE.* 1998

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