

多类别无监督异常检测中的极简主义设计：Dinomally 模型研究

汇报人：陈柏彤

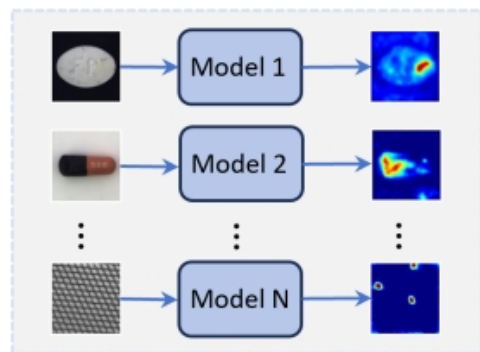
2026/06/17

文献：《Dinomally: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection》

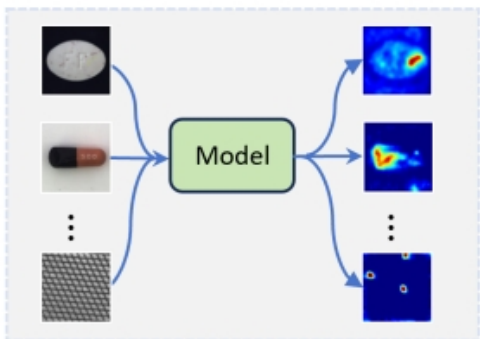
出处：CVPR 2025 (Open Access version, Computer Vision Foundation)

研究背景：从单类别走向多类别统一模型

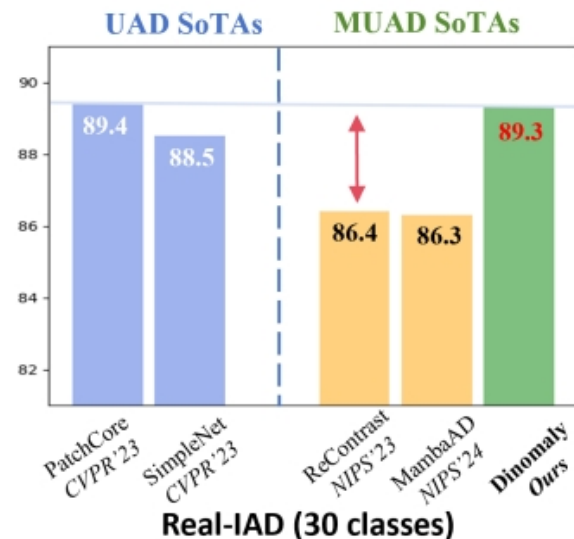
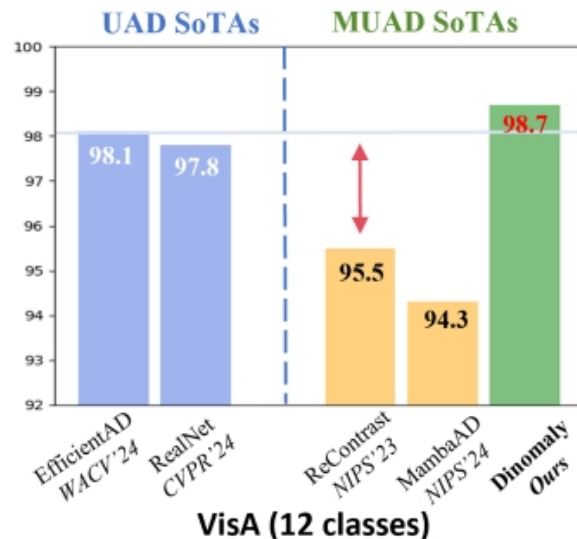
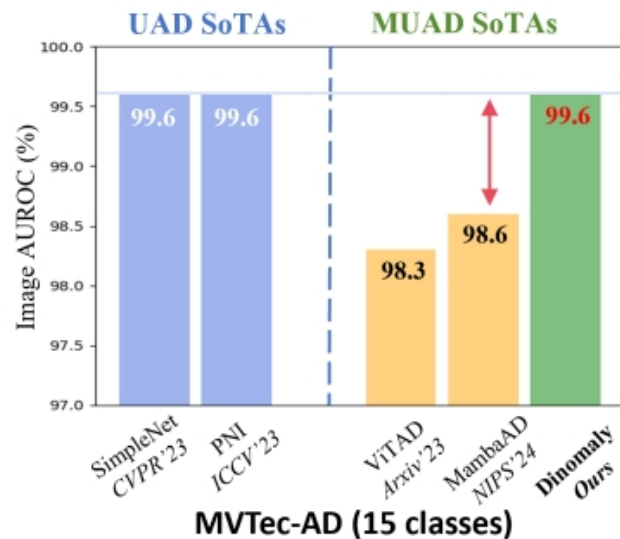
- **无监督异常检测 (UAD) 定义：** 仅利用正常样本进行模型训练，在推理阶段识别并定位图像中的异常区域。
- **传统 UAD 的局限：** 采用“一类一模型 (Class-Separated)”设定，当应用场景包含大量产品类别时，面临巨大的模型存储与维护开销。
- **多类别统一模型 (MUAD) 设定：** 仅使用单个统一模型处理包含多类不同物体的图像数据集。



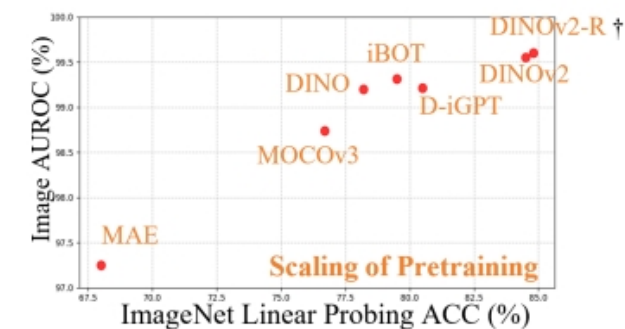
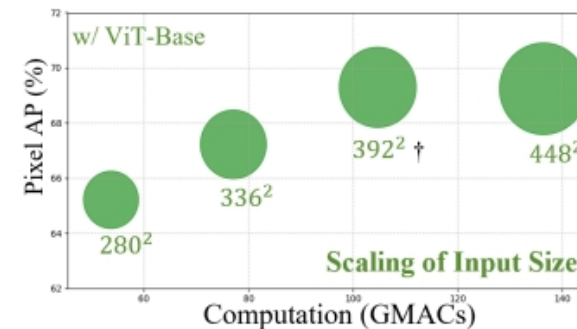
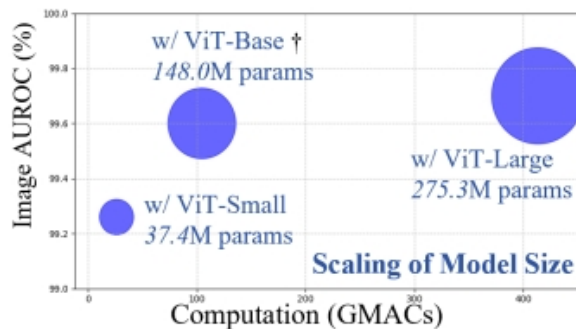
(a) **Class-Separated UAD Setting**



(b) **Model-Unified Multi-Class UAD Setting**



(c) **Image-Level AUROC on 3 Benchmarks**



(d) **Scaling Properties of Dinomally on MVTec-AD.** †: default setting.

Figure 1. Setting, benchmarking, and scaling of Dinomally. (a) Task setting of class-separated UAD. (b) Task setting of MUAD. (c) Comparison with previous SoTA methods on MVTec-AD [3], VisA [70], and Real-IAD [54]. (d) Scaling properties of Dinomally.

核心痛点：多类别设定下的“恒等映射”问题

- **现象根源：**多类别数据集内正常的纹理和结构特征极其多元化，促使网络产生极强的通用泛化能力。
- **恒等捷径：**由于网络泛化能力过强，解码器在推理时倾向于直接复制输入特征，导致不仅能完美重建正常区域，也能无损重建未见过的异常区域，最终引发异常检测失效。
- **现有方案的局限：**现有的 MUAD 方案（如合成伪异常、复杂的特征编辑或扩散模型）多依赖启发式、手工设计的技巧，系统复杂度高且泛化性弱。

Dinomally 极简重建框架概述

- **设计哲学：**“少即是多”，摒弃复杂的模块堆叠，完全由纯 Transformer 模块（Self-Attention 与 MLP）构建特征重建流水线。
- **系统基本拓扑：**包含自监督预训练编码器（Encoder）、中介瓶颈层（Bottleneck）和重建解码器（Decoder）。
- **工作流：**解码器通过最大化余弦相似度，学习恢复编码器的中层特征分布。

双击此处添加标题

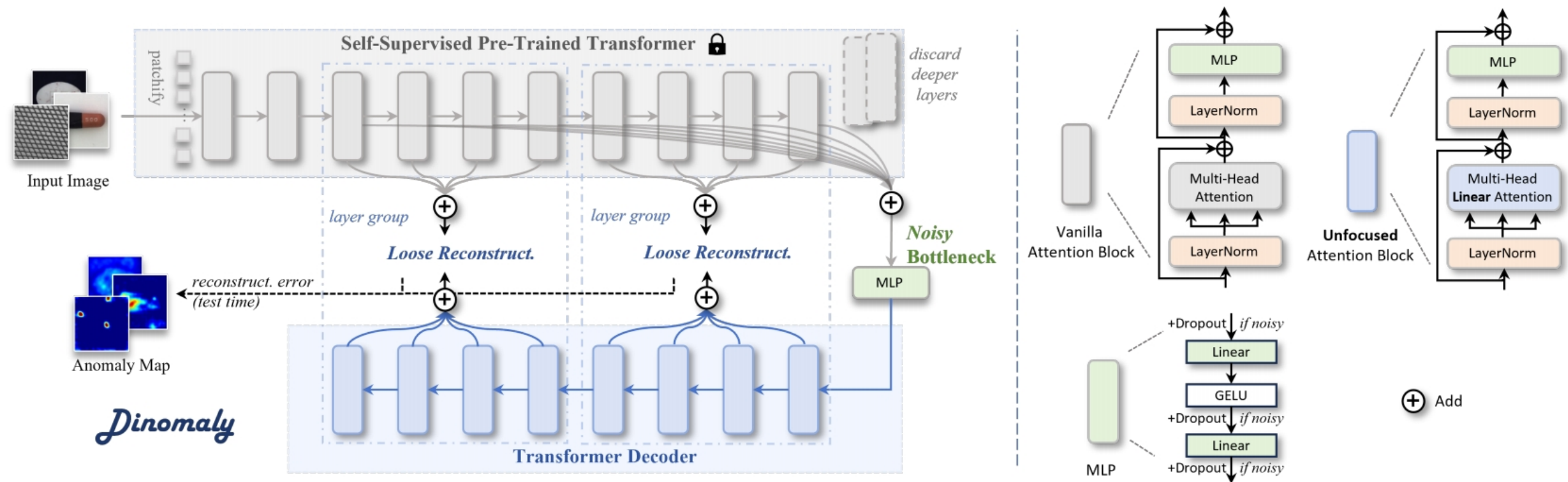


Figure 2. The framework of Dinomaly, built by simple and pure Transformer building blocks.

核心机制一：可扩展的基础 Transformer 编码器

- **特征源选型**：针对无监督场景，采用在海量数据上通过自监督策略预训练的 Vision Transformer (ViT)。
- **默认骨干网络**：综合考量精度与效率，默认采用基于 DINOv2-Register 预训练的 ViT-Base/14 架构。
- **主要发现**：实验证实，无监督异常检测性能与基础模型在 ImageNet 上的线性探测 (Linear Probing) 准确率呈现出强正相关性。

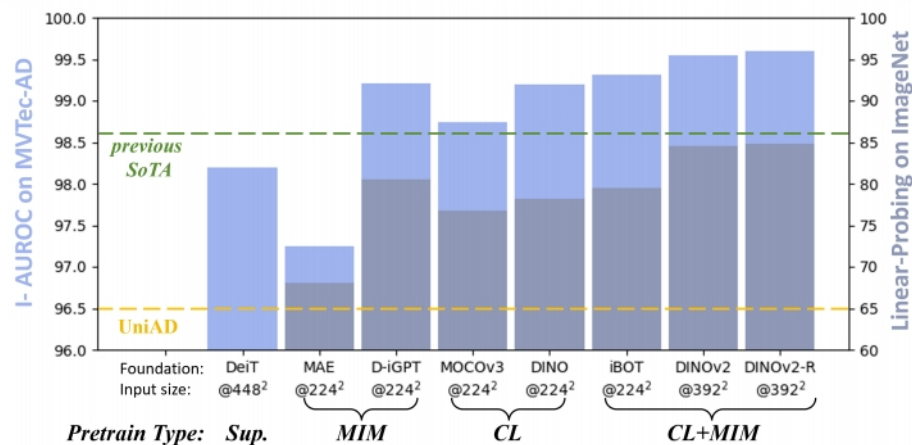


Figure 5. Image-level AUROC of Dinomaly equipped with various ViT foundations, and their linear-probing accuracy on ImageNet. MIM: Masked Image Modeling. CL: Contrastive Learning.

核心机制二：基于 Dropout 的噪声瓶颈

- **设计意图：** 将“特征重建”转换为“特征恢复（Restoration）”，打破恒等映射路径。
- **实现方式：** 摒弃手工设计特征扰动的传统做法，直接激活瓶颈层 MLP 中原有的 Dropout 组件（丢弃率设为 0.2-0.4）。
- **物理机制：** Dropout 通过随机丢弃神经元激活值，在线产生特征域的伪异常噪声（类似降噪自编码器），强迫解码器必须执行去噪恢复，使其丧失直接复制异常特征的能力。

双击此处添加标题

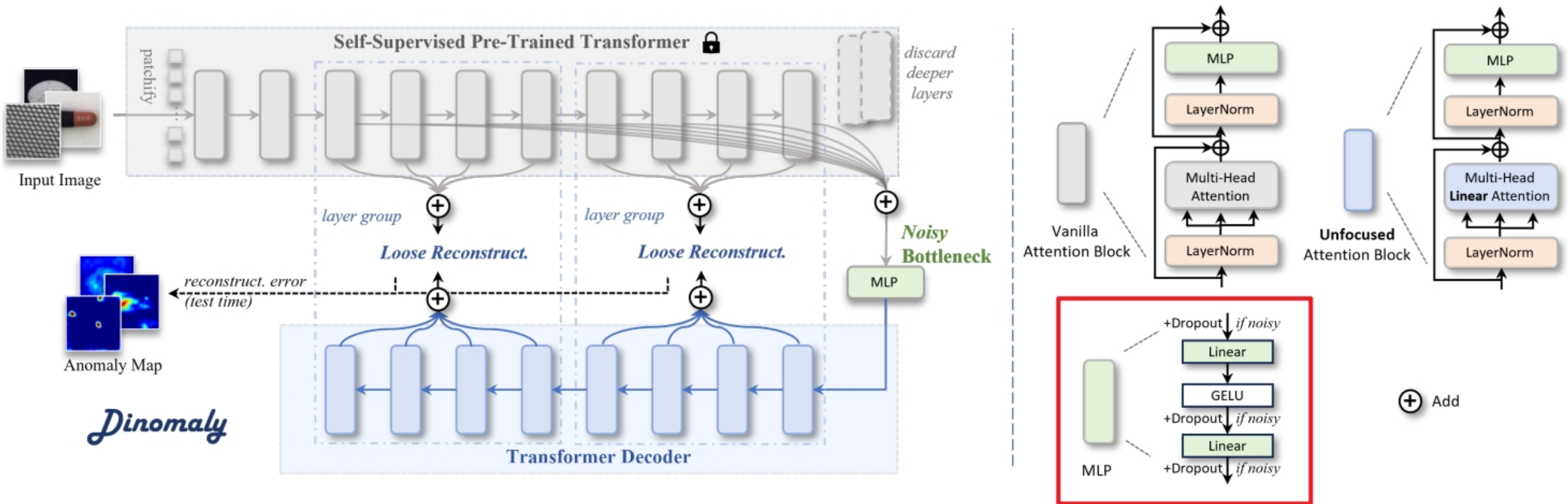
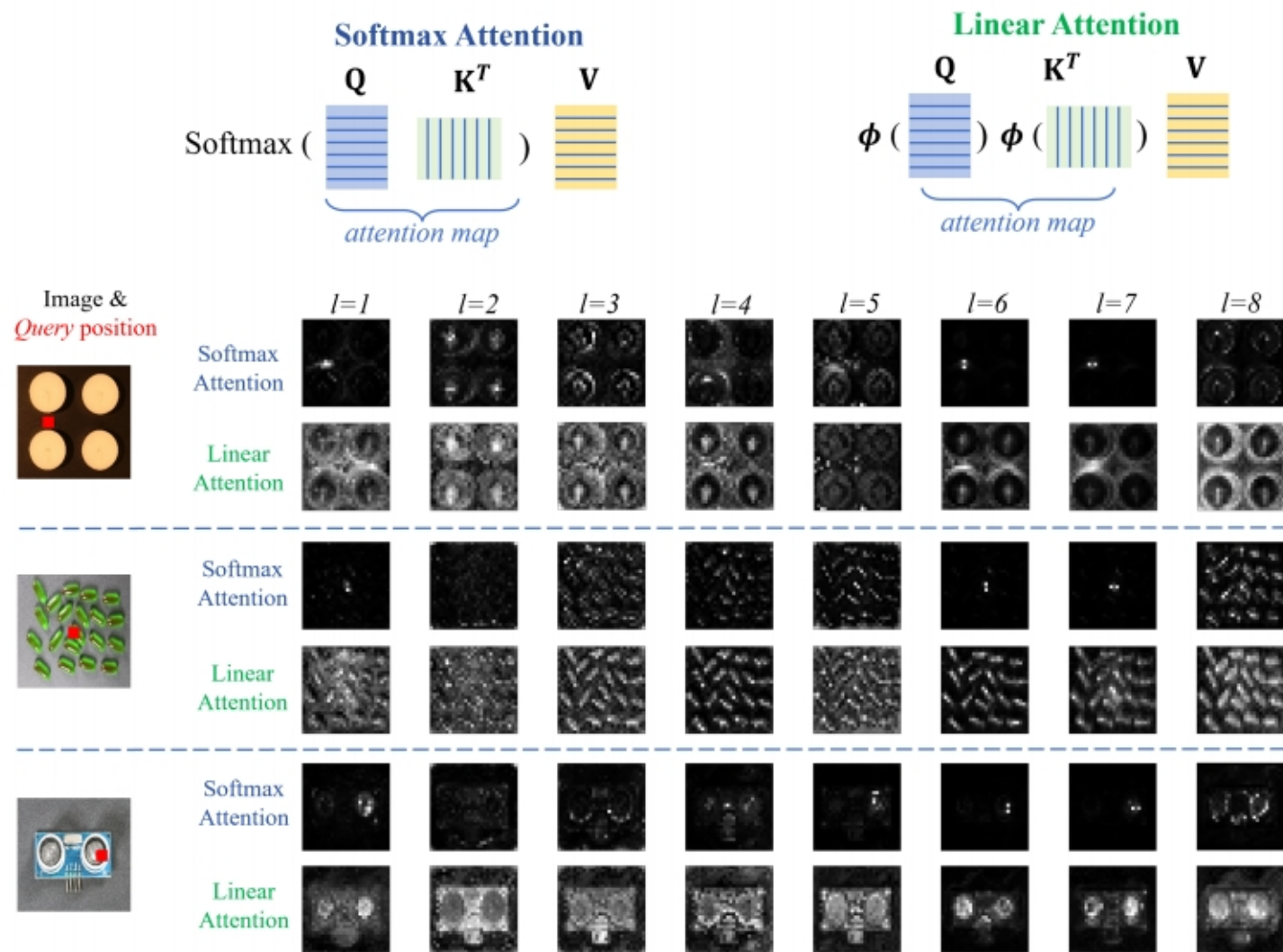


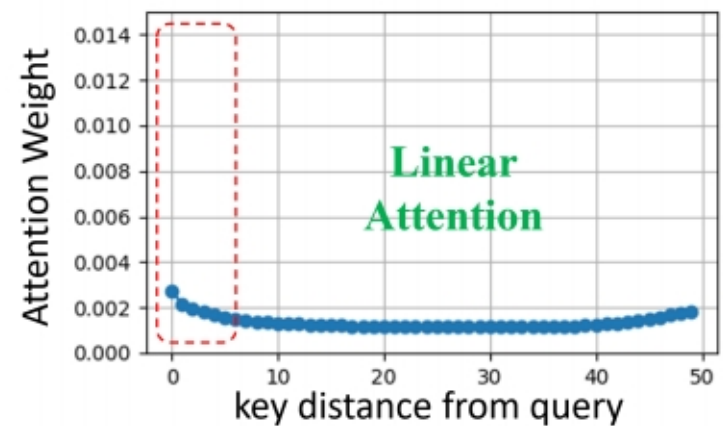
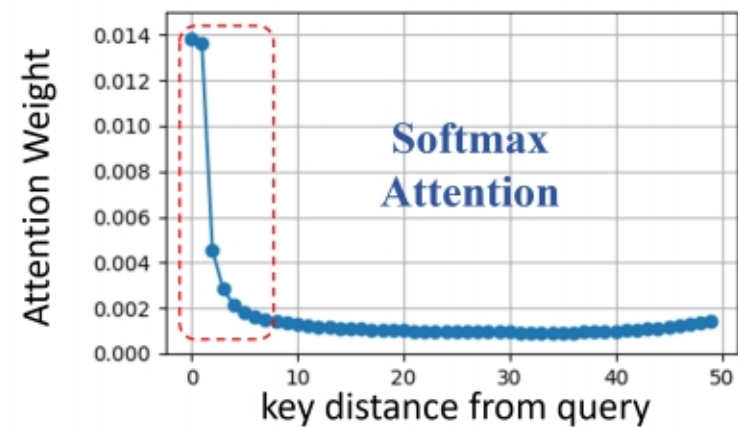
Figure 2. The framework of Dinomaly, built by simple and pure Transformer building blocks.

核心机制三：失焦的线性注意力机制

-



(a) Attention map of decoder layer l (mean of 12 heads)



(b) Attention distribution across query-key distances

Figure 3. Softmax Attention vs. Linear Attention. (a) Visualization of attention maps. (b) Attention distribution.

核心机制四：宽松重建约束与损失优化

- **层约束放宽 (Loose Constraint)**：传统方案采用严格的层对层逐一蒸馏，这使得解码器极易完美模仿编码器。Dinomaly 将感兴趣的多个连续层特征进行宏观求和和加总 (Group-to-group)，放宽层级对应约束，赋予解码器在面对未见过模式时更大的行为自由度。
- **点约束放宽 (Loose Loss)**：引入困难样本挖掘全局余弦损失 (Hard-mining global cosine loss)。在训练过程中，对重建已经非常完美的正常特征点执行梯度缩减 (收缩至 $1/10$)，驱使模型聚焦优化难以重建的复杂边界。

双击此处添加标题

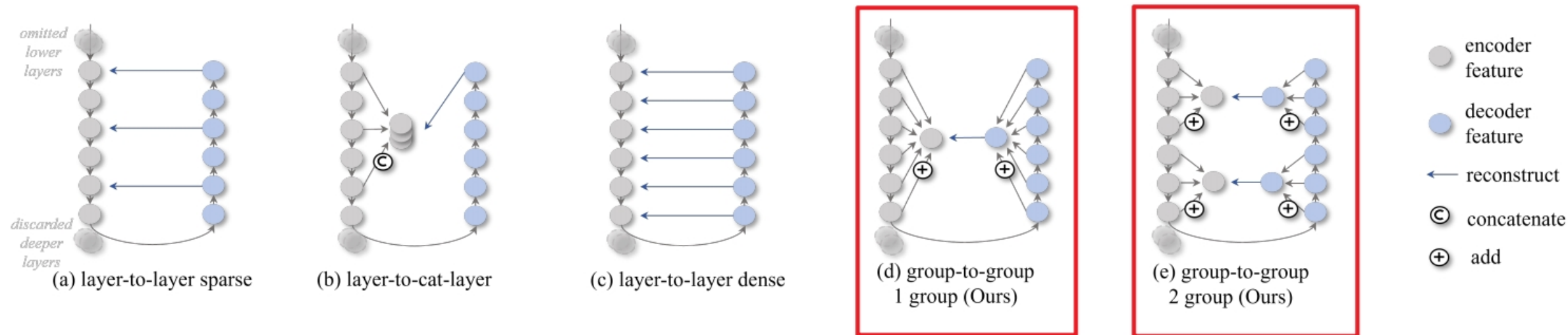


Figure 4. Schemes of reconstruction constraint. (a) Layer-to-layer (sparse). (b) Layer-to-cat-layer. (c) Layer-to-layer (dense). (d) Loose group-to-group, 1-group (Ours). (e) Loose group-to-group, 2-group (Ours).

实验部分 - 基准数据集性能评测

- **测试数据集：** 包含多类别统一工业异常检测三大权威基准：MVTec-AD (15类), VisA (12类), Real-IAD (30类、多视角)。
- **评估指标：** 覆盖图像级 (Image-level) 与像素级 (Pixel-level) 的 AUROC, AP, F1-max 及 AUPRO。
- **性能表现：** 在 MVTec-AD 上实现创纪录的 **99.6%** 图像级 AUROC。
 - 在最复杂的 Real-IAD 数据集上全面刷新多类别统一模型 (MUAD) 的性能上限。

Table 1. Performance under **multi-class** UAD setting (%). †: method designed for MUAD. Dinomaly†: training schedule is scaled up to 20,000, 20,000, and 100,000 iterations (original: 10,000/10,000/50,000).

	Dataset	Method	Image-level			Pixel-level			
			AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUPRO
双	MVTEC-AD [3]	RD4AD [10]	94.6	96.5	95.2	96.1	48.6	53.8	91.1
		SimpleNet [34]	95.3	98.4	95.8	96.9	45.9	49.7	86.5
		DeSTSeg [67]	89.2	95.5	91.6	93.1	54.3	50.9	64.8
		UniAD [60]†	96.5	98.8	96.2	96.8	43.4	49.5	90.7
		ReContrast [14]†	98.3	99.4	97.6	97.1	<u>60.2</u>	<u>61.5</u>	<u>93.2</u>
		DiAD [18]†	97.2	99.0	96.5	96.8	52.6	55.5	90.7
		ViTAD [65]†	98.3	99.4	97.3	<u>97.7</u>	55.3	58.7	91.4
		MambaAD [17]†	<u>98.6</u>	<u>99.6</u>	<u>97.8</u>	<u>97.7</u>	56.3	59.2	93.1
		Dinomaly (Ours)	99.6	99.8	99.0	98.4	69.3	69.2	94.8
		Dinomaly†	99.7	99.8	99.2	98.4	69.3	69.2	94.7
双	VisA [70]	RD4AD [10]	92.4	92.4	89.6	98.1	38.0	42.6	91.8
		SimpleNet [34]	87.2	87.0	81.8	96.8	34.7	37.8	81.4
		DeSTSeg [67]	88.9	89.0	85.2	96.1	39.6	43.4	67.4
		UniAD [60]†	88.8	90.8	85.8	98.3	33.7	39.0	85.5
		ReContrast [14]†	<u>95.5</u>	<u>96.4</u>	<u>92.0</u>	<u>98.5</u>	<u>47.9</u>	<u>50.6</u>	<u>91.9</u>
		DiAD [18]†	86.8	88.3	85.1	96.0	26.1	33.0	75.2
		ViTAD [65]†	90.5	91.7	86.3	98.2	36.6	41.1	85.1
		MambaAD [17]†	94.3	94.5	89.4	<u>98.5</u>	39.4	44.0	91.0
		Dinomaly (Ours)	98.7	98.9	96.2	98.7	53.2	55.7	94.5
		Dinomaly†	98.9	99.0	96.4	98.8	53.8	55.8	94.5
	Real-IAD [54]	RD4AD [10]	82.4	79.0	73.9	97.3	25.0	32.7	89.6
		SimpleNet [34]	57.2	53.4	61.5	75.7	2.8	6.5	39.0
		DeSTSeg [67]	82.3	79.2	73.2	94.6	<u>37.9</u>	<u>41.7</u>	40.6
		UniAD [60]†	83.0	80.9	74.3	97.3	21.1	29.2	86.7
		ReContrast [14]†	<u>86.4</u>	84.2	<u>77.4</u>	97.8	31.6	38.2	<u>91.8</u>
		DiAD [18]†	75.6	66.4	69.9	88.0	2.9	7.1	58.1
		ViTAD [65]†	82.7	80.2	73.7	97.2	24.3	32.3	84.8
		MambaAD [17]†	86.3	<u>84.6</u>	77.0	<u>98.5</u>	33.0	38.7	90.5
		Dinomaly (Ours)	89.3	86.8	80.2	98.8	42.8	47.1	93.9
		Dinomaly†	89.5	86.9	80.4	98.9	43.3	47.4	94.2

实验部分 - 软实力对决与可扩展性验证

- **逼近单类模型上限：**实验数据表明，多类别统一训练的 Dinomaly 性能不仅远超历史同设定模型，甚至近乎无损地逼近了那些为每个类别单独训练一个模型的顶尖一类一模型（如 PatchCore）。
- **模型的可扩展性 (Scaling Law)：**首次打破了无监督模型不遵循扩展定律的魔咒，证明随 ViT 骨干网络规模从 Small 扩展至 Base、Large，缺陷检测准确率稳步爬升。

Table 2. Performance under conventional **class-separated** UAD setting (%). n/a: not available.

Method	MVTec-AD [3]			VisA [70]			Real-IAD [54]		
	I-AUROC	P-AUROC	P-AUPRO	I-AUROC	P-AUROC	P-AUPRO	I-AUROC	P-AUROC	P-AUPRO
<i>Dinomaly (MUAD)</i>	99.6	98.4	94.8	98.7	98.7	94.5	89.3	98.8	93.9
Dinomaly	99.7	99.9	95.0	98.9	98.9	95.1	92.0	99.1	95.1
RD4AD [10]	98.5	97.8	<u>93.9</u>	96.0	90.1	70.9	87.1	n/a	<u>93.8</u>
PatchCore [45]	99.1	<u>98.1</u>	93.5	94.7	<u>98.5</u>	91.8	<u>89.4</u>	n/a	<u>91.5</u>
SimpleNet [34]	<u>99.6</u>	<u>98.1</u>	90.0	<u>97.1</u>	98.2	<u>90.7</u>	88.5	n/a	84.6

Table 4. Scaling of ViT model sizes on MVTec-AD (%). Im/s (Throughput, image per second) is measured on NVIDIA RTX3090 with batch size=16. Results on VisA and Real-IAD are shown in Table A3. †:default.

Arch.	Params	MACs	Im/s	Image-level			Pixel-level			
				AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUPRO
ViT-Small	37.4M	26.3G	153.6	99.26	99.67	98.72	98.07	68.29	67.78	94.36
ViT-Base†	148.0M	104.7G	58.1	<u>99.60</u>	<u>99.78</u>	<u>99.04</u>	<u>98.35</u>	<u>69.29</u>	<u>69.17</u>	<u>94.79</u>
ViT-Large	275.3M	413.5G	24.2	99.77	99.92	99.45	98.54	70.53	70.04	95.09

消融实验

Table 3. Ablations of Dinomaly elements on MVTec-AD (%). NB: Noisy Bottleneck. LA: Linear Attention. LC: Loose Constraint (2 groups). LL: Loose Loss. As MVTec-AD has reached saturation, we also present the results on VisA (Table A4).

NB	LA	LC	LL	Image-level			Pixel-level			
				AUROC	AP	F_1 -max	AUROC	AP	F_1 -max	AUPRO
✓	✓	✓	✓	98.41	99.09	97.41	97.18	62.96	63.82	92.95
				99.06	99.54	98.31	97.62	66.22	66.70	93.71
				98.54	99.21	97.62	97.20	62.94	63.73	93.09
				98.35	99.04	97.43	97.10	61.05	62.73	92.60
✓	✓	✓	✓	99.03	99.45	98.19	97.62	64.10	64.96	93.34
				99.27	99.62	98.63	97.85	67.36	67.33	94.16
				99.50	99.72	98.87	98.14	68.16	68.24	94.23
				99.52	<u>99.73</u>	98.92	<u>98.20</u>	<u>68.25</u>	<u>68.34</u>	94.17
✓	✓	✓	✓	<u>99.57</u>	99.78	<u>99.00</u>	<u>98.20</u>	67.93	68.21	<u>94.50</u>
✓	✓	✓	✓	99.60	99.78	99.04	98.35	69.29	69.17	94.79

消融实验分析

- 关键结论：

1. 噪声瓶颈（NB）与宽松损失（LL）能独立对模型带来直接增益。
2. 线性注意力（LA）与宽松重建约束（LC）在噪声瓶颈（NB）存在的前提下能引发强烈的协同催化效应，进一步大幅拉升检测精度。
3. 单独使用宽松约束（LC）会导致模型由于重建过于容易而引发性能下滑，证实了“加噪”与“放宽约束”软硬组合的必要性。

结论与展望

- **研究总结：** 本文成功证实了仅通过 Attention 与 MLP 的纯极简 Transformer 组合，在辅以合理的去噪、失焦与宽松约束后，能彻底攻克多类别 UAD 任务中的恒等映射难关。
- **主要贡献：** 打破了特征检测网络不遵循扩展定律的传统局限，实现了单统一模型在复杂、多门类工业场景中的高精度实时化部署可能。

数据类别 (Category)	模型与策略 (Model & Setting)	图像级 I-AUROC	图像级 I-AP	像素级 P-AUROC	像素级 P-AP	像素级 P-AUPRO
整体平均 (Mean)	Dinomally (Sep)	0.9946	0.9982	0.9822	0.5843	0.9357
	Dinomally (Uni)	0.9745	0.9889	0.9735	0.5629	0.9199
	ReContrast (Uni)	0.9347	0.9763	0.945	0.51	0.8882
地毯 (Carpet)	Dinomally (Sep)	1	1	0.991	0.6097	0.9561
	Dinomally (Uni)	0.9996	0.9999	0.9903	0.6516	0.9491
	ReContrast (Uni)	0.9992	0.9997	0.9896	0.6454	0.9613
网格 (Grid)	Dinomally (Sep)	1	1	0.9885	0.3904	0.9577
	Dinomally (Uni)	0.9983	0.9994	0.9843	0.2663	0.9411
	ReContrast (Uni)	0.9031	0.9701	0.8748	0.2812	0.7792
皮革 (Leather)	Dinomally (Sep)	1	1	0.99	0.3933	0.9675
	Dinomally (Uni)	1	1	0.9898	0.4495	0.9493
	ReContrast (Uni)	1	1	0.9927	0.5716	0.9833
瓷砖 (Tile)	Dinomally (Sep)	1	1	0.9685	0.6105	0.8819
	Dinomally (Uni)	1	1	0.9688	0.6279	0.8857
	ReContrast (Uni)	0.991	0.9967	0.9255	0.4795	0.8138
牙刷 (Toothbrush)	Dinomally (Sep)	0.9833	0.9934	0.9879	0.4615	0.9241
	Dinomally (Uni)	1	1	0.9895	0.5523	0.9217
	ReContrast (Uni)	0.9694	0.9885	0.986	0.5682	0.8904