Supplementary Material for Normal and Abnormal Pathology Knowledge-Augmented Vision-Language Model for Anomaly Detection in Pathology Images

1. Detailed Experimental Setup

1.1. Pathology term pools

We present the pathology terms that we collected and used for lymph node metastasis detection in Tab. S1. The curated term pools consist of 92 normal terms and 48 abnormal terms.

1.2. Implementation details of competing models

Seven competing models are involved in this study. For GANomaly, STFPM, Fastflow, CFA, and EfficientAD, we followed the implementation of anomalib [2]. Based on the benchmark results, we selected the specific backbone architectures and implementation strategies. These models were trained for 1 epoch and evaluated on an NVIDIA RTX 3090 GPU. Detailed implementations settings are as follows:

GANomaly. We set the batch size to 100 for training. Following the original work [1], we adopted Adam optimizer with an initial learning rate of 0.0002 and momentum parameters β_1 = 0.5, β_2 = 0.999. The size of the autoencoder latent vector was set to 100. The weights for the loss functions were set to w_{bce} = 1, w_{rec} = 50, and w_{enc} = 1.

STFPM. The batch size was set to 100 for training. For the remaining settings, we followed the optimized configurations from the original work [9], utilizing the first three blocks of ResNet18 [4] as the feature extractor and Stochastic Gradient Descent optimizer with a learning rate of 0.4.

Fastflow. We set the batch size to 50 for training due to computational limitations. Following the optimized configurations provided in the original work [11], we set all other settings based on WideResNet50 [12]: Adam optimizer with a learning rate of 0.001 and weight decay of 0.00001, and 8-step flows.

CFA. We set the batch size to 50 for training because of computational limitations. For other settings, we followed the optimized configurations provided in the CFA [5] paper, based on the WideResNet50 [12] feature extractor: AdamW optimizer applied with amsgrad, a learning rate of 0.001 and weight decay of 0.0005, the number of nearest neighbors for each patch feature to 3, and γ_c and γ_d to 1.

EfficientAD. Following the original work [3], we adopted EfficientAD-S and utilized Adam optimizer with a learning rate of 0.0001 and a weight decay of 0.00001, a batch size of 1, and 384 convolution output channels.

The remaining two models (AnoDDPM and AnomalyCLIP) are implemented as follows:

AnoDDPM. [10] We used the official implementation of AnoDDPM and adopted the training settings of AnoDDPM-PNDM [6]. Specifically, we employed 100 timesteps $t \in [10, 20, 30, ..., 990, 1000]$, each representing a specific noise level in the denoising process, using the PNDM sampler [7]. The model was trained using Adam optimizer with a batch size of 28 for 1 million iterations (approximately 20 epochs). The optimal model was selected based on a signal-to-noise ratio (SNR) of 0.1, which showed the best performance on **Camelyon16** as reported in [6]. Evaluation was conducted on an NVIDIA RTX A6000 GPU.

AnomalyCLIP. [13] We employed the official implementation of AnomalyCLIP with CLIP(ViT-L/14) [8] as the backbone. The model was trained for one epoch with a batch size of 64. To tailor the approach to lymph node detection, the prompt learner was initialized with the class name "lymph node tissue". For the remaining settings, we followed the optimized configurations outlined in [13]. Specifically, we adopted Adam optimizer with a learning rate of 0.001. The number of learnable text prompts was set to 12. Trainable text tokens were attached to the first 9 layers of the text encoder. Each text token had a length of 4.

1.3. Computational efficiency evaluation

Latency and peak GPU memory usage were measured with a batch size of 28, and dummy iterations were performed beforehand to stabilize execution. GPU synchronization using `torch.cuda.synchronize` was applied before and after the inference to ensure accurate timing. Peak GPU memory usage was tracked with `torch.cuda.max_memory_reserved`, which records the highest reserved memory during inference.

Normal terms	36. Macrophages	72. Prominent single nucleoli	13. Microacinar architecture along its advancing edge
1. Helper T lymphocyte	37. Plasmacytoid lymphocytes	73. Efferent vessels	14. Tumor budding
2. Small dormant lymphocytes	38. Capillaries	74. Secondary lymphoid follicles	15. Large apocrine-like pleomorphic cells with large nuclei and prominent nucleoli
3. Large B lymphocytes	39. Plasmablasts	75. Primary follicle	16. Discohesive cells
4. Littoral cells	40. Distinct cytoplasmic boundaries	76. Frequent mitotic figures	17. Isolated or small clusters of malignant cells in the stroma
5. Large and small cleaved follicular center cells scant cytoplasm and Inconspicuous nucleoli	41. Follicle	77. Mature T cells	18. Glands with cells that have nuclear pseudostratification
6. Sclerosis in an inguinal lymph node	42. Histiocytes	78. Trabeculae	19. Cuboidal cells with eosinophilic cytoplasm and central nucleus
7. Sinuses	43. Small B and T lymphocytes	79. B cells	20. Thickened capsule
8. Marginal zone	44. Tightly packed anastomosing networks	80. Memory B cells	21. Central lumenal spaces of some small glands are filled by tumor cells producing small solid areas
9. Large B cells scattered throughout the paracortex	45. Recirculating cells	81. Arterioles	22. Caseous necrosis
10. Germinal center	46. Abundant cytoplasm with medium to large nuclei with vesicular chromatin	82. Cells are elongated and resemble fibroblasts	23. Sinusoidal permeation
11. Squamous endothelium	47. Paracortex	83. Coarse network of reticulin fibers	24. Microacinar structures
12. Interdigitating dendritic cells	48. Faintly granular cytoplasm	84. Postfollicular memory B cells	25. Medium to small glands have an internal structure formed by microacini
13. Histiocytes and high endothelial venules	49. Distinct group of non T and non B lymphocytes	85. Lymphatic vessels	26. Epithelioid cell clusters
14. Lymphocytes	50. Subcapsular sinus	86. T lymphocytes	27. Glandular arrangement
15. Large round nuclei	51. Afferent lymph vessels	87. T cells	28. Sinus involvement
16. Circulatory monocytes	52. Capsule	88. Plasmacytoid dendritic cells	29. Small solid clusters or buds of tumor cells
17. Lymphatic artery	53. Trabecular sinuses	89. Cortex	30. Irregularly folded, distorted, and small tubules
18. Discontinuous endothelium	54. Open chromatin	90. B-cell-rich non-Germinal center	31. Comedo, trabecular and papillary patterns
19. Smooth muscle	55. B immunoblasts	91. Helper T cells	32. Advancing edge of the adenocarcinomas
20. Small unchallenged B cells	56. Follicles	92. Medulla	33. Single and small clusters of undifferentiated cells are admixed
21. Medullary sinuses	57. Tingible body macrophages		34. May expand sinuses
22. Basophilic cytoplasm	58. Smooth muscle proliferation in lymph node hilum		35. Glandular structure is completely or almost completely lost
23. Erythrocytes	59. Large pale nuclei	**Abnormal terms**	36. Paracortical expansion
24. Centrocytes	60. Plasma cells	1. Tumor buds that emerge from medium-sized tubules	37. Cells grow predominantly in solid masses or cords
25. Tingible body macrophages contain apoptotic bodies	61. Straight branches	2. Completely destroyed sinus architecture	38. Glands acquire a haphazard arrangement with marked variation in size, shape, and outline
26. Endothelial cell	62. Blood vessel	3. Cells are discontinuous from the more superficial malignant glands	39. Cytokeratin
27. Peripheral nucleoli	63. Medullary cords, sinuses and vessel	4. Small isolated round tubules within the stroma	40. Complex, irregular, cribriform glands and small solid areas
28. Cortical sinus	64. Abundant basophilic cytoplasm	5. Balloon cell variant resembles histiocytes although nuclei are atypical	41. Apoptosis
29. Mantle zone	65. Primary follicles	6. Partially destroyed sinus architecture	42. Large apocrine-like pleomorphic cells with pink, granular cytoplasm
30. Follicular dendritic cells	66. Intranodal vessels	7. Cytoplasmic mucin	43. Glands are loosely and irregularly arranged
31. Plasma cell-rich Germinal center	67. Lymphoid nodules	8. Small tubules that formed cribriform structures within medium or large gland	44. Complex or simple tubules with a compact glandular structure
32. Large noncleaved follicular center cells	68. Centroblasts	9. Undifferentiated cells	45. Grade 3 adenocarcinoma
33. Immunoblasts	69. Quiescent B cells	10. Diffuse growth pattern	46. Glands are lined by 3 or more layers
34. Mast cells	70. Thin connective tissue capsule	11. Glands are small, round, and microacinar	47. Large, highly irregular glands that frequently have outpouchings and microacinar forms
35. Dense connective tissue	71. Tingible body macrophages have clear cytoplasm	12. Nodular growth pattern	48. Acini are lined by 2 or 3 layers of cells with basally oriented nuclei

Table S1. Pathology terms in the curated term pools.

	Method	CONCH-class	CONCH-pool	CONCH($\mathbf{v}^I$)	Ano-NAViLa
GastricLN	AUROC ($A^{\max}_{score}$)	0.9068 [0.85, 0.96]	0.7619 [0.67, 0.85]	0.9902 [0.97, 1.00]	**0.9967 [0.99, 1.00]**
	AUPR ($A^{\max}_{score}$)	0.8987 [0.81, 0.96]	0.6902 [0.57, 0.82]	0.9922 [0.98, 1.00]	**0.9971 [0.99, 1.00]**
	AUROC ($A^{\text{top1\%}}_{score}$)	0.8896 [0.82, 0.95]	0.7166 [0.61, 0.81]	0.9828 [0.96, 1.00]	**0.9894 [0.97, 1.00]**
	AUPR ($A^{\text{top1\%}}_{score}$)	0.8607 [0.74, 0.95]	0.6107 [0.49, 0.76]	0.9844 [0.96, 1.00]	**0.9904 [0.98, 1.00]**
	Patch (AUROC)	0.8942 [0.89, 0.90]	0.8300 [0.83, 0.83]	0.9651 [0.96, 0.97]	**0.9681 [0.97, 0.97]**
Camelyon16	AUROC ($A^{\max}_{score}$)	0.7459 [0.66, 0.83]	0.5805 [0.48, 0.68]	0.7857 [0.70, 0.86]	**0.8594 [0.79, 0.92]**
	AUPR ($A^{\max}_{score}$)	0.6762 [0.54, 0.80]	0.4257 [0.32, 0.57]	0.6844 [0.54, 0.80]	**0.8309 [0.73, 0.91]**
	AUROC ($A^{\text{top1\%}}_{score}$)	0.7681 [0.67, 0.85]	0.5135 [0.41, 0.61]	**0.7898 [0.70, 0.87]**	0.7702 [0.67, 0.86]
	AUPR ($A^{\text{top1\%}}_{score}$)	0.7338 [0.61, 0.84]	0.3837 [0.28, 0.50]	0.7569 [0.64, 0.85]	**0.7941 [0.69, 0.87]**
Camelyon16$_{\text{macro}}$	AUROC ($A^{\max}_{score}$)	0.8358 [0.74, 0.92]	0.6338 [0.49, 0.77]	0.8940 [0.82, 0.95]	**0.9858 [0.96, 1.00]**
	AUPR ($A^{\max}_{score}$)	0.6255 [0.39, 0.81]	0.3114 [0.18, 0.51]	0.6931 [0.47, 0.86]	**0.9547 [0.88, 1.00]**
	AUROC ($A^{\text{top1\%}}_{score}$)	0.9000 [0.80, 0.97]	0.5051 [0.37, 0.64]	0.9557 [0.90, 0.99]	**0.9761 [0.92, 1.00]**
	AUPR ($A^{\text{top1\%}}_{score}$)	0.7821 [0.59, 0.91]	0.2256 [0.13, 0.36]	0.8847 [0.74, 0.98]	**0.9699 [0.89, 1.00]**

Table S2. AD performance compared to the VLM baseline.

2. Additional Experiments and Analyses

2.1. AD performance of baseline VLM

To compare Ano-NAViLa and the zero-shot performance of the VLM, we employed CONCH in its original form to conduct AD using the text prompt "`an image showing KEYWORD`" by replacing `KEYWORD` with 1) CONCH-class: either normal or metastasis lymph node, and 2) CONCH-pool: each of normal and abnormal terms in the pool. Both variants obtained substantially lower performance than Ano-NAViLa (Tab. S2). These results suggest that while the performance gain can be partially attributed to the use of CONCH, the strong performance, particularly on **Camelyon16**, is largely due to the proposed text-augmented embeddings.

2.2. AD performance without erosion operation

We repeated the experiments without the 3x3 erosion operation. The result are illustrated in Tab. S3. The performance of Ano-NAViLa were less sensitive to the erosion operation than other competing models. Ano-NAViLa outperformed others for all metrics except AUROC using $A^{\text{top1\%}}_{score}$ on **Camelyon16**. For this metric, CFA achieved the best performance, followed by STFPM. Ano-NAViLa ranked third with an AUROC of 0.8092. Nevertheless, Ano-NAViLa obtained a higher AUROC of 0.8543 using $A^{\max}_{score}$, surpassing CFA, which achieved 0.8176 AUROC using $A^{\text{top1\%}}_{score}$.

2.3. Anomaly localization

In Fig. S1, we provide additional heatmap visualizations of WSI-level AD results from abnormal WSIs in **Camelyon16**. The visualizations include heatmaps from the ground truth, Ano-NAViLa, and STFPM, achieving the second-best performance in AD and localization.

2.4. Distribution of image-text similarities

Fig. S2 and Fig. S3 show the distribution of image-text similarities in the normal and abnormal term pool, respectively. The numbers on the x-axis correspond to the indices of the pathology terms in Tab. S1. For (a) and (b) in both Fig. S2 and Fig. S3, we calculated the similarity scores of the normal or abnormal-labeled patches of each dataset and the pathology terms in each pool. For (c) and (d), we analyzed the representative patches (the patches with the highest anomaly scores in the WSIs) of normal or abnormal WSIs in each dataset. We averaged the similarity scores by dataset and label, and then visualized the distributions. The results reveal clear differences in similarity distributions between normal and abnormal images when combined with the normal and abnormal term pools, indicating strong discriminative alignment between image features and pathology-specific textual descriptions.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Computer Vision – ACCV 2018*, pages 622–637, 2019. 1

[2] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, and Utku Genc. Anomalib: A deep learning library for anomaly detection. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1706–1710. IEEE, 2022. 1

[3] Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 128–138, 2024. 1

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[5] Sungwook Lee, Seunghyun Lee, and Byung Cheol Song. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*, 10: 78446–78454, 2022. 1

[6] Jasper Linmans, Gabriel Raya, Jeroen van der Laak, and Geert Litjens. Diffusion models for out-of-distribution detection in digital pathology. *Medical Image Analysis*, 93: 103088, 2024. 1

[7] Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In *International Conference on Learning Representations*, 2022. 1

[8] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry,

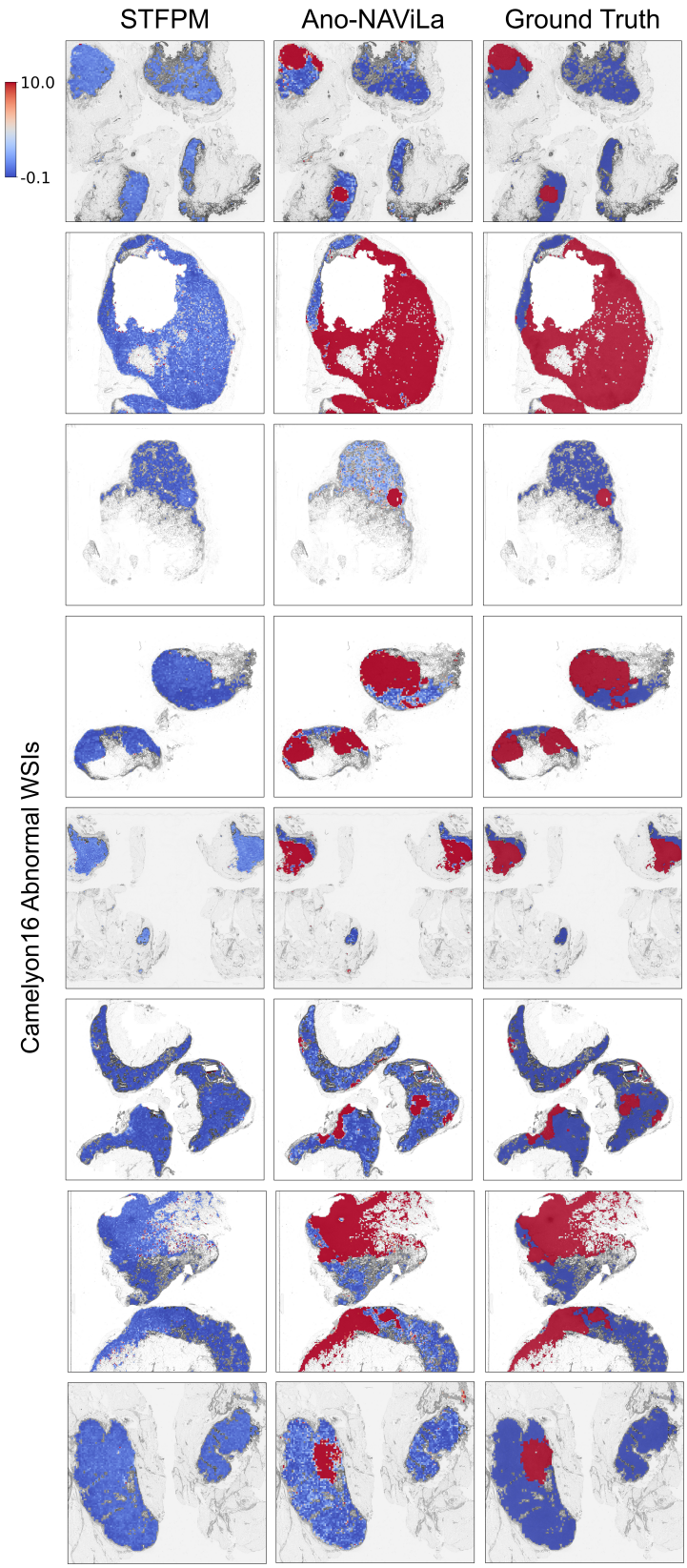

Figure S1. Visualization of anomaly localization in abnormal WSIs from Camelyon16.

Method	GastricLN (WSI)				Patch	Camelyon16				Camelyon16$_{\text{macro}}$			
	AUROC ($A^{\max}_{score}$)	AUPR ($A^{\max}_{score}$)	AUROC ($A^{\text{top1\%}}_{score}$)	AUPR ($A^{\text{top1\%}}_{score}$)	AUROC	AUROC ($A^{\max}_{score}$)	AUPR ($A^{\max}_{score}$)	AUROC ($A^{\text{top1\%}}_{score}$)	AUPR ($A^{\text{top1\%}}_{score}$)	AUROC ($A^{\max}_{score}$)	AUPR ($A^{\max}_{score}$)	AUROC ($A^{\text{top1\%}}_{score}$)	AUPR ($A^{\text{top1\%}}_{score}$)
GANomaly	0.4153 [0.31, 0.52]	0.4769 [0.36, 0.61]	0.3143 [0.22, 0.42]	0.4129 [0.31, 0.52]	0.4182 [0.42, 0.42]	0.5717 [0.47, 0.67]	0.4962 [0.35, 0.63]	0.5827 [0.48, 0.68]	0.4713 [0.34, 0.61]	0.6193 [0.48, 0.75]	0.3535 [0.18, 0.54]	0.6091 [0.48, 0.73]	0.3348 [0.18, 0.52]
STFPM	0.9410 [0.88, 0.99]	0.8855 [0.78, 0.98]	0.9940 [0.98, 1.00]	0.9926 [0.98, 1.00]	0.9538 [0.95, 0.96]	0.7398 [0.65, 0.82]	0.5568 [0.43, 0.72]	0.7954 [0.71, 0.88]	0.6914 [0.55, 0.85]	0.7665 [0.67, 0.85]	0.3544 [0.22, 0.57]	0.8250 [0.71, 0.92]	0.5584 [0.35, 0.82]
FastFlow	0.9525 [0.90, 0.99]	0.9175 [0.81, 0.99]	0.9758 [0.94, 1.00]	0.9320 [0.82, 1.00]	0.9242 [0.92, 0.93]	0.7814 [0.70, 0.86]	0.6628 [0.51, 0.80]	0.8120 [0.73, 0.89]	0.7371 [0.61, 0.85]	0.8472 [0.75, 0.93]	0.5980 [0.38, 0.81]	0.8676 [0.76, 0.95]	0.7014 [0.48, 0.89]
CFA	0.9598 [0.91, 1.00]	0.8898 [0.77, 1.00]	0.9828 [0.94, 1.00]	0.9274 [0.82, 1.00]	0.8881 [0.89, 0.89]	0.8306 [0.75, 0.91]	0.7667 [0.64, 0.87]	**0.8176 [0.73, 0.89]**	0.7774 [0.66, 0.87]	0.9034 [0.81, 0.97]	0.7603 [0.56, 0.91]	0.8898 [0.78, 0.97]	0.7915 [0.62, 0.92]
AnoDDPM	0.7816 [0.69, 0.86]	0.7200 [0.60, 0.85]	0.8454 [0.77, 0.91]	0.8415 [0.75, 0.92]	0.8860 [0.88, 0.89]	0.5311 [0.37, 0.67]	0.5573 [0.23, 0.59]	0.4990 [0.33, 0.65]	0.5107 [0.18, 0.53]	0.6869 [0.54, 0.83]	0.5934 [0.41, 0.77]	0.7034 [0.55, 0.83]	0.5640 [0.36, 0.76]
EfficientAD	0.8584 [0.78, 0.93]	0.7823 [0.65, 0.90]	0.8590 [0.78, 0.94]	0.7441 [0.62, 0.89]	0.8432 [0.84, 0.85]	0.6342 [0.54, 0.73]	0.5312 [0.39, 0.66]	0.6224 [0.52, 0.72]	0.5243 [0.39, 0.65]	0.6528 [0.53, 0.77]	0.2939 [0.17, 0.49]	0.6193 [0.49, 0.74]	0.3150 [0.17, 0.49]
AnomalyCLIP	0.3959 [0.30, 0.50]	0.4086 [0.32, 0.52]	0.2241 [0.14, 0.31]	0.3494 [0.27, 0.45]	0.3314 [0.33, 0.33]	0.6217 [0.52, 0.72]	0.5216 [0.39, 0.66]	0.5796 [0.48, 0.68]	0.4817 [0.34, 0.61]	0.6500 [0.53, 0.76]	0.3575 [0.19, 0.54]	0.5750 [0.44, 0.71]	0.3027 [0.16, 0.48]
Ano-NAViLa	**0.9994 [1.00, 1.00]**	**0.9994 [1.00, 1.00]**	**0.9976 [0.99, 1.00]**	**0.9978 [0.99, 1.00]**	**0.9681 [0.97, 0.97]**	**0.8543 [0.78, 0.92]**	**0.8413 [0.75, 0.91]**	0.8092 [0.72, 0.90]	**0.8383 [0.75, 0.91]**	**0.9977 [0.99, 1.00]**	**0.9923 [0.97, 1.00]**	**1.0000 [1.00, 1.00]**	**1.0000 [1.00, 1.00]**

Table S3. AD performance without the erosion operation on GastricLN and Camelyon16 datasets.

Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 1

[9] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. *The British Machine Vision Conference (BMVC)*, 2021. 1

[10] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 650–656, 2022. 1

[11] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 1

[12] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *CoRR*, abs/1605.07146, 2016. 1

[13] Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2023. 1

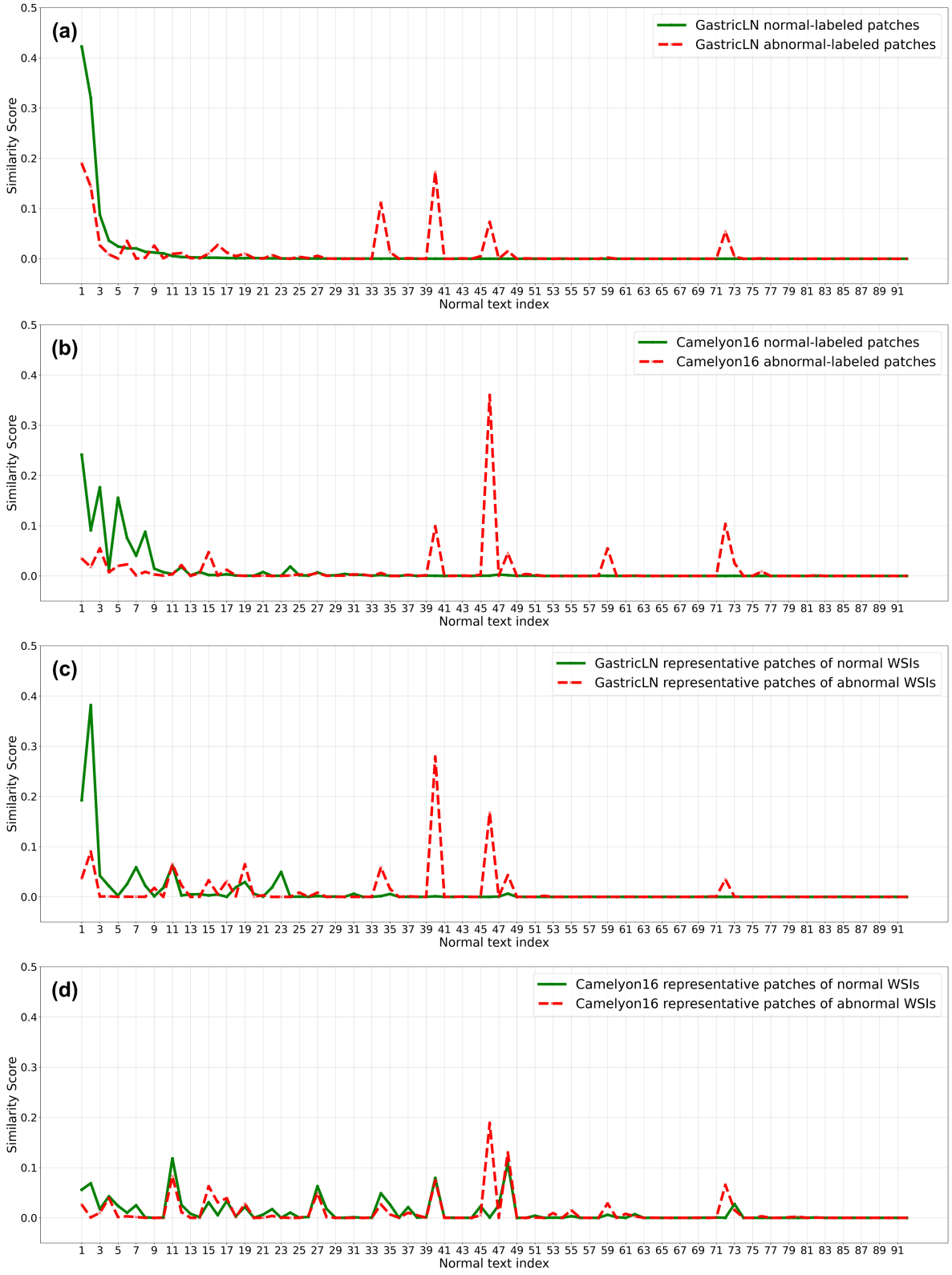

Figure S2. Distribution of image-text similarities in the normal term pool. For both GastricLN and Camelyon16 datasets, (a) and (b) are patch-level analysis results, and (c) and (d) are WSI-level analysis results.

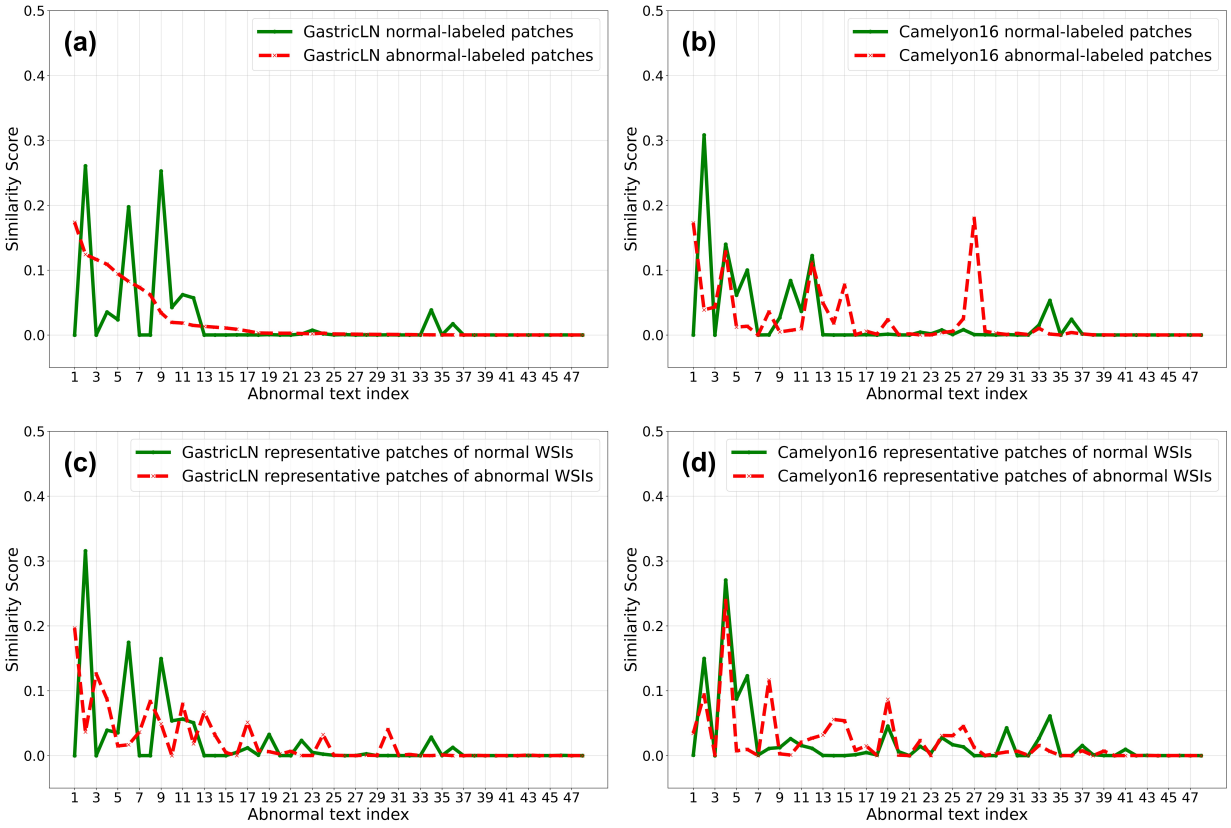

Figure S3. Distribution of image-text similarities in the abnormal term pool. For both GastricLN and Camelyon16 datasets, (a) and (b) are patch-level analysis results, and (c) and (d) are WSI-level analysis results.