Diverse Normal Prototypes-Guided Contrastive Reconstruction for Medical Anomaly Detection

Luhu Li[1], Bin Liu[2], Bowen Lin[2] (✉), Zihan Shen[1], Chengwei Wang[3], and Shujun Fu[1] (✉)

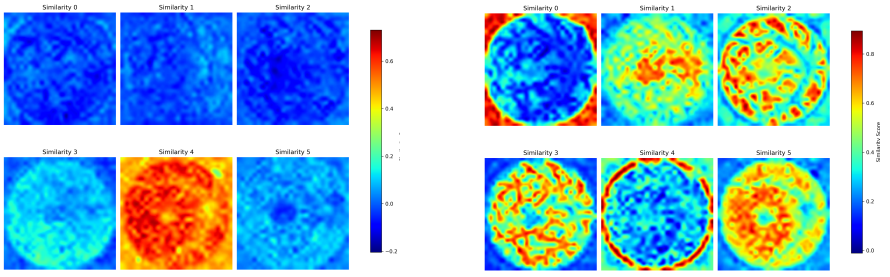
[1] School of Mathematics, Shandong University, Jinan 250100, China
shujunfu@163.com
[2] Department of Interventional Medicine and Minimally Invasive Oncology, The Second Qilu Hospital of Shandong University, Jinan 250033, China
linbowencore7@outlook.com
[3] Department of Neurosurgery, The Second Qilu Hospital of Shandong University, Jinan 250033, China

Abstract. Anomaly detection in medical images is challenging due to limited annotations and the domain gap. Existing reconstruction-based methods often rely on frozen pre-trained encoders, restricting adaptation to domain-specific patterns and degrading localization accuracy. Meanwhile, prototype-based learning offers interpretable representations but commonly suffers from prototype collapse, where a few prototypes dominate training and reduce diversity. To address these issues, we propose DNP-ConFormer, a unified framework that integrates a trainable encoder with prototype-guided reconstruction and a Diversity-Aware Alignment Loss. A momentum encoder enables stable domain-adaptive representation learning, while a lightweight Prototype Extractor discovers informative normal prototypes and injects them into the decoder via attention to guide reconstruction. The proposed alignment objective further encourages balanced feature-to-prototype assignments, effectively mitigating prototype collapse. Extensive experiments on multiple medical imaging benchmarks demonstrate improved representation quality and anomaly localization compared with prior methods. Visualization and prototype assignment analyses further validate the effectiveness and interpretability of our approach. The source code is available at https://github.com/liluhu0/DNP-ConFormer.

Keywords: Medical anomaly detection· Prototype collapse · Diversity-aware alignment· Contrastive domain adaptation

1 Introduction

Unsupervised anomaly detection (UAD) provides an alternative to supervised deep learning for medical image analysis, where obtaining pixel-level annotations is costly and anomalies are often rare [1,7,13,15,17]. By training only on normal images, UAD methods detect anomalies as deviations at inference time, which can improve scalability for tasks such as early screening [6].



(a) INP-Former (b) Ours

Fig. 1: Prototype visualizations on fundus image: (a) INP-Former with coherence loss suffers prototype collapse; (b) Our method with Diversity-Aware Alignment loss achieves diverse semantic assignments.

Among these approaches, reconstruction-based models assume anomalies yield higher reconstruction errors. However, image-level strategies often fail to capture fine-grained local anomalies. Recent teacher-student reverse distillation frameworks address this by training a student decoder to reconstruct multi-scale features from a frozen teacher encoder [6]. Representations produced by the teacher are further refined using contrastive learning (CL) [8, 9, 18]. Notably, self-supervised Vision Transformers (ViTs) provide richer semantic features for anomaly localization than CNNs or supervised ViTs [2, 5, 16]. Building on these advances, INP-Former [14] employs a self-supervised ViT teacher to extract Intrinsic Normal Prototypes (INPs) that guide student reconstruction via a coherence objective. While effective in industrial inspection, two challenges remain when applying such methods to medical images: (1) limited domain adaptability of the frozen encoder when the source (natural images) and target (medical images) domains differ substantially, and (2) prototype collapse under the coherence objective, where low-contrast medical features tend to converge to a dominant prototype, reducing the diversity of normal representations (Fig. 1a, 1b).

In this work, we propose **DNP-ConFormer** (Diverse Normal Prototypes-guided Contrastive Reconstruction), a dual-branch framework designed for domain-adaptive anomaly detection. Unlike reverse distillation approaches that rely on a frozen teacher encoder, our framework employs a momentum-updated teacher that evolves during training. The student branch combines a trainable encoder with a DNP-guided decoder, where Diverse Normal Prototypes (DNPs) extracted from encoder features are injected via cross-attention to guide reconstruction. The teacher branch provides relatively stable reference features through exponential moving average updates. Anomaly maps are derived from feature discrepancies between the two branches, which helps improve robustness under domain shifts. To mitigate prototype collapse, we consider prototype learning from a long-tail perspective by treating each prototype as a class center and patch-level features as samples. Based on this observation, we introduce a Diversity-Aware

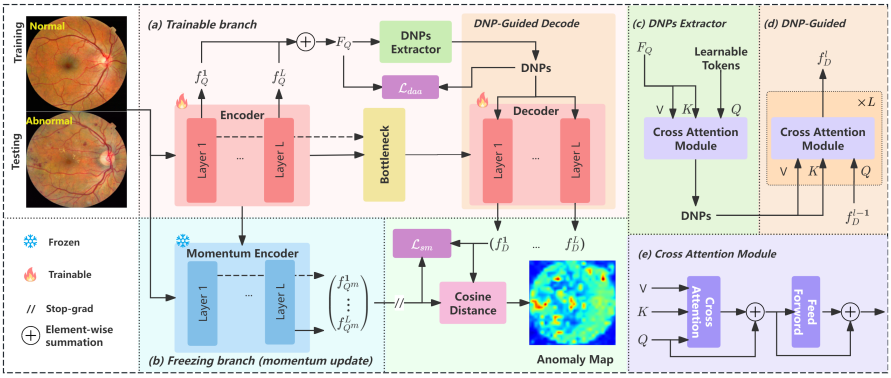


Fig. 2: Overview of the proposed DNP-ConFormer. (a) The trainable student branch. (b) The momentum teacher branch. (c) The DNP extraction module. (d) The DNP-guided decoding module. (e) The cross-attention module.

Alignment (DAA) loss that encourages balanced feature–prototype assignments while maintaining inter-prototype distinctiveness.

2 Method

We present the overall pipeline of **DNP-ConFormer**, illustrated in Fig. 2. The framework consists of a **student branch** for representation learning and reconstruction, and a **teacher branch** that provides stable reference features via momentum updating. Given an input image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, the student encoder $\mathcal{Q}$ extracts multi-scale features $f_{\mathcal{Q}} = \{f_{\mathcal{Q}}^1, \ldots, f_{\mathcal{Q}}^L\}$ with $f_{\mathcal{Q}}^l \in \mathbb{R}^{N\times C}$, which are aggregated into a unified representation $\mathbf{F}_{\mathcal{Q}} = \frac{1}{L}\sum_{l=1}^{L} f_{\mathcal{Q}}^l$. To capture diverse normal patterns, we introduce a set of M learnable tokens $\{t_1, \ldots, t_M\}$ and employ a cross-attention module to extract **Diverse Normal Prototypes** $\mathbf{P} = \mathcal{E}(\mathbf{F}_{\mathcal{Q}}, \{t_m\})$, where $\mathbf{P} \in \mathbb{R}^{M\times C}$ represents different modes of normality in the feature space. These prototypes are injected into the DNP-guided decoder $\mathcal{D}$ via cross-attention with the bottleneck representation $F_{\mathcal{B}} = \mathcal{B}(f_{\mathcal{Q}})$, where $\mathcal{B}$ denotes a bottleneck aggregation module, enabling the decoder to reconstruct feature representations aligned with normal patterns. The **teacher branch** maintains a momentum encoder $\mathcal{Q}^m$ whose parameters $\theta_{\mathcal{Q}^m}$ are updated as an exponential moving average (EMA) of the student encoder parameters $\theta_{\mathcal{Q}}$, producing temporally smoothed and stable reference features. During training, we compute the cosine distance between student and teacher features to measure representation discrepancy, which is used to generate the anomaly map at the patch level. The overall training objective consists of two components: a **soft-mining loss** $\mathcal{L}_{sm}$ that enforces consistency between the student and teacher representations while emphasizing informative samples, and a **Diversity-Aware Alignment loss** $\mathcal{L}_{daa}$ that encourages balanced and discriminative prototype assignments.

2.1 Encoder Adaptation Frameworks

Frozen natural-image encoders $\mathcal{Q}$ lack adaptability to medical imaging data, leading to domain inflexibility, representation misalignment, and uncorrectable reconstruction errors. A direct approach is to unfreeze $\mathcal{Q}$ and jointly optimize it with the decoder $\mathcal{D}$ (**M1** in Fig. 3). However, with limited medical data, end-to-end fine-tuning often causes gradient instability, overfitting, and catastrophic forgetting of useful pretrained priors. Inspired by SimSiam [3], we adopt an asymmetric twin-path framework (**M2** in Fig. 3) for stable adaptation, comprising a trainable online path ($x \to \mathcal{Q} \to \mathcal{B} \to \mathcal{D} \Rightarrow \hat{f}_{\mathcal{D}}$) and a stable frozen target path ($x \to \mathcal{Q}^{\text{frz}} \Rightarrow f_{\mathcal{Q}^{\text{frz}}}$). This asymmetric design mitigates representation collapse and enables gradual alignment between the online and target representations via the following objective:

$$\mathcal{L}_{sm}^{\text{M2}} = 1 - \cos\left(\text{sg}(f_{\mathcal{Q}^{\text{frz}}}), \hat{f}_{\mathcal{D}}\right). \tag{1}$$

Nevertheless, the statically frozen target encoder cannot evolve with the online network, which may limit adaptation capacity and lead to performance saturation. To overcome this limitation, we further propose $\mathbf{M2}^{+}$ (Fig. 3), where the target encoder is replaced with a momentum encoder $\mathcal{Q}^{m}$ updated using an exponential moving average:

$$\theta_{\mathcal{Q}^m} \leftarrow \beta\theta_{\mathcal{Q}^m} + (1-\beta)\theta_{\mathcal{Q}}. \tag{2}$$

This design provides a slowly evolving target representation that improves temporal consistency, preserves pretrained knowledge, and stabilizes optimization, thereby enabling stable convergence across diverse medical datasets as illustrated in Fig. 3. Accordingly, we adopt M2^{+} as our final design, and the training objective is defined as

$$\mathcal{L}_{sm}^{\text{M2}^{+}} = 1 - \cos\left(\text{sg}(f_{\mathcal{Q}^m}), \hat{f}_{\mathcal{D}}\right). \tag{3}$$

2.2 Diversity-Aware Normal Prototypes

Prototype collapse. The baseline coherence objective often leads to single-prototype dominance, where most patches are assigned to one prototype while the others remain underutilized, thereby limiting the representational capacity of the model. This issue is particularly pronounced in medical images due to their low contrast and limited texture variability. The collapse arises because the consistency objective enforces feature alignment but does not impose any constraint on prototype diversity; consequently, early advantages gained by certain prototypes become self-reinforcing, resulting in imbalanced prototype utilization.

Diversity-Aware Alignment Loss: A Long-Tail Learning Perspective. From a long-tail learning perspective, prototype collapse resembles class-imbalance behavior, where a small subset of prototypes dominates feature assignments. Existing consistency objectives focus on patch-level alignment but overlook prototype-level distribution imbalance. To address this issue, we encourage

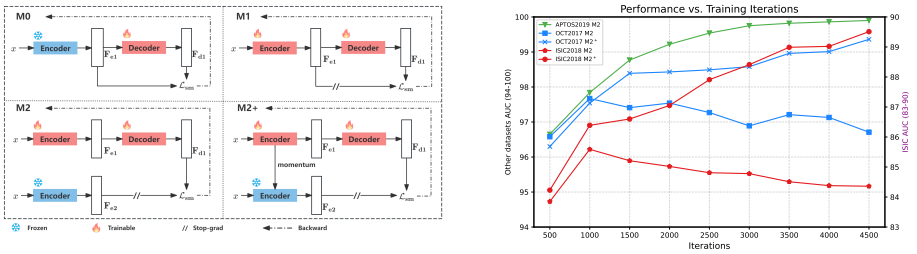


Fig. 3: (a) Model variants (M0 to M2$^+$): Progressive design for domain adaptation in anomaly detection. (b) Performance trends over training iterations on different datasets. Under the M2 framework, performance steadily improves on APTOS2019, while OCT2017 peaks at 1000 iterations before declining. With the M2$^+$ framework, OCT2017 also shows continuous improvement.

balanced participation across prototypes by promoting diverse and informative feature alignment, leading to the proposed **Diversity-Aware Alignment Loss**. For each prototype p_j, we define the assigned feature set $\mathcal{M}_j$ as

$$\mathcal{M}_j = \left\{ i \in \{1, \ldots, N\} \;\middle|\; \arg\min_m \mathcal{S}(\mathbf{F}_{\mathcal{Q}}(i), p_m) = j \right\}, \tag{4}$$

where $\mathcal{S}(\cdot, \cdot)$ denotes cosine distance. Each patch feature is assigned to its nearest prototype, and a distance-based loss is computed within each prototype group. To ensure robustness, we only include prototypes that have been assigned at least one feature. The Diversity-Aware Alignment Loss is defined as

$$\mathcal{L}_{\text{daa}} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}_{|\mathcal{M}_j|>0} \cdot \frac{1}{|\mathcal{M}_j|} \sum_{i \in \mathcal{M}_j} \mathcal{S}(\mathbf{F}_{\mathcal{Q}}(i), p_j). \tag{5}$$

This formulation promotes a more balanced assignment distribution across prototypes, analogous to maximizing entropy in prototype utilization, thereby mitigating collapse. The indicator function $\mathbf{1}_{|\mathcal{M}_j|>0}$ ensures that only non-empty prototype groups contribute to the loss, avoiding invalid gradients. The final objective is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{sm}^{\text{M2}^+} + \lambda \mathcal{L}_{daa}. \tag{6}$$

3 Experiments and Results

3.1 Experimental Setup

Datasets. We evaluate our method on three public benchmarks and one in-house clinical collection: **OCT2017** [12] , **APTOS2019** [11], **ISIC2018** [4], and an in-house brain MRI dataset (**BrainMRI**). All images are resized to 224×224.

Metrics. We evaluate performance using image-level and pixel-level metrics. Image-level performance is measured by AUC, F1-score, and accuracy (ACC), with AUC as the primary, threshold-independent metric and F1/ACC computed at the F1-optimal threshold, while pixel-level evaluation on the BrainMRI dataset reports AUC, Average Precision (AP), pixel-level F1, and Area Under the Per-Region Overlap curve (AUPRO) for anomaly localization.

Implementation details. We employ ViT-Small/14 with DINOv2-R [5] weights as the teacher and an 8-layer Transformer as the student decoder. The bottleneck module consists of a two-layer MLP. We further report the backbones used by the comparison methods. Specifically, ViT-based methods (INP-Former [14], Dinomaly [10]) use ViT-Small/14, and convolution-based methods are evaluated with WideResNet50 (RD4AD [6], ReContrast [8]) and ResNet50 (EDC [9], EA2D [18]). Models are trained on a single NVIDIA RTX 3090 using StableAdamW [19] with a batch size of 32. Learning rates are set to 10^{-4} for the decoder, bottleneck, and DNPs extractor, and 10^{-5} for the encoder. Training iterations vary by dataset: 8k (OCT2017), 5k (APTOS, BrainMRI), and 4k (ISIC2018). Hyperparameters are fixed at $M = 6$, $\beta = 0.999$, and $\lambda = 0.2$. Results are averaged over five runs.

3.2 Main Result

We evaluate DNP-ConFormer against state-of-the-art (SOTA) methods on three public benchmarks (APTOS2019, OCT2017, ISIC2018) and an in-house BrainMRI dataset. All experimental results are averaged over five independent runs. Table 1 and Table 2 report the comparative results, where **bold** and underlined indicate the best and second-best performance, respectively. Quantitative results and qualitative visualizations demonstrate the superiority and robustness of our model across diverse modalities.

Performance Comparison. As shown in Table 1, DNP-ConFormer demonstrates superior performance, achieving the top-ranked AUC, F1-score, and ACC across all public benchmarks, outperforming the runner-up Recontrast on APTOS2019 and reaching near-saturated results on OCT2017. For the challenging ISIC2018, our method significantly exceeds competitors, confirming its strong generalization capability. On the clinical BrainMRI dataset (Table 2), DNP-ConFormer demonstrates exceptional robustness. It leads in image-level detection (92.52% AUC, 87.10% F1-score, and 85.39% ACC) over Dinomaly and significantly outperforms INP-Former in pixel-level localization across all metrics, including 98.53% AUC, 46.03% AP, 51.10% F1-score, and 89.27% AUPRO. These results highlight the model's effectiveness in real-world clinical scenarios.

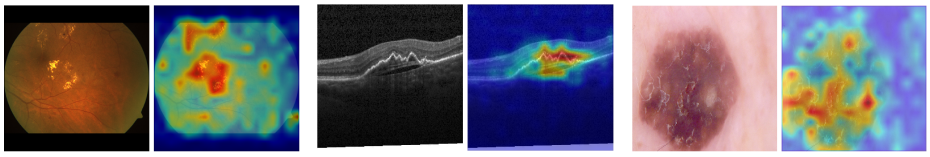
Qualitative Analysis. To assess localization precision, Fig. 4 and Fig. 5 visualize anomaly maps across various modalities. Despite structural complexities, DNP-ConFormer consistently generates focused heatmaps that closely align with pathological regions while suppressing background noise. Particularly in BrainMRI, the high-response regions show a high degree of overlap with pixel-level ground truth, confirming the model's ability to learn clinically meaningful representations.

Table 1: Performance comparison on public benchmarks.

Dataset →	APTOS2019			OCT2017			ISIC2018		
Method ↓	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC
RD4AD [6]	94.40	92.06	88.88	99.29	97.61	96.40	76.76	70.47	67.43
EDC [9]	95.70	93.22	90.61	99.59	98.67	98.00	90.53	81.94	86.53
INP-Former [14]	96.00	93.39	90.65	97.82	96.21	94.30	90.75	80.50	83.94
EA2D [18]	96.83	94.45	92.25	99.82	98.86	98.28	88.48	79.29	84.04
Dinomaly [10]	95.87	93.05	90.57	98.62	96.99	95.50	89.31	80.48	83.42
Recontrast [8]	97.61	95.22	93.28	99.57	98.39	97.60	90.70	82.16	86.70
Ours	**97.92**	**95.69**	**93.91**	**99.83**	**99.13**	**98.70**	**91.73**	**82.61**	**87.05**

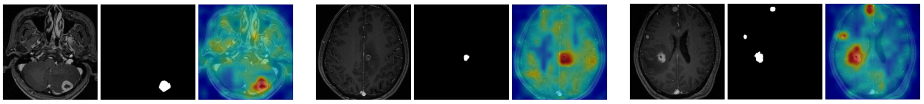
Table 2: Performance comparison on the in-house BrainMRI dataset.

Method	Image-level			Pixel-level			
	AUC	F1	ACC	AUC	AP	F1	AUPRO
RD4AD [6]	73.33	75.93	67.03	89.77	5.34	11.08	62.18
EDC [9]	88.10	83.25	80.03	96.63	27.12	33.95	84.75
EA2D [18]	78.68	77.54	71.14	89.30	6.23	12.36	62.44
Recontrast [8]	79.06	76.97	70.73	96.93	25.88	33.72	84.69
Dinomaly [10]	91.81	86.22	84.31	97.46	38.66	44.32	84.28
INP-Former [14]	91.17	85.41	83.64	98.25	43.12	49.05	87.92
Ours	**92.52**	**87.10**	**85.39**	**98.53**	**46.03**	**51.10**	**89.27**



(a) APTOS2019 (b) OCT2017 (c) ISIC2018

Fig. 4: Visualization of anomaly maps on three public benchmarks.



(a) (b) (c)

Fig. 5: Visualization of anomaly maps on the in-house BrainMRI dataset.

3.3 Ablation Study

To evaluate the contribution of each component, we conduct component ablation experiments focusing on encoder adaptation strategies (M0, M1, M2 and M2$^+$),

Table 3: Ablation study on encoder adaptation.

Variant	OCT2017		ISIC2018	
	AUC	Δ_{M0}	AUC	Δ_{M0}
M0 (Baseline)	98.06	–	88.46	–
M1 (+Unfrz.)	60.09	$-37.97\downarrow$	83.80	$-4.66\downarrow$
M2 (+CL)	97.41	$-0.65\downarrow$	85.21	$-3.25\downarrow$
Ours (+Mom.)	99.83	$+1.77\uparrow$	91.73	$+3.27\uparrow$

Table 4: Ablation study on momentum coefficient β.

Momentum β	0.99	0.999	0.9999	1
Last AUC	85.44	90.00	88.96	84.58
Best AUC	88.00	90.91	89.02	84.71

Table 5: Effect of diversity-aware alignment loss $\mathcal{L}_{\text{daa}}$.

Strategy	APTOS2019		BrainMRI	
	AUC	Δ	AUC	Δ
w $\mathcal{L}_{\text{daa}}$	95.97	–	92.52	–
w/o $\mathcal{L}_{\text{daa}}$	95.24	$-0.73\downarrow$	91.46	$-1.06\downarrow$

sensitivity to the momentum coefficient, and Diversity-Aware Alignment loss. As summarized in Table 3–Table 5, our complete model consistently outperforms all variants across public benchmarks and the clinical BrainMRI dataset.

Starting with **Encoder Adaptation**, a frozen encoder (M0) provides a robust baseline, whereas direct fine-tuning (M1) leads to a catastrophic performance drop ($\Delta = -37.97\%$ AUC on OCT2017), indicating severe overfitting on limited anomalous samples. While the addition of contrastive loss (M2) partially mitigates this, our momentum-based strategy (**Ours**) effectively balances domain adaptation and stability, improving AUC by 1.77% and 3.27% on OCT2017 and ISIC2018, respectively. Regarding the **Momentum Coefficient** β, we find that $\beta = 0.999$ yields the optimal trade-off, achieving the highest best-epoch AUC (90.91%). Lower values (e.g., 0.99) induce training instability, shown by the larger gap between last and best AUC, while a static target ($\beta = 1$) significantly limits adaptation capacity. Furthermore, removing the **Diversity-Aware Alignment Loss ($\mathcal{L}_{\text{daa}}$)** results in a clear performance degradation (-0.73% on APTOS2019 and -1.06% on BrainMRI), confirming its critical role in preventing prototype collapse and ensuring the learning of discriminative features across clinical scenarios.

The ablation results collectively validate that the performance gains of DNP-ConFormer stem from the synergy between its core components. While the momentum-based encoder adaptation provides a stable foundation for domain-

specific feature extraction, $\mathcal{L}_{\text{daa}}$ ensures that these features remain diverse and discriminative. Together, these mechanisms ensure precise and reliable anomaly detection across a wide range of medical imaging modalities.

4 Conclusion

The robust performance of DNP-ConFormer across diverse datasets establishes its potential as a unified framework for medical anomaly analysis. Its stability on both structured (e.g., OCT, MRI) and unstructured (e.g., fundus, dermoscopic) modalities confirms that the DNPs effectively preserve rich local semantics and preclude prototype collapse, while the dual-branch comparative reconstruction facilitates powerful global representation. Furthermore, the superior results on the in-house BrainMRI dataset underscore its reliability for deployment in real-world clinical settings. Overall, DNP-ConFormer provides a robust and generalizable framework, offering a high-performance alternative for diverse medical anomaly analysis tasks.

Acknowledgments. The research has been supported in part by the National Natural Science Foundation of China (NSFC) (Grants Nos. 12426310, 12571534, 62571299, 12371492, and 12301543), and the Shandong Provincial Natural Science Foundation (Grant Nos. ZR2025MS70 and ZR2023QA041).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint (2021), arXiv:2107.02314
2. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9630–9640 (2021). https://doi.org/10.1109/ICCV48922.2021.00951
3. Chen, X., He, K.: Exploring simple siamese representation learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15745–15753 (2021). https://doi.org/10.1109/CVPR46437.2021.01549
4. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint (2019), arXiv:1902.03368
5. Darcet, T., Oquab, M., Mairal, J., Bojanowski, P.: Vision transformers need registers. arXiv preprint (2024), arXiv:2309.16588
6. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9727–9736 (2022). https://doi.org/10.1109/CVPR52688.2022.00951

7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations. pp. 1–21 (2021)
8. Guo, J., Lu, S., Jia, L., Zhang, W., Li, H.: Recontrast: Domain-specific anomaly detection via contrastive reconstruction. In: Advances in Neural Information Processing Systems. vol. 36, pp. 10721–10740 (2023)
9. Guo, J., Lu, S., Jia, L., Zhang, W., Li, H.: Encoder-decoder contrast for unsupervised anomaly detection in medical images. IEEE Transactions on Medical Imaging **43**(3), 1102–1112 (2024). https://doi.org/10.1109/TMI.2023.3327720
10. Guo, J., Lu, S., Zhang, W., Chen, F., Li, H., Liao, H.: Dinomaly: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20405–20415 (2025). https://doi.org/10.1109/CVPR52734.2025.01900
11. Karthik, Maggie, Dane, S.: Aptos 2019 blindness detection. Kaggle Competition (2019), https://www.kaggle.com/competitions/aptos2019-blindness-detection
12. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–1131.e9 (2018). https://doi.org/10.1016/j.cell.2018.02.010
13. Li, L., Liu, X., Hou, X., Chen, L., Zhou, Y., Fu, S.: Kmocol: k-positive momentum contrastive learning for imbalanced diabetic retinopathy grading. IEEE Transactions on Instrumentation and Measurement **74**, 1–12 (2025). https://doi.org/10.1109/TIM.2025.3542859
14. Luo, W., Cao, Y., Yao, H., Zhang, X., Lou, J., Cheng, Y., Shen, W., Yu, W.: Exploring intrinsic normal prototypes within a single image for universal anomaly detection. arXiv preprint (2025), arXiv:2503.02424
15. Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T.T., Dinh, D.H., et al.: Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. Scientific Data **9**(1), 429 (2022). https://doi.org/10.1038/s41597-022-01498-w
16. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. Transactions on Machine Learning Research (2024)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Tang, P., Yan, X., Hu, X., Wu, K., Lasser, T., Shi, K.: Anomaly detection in medical images using encoder-attention-2decoders reconstruction. IEEE Transactions on Medical Imaging **44**(8), 3370–3382 (2025). https://doi.org/10.1109/TMI.2025.3563482
19. Wortsman, M., Dettmers, T., Zettlemoyer, L., Morcos, A., Farhadi, A., Schmidt, L.: Stable and low-precision training for large-scale vision-language models. In: Advances in Neural Information Processing Systems. vol. 36, pp. 10271–10298 (2023)