Histopath-C: Towards Realistic Domain Shifts for Histopathology Vision-Language Adaptation - Supplementary Material

Mehrdad Noori*   Gustavo A. Vargas Hakim*   David Osowiechi*
Fereshteh Shakeri   Ali Bahri   Moslem Yazdanpanah   Sahar Dastani
Ismail Ben Ayed   Christian Desrosiers

LIVIA, ÉTS Montreal, Canada
International Laboratory on Learning Systems (ILLS)

1. Implementation Details

We conduct all experiments using the Quilt VLM (QuiltNet-B-32) [4], a ViT-B/32–based model with 32×32 px patches and 224×224 px input resolution. We also evaluate on Path-Gen [9], which uses a ViT-B/16 backbone with 16×16 px patches and a 224×224 px input resolution, trained on the large-scale PathGen pathology image–text dataset. Finally, we evaluate on the CONCH VLM [7], which employs a ViT-B/16 backbone with 16×16 px patches and a 448×448 px input resolution, paired with a GPT-style text encoder for captioning and contrastive alignment. Following prior work [10], both our method (LATTE) and all baseline TTA approaches are fine-tuned at test time using the Adam optimizer with a fixed learning rate of 1×10^{-3} for 10 adaptation steps and a batch size of 128.

All runs use PyTorch on NVIDIA V100 GPUs with 32 GB of memory. To ensure statistical robustness, each experiment is repeated three times; we report the mean and standard deviation of each metric.

We provide full reproduction scripts at https://github.com/Mehrdad-Noori/Histopath-C, including on-the-fly corruption generation for Histopath-C, test-time adaptation routines for LATTE and all baselines, and evaluation code for computing zero-shot and post-adaptation metrics. For fairness, baselines use the same general hyperparameters above; any method-specific settings are set as recommended in their original publications.

2. Dataset Details

To assess the effectiveness and generalization capabilities of LATTE, we perform extensive evaluations across a suite of diverse and challenging histopathology datasets, encompassing various organs, cancer types, imaging conditions, and annotation granularities.

- **NCT-7K and NCT-100K** [5] are large-scale colorectal cancer datasets composed of tissue patches extracted from whole-slide images. NCT-7K contains 7,180 non-overlapping patches, while NCT-100K extends this to 100,000 patches. Both datasets are annotated into nine tissue categories and can be grouped into broader classes of cancerous vs. normal tissues. These datasets serve as standard benchmarks for evaluating performance in colorectal cancer classification.

- **LC25000** [1] comprises digitized histology images from two organs—lung and colon—and includes binary classification tasks distinguishing adenocarcinoma from normal tissue. We evaluate LATTE on the LC25K-Lung and LC25K-Colon subsets individually, as well as on a combined version (LC25K-All) to introduce additional heterogeneity and assess cross-organ generalization under stronger distributional shifts.

- **SkinCancer** [6] is derived from the publicly available BCN20000 dataset and includes tissue patches extracted from skin biopsies. The dataset spans 12 anatomical compartments and 4 tumor types, such as basal cell carcinoma and melanoma, which we consolidate into the SkinTumor subset. This dataset introduces both intra-class variability and inter-class visual overlap, reflecting real-world diagnostic challenges in dermatopathology.

- **RenalCell** [3] includes histological images of clear cell renal cell carcinoma, annotated into five distinct tissue textures, including tumor, normal, and other microenvironmental structures. This dataset enables evaluation of LATTE on texture-centric classification tasks in renal oncology.

- **MHIST** [12] is a curated dataset of colorectal polyp images, specifically labeled as hyperplastic or sessile serrated adenomas (SSAs)—two classes with subtle morphological differences that pose a challenge even for ex-

---
*Equal contribution.
Correspondence to mehrdad.noori.1@ens.etsmtl.ca

	Template
T^1	"a histopathology slide showing {class k}"
T^2	"histopathology image of {class k}"
T^3	"pathology tissue showing {class k}"
T^4	"presence of {class k} tissue on image"
T^5	"a photomicrograph showing {class k}"
T^6	"a photomicrograph of {class k}"
T^7	"an image of {class k}"
T^8	"an image showing {class k}"
T^9	"an example of {class k}"
T^{10}	"{class k} is shown"
T^{11}	"this is {class k}"
T^{12}	"there is {class k}"
T^{13}	"a histopathological image showing {class k}"
T^{14}	"a histopathological image of {class k}"
T^{15}	"a histopathological photograph of {class k}"
T^{16}	"a histopathological photograph showing {class k}"
T^{17}	"shows {class k}"
T^{18}	"presence of {class k}"
T^{19}	"{class k} is present"
T^{20}	"an H&E stained image of {class k}"
T^{21}	"an H&E stained image showing {class k}"
T^{22}	"an H&E image showing {class k}"
T^{23}	"an H&E image of {class k}"
T^{24}	"{class k}, H&E stain"
T^{25}	"{class k}, H&E"

Table 1. Full list of text templates used in LATTE. These templates provide diverse linguistic expressions of each class and are used during adaptation for loss-level ensembling.

pert pathologists. This binary classification task assesses model sensitivity to fine-grained structural features in colorectal screening.

For all datasets, we apply the ten corruption types defined in Histopath-C and refer to the resulting corrupted versions as {*DatasetName*}-*C* throughout the paper.

3. Text Templates

To account for the sensitivity of vision-language models to prompt phrasing, we evaluate our method using a diverse set of 25 text templates that describe each class from different linguistic perspectives. These templates range from generic image captions (e.g., "an image of class") to domain-specific phrasings that reference histopathological context and H&E staining.

Table 1 lists all templates used in our experiments. During adaptation, we employ loss-level ensembling across these prompts to improve prediction stability and mitigate prompt-induced variability.

4. Detailed Results of the Main Paper

This section provides complete versions of the experimental results that were summarized or partially reported in the main paper. The following tables report the detailed results when Quilt is used as the base VLM with different adaptation methods, TENT [11], LAME [2], TPT [8], and CLIPArTT [10], across various datasets and evaluation settings. More specifically, Table 2 reports detailed results for the

NCT100K dataset, Table 3 for LC25K-Lung, Table 4 for LC25K-Colon, Table 5 for LC25K-All, Table 6 for Skin, Table 7 for Renal, and Table 8 for MHIST. These results provide a more comprehensive assessment of the effectiveness and robustness of our approach, demonstrating that while LATTE may be outperformed by other methods on specific corruptions, it remains the most consistent overall across datasets and corruption types.

To further assess the generality of our approach, we conducted experiments on two additional vision–language models, PathGen [9] and CONCH [7]. The detailed results, reported in Tables 9 - 24, consistently demonstrate that LATTE not only improves upon the source baseline but also frequently outperforms existing state-of-the-art methods by a significant margin. These findings highlight the robustness and adaptability of LATTE across diverse model architectures and corruption scenarios.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
NCT100K	55.98	41.42 $\pm$0.04	64.06 $\pm$0.21	52.83 $\pm$0.00	59.86 $\pm$0.01	**68.14** $\pm$**0.03**

NCT100K-C		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
	Stain-Light	47.35	37.53 $\pm$0.05	48.79 $\pm$0.07	44.74 $\pm$0.00	50.12 $\pm$0.05	**59.54** $\pm$**0.00**
	Stain-Heavy	38.51	17.23 $\pm$0.02	35.33 $\pm$0.08	37.19 $\pm$0.01	52.30 $\pm$0.04	**58.61** $\pm$**0.01**
	Dust	54.64	47.59 $\pm$0.14	60.69 $\pm$0.11	51.79 $\pm$0.04	61.33 $\pm$0.03	**64.47** $\pm$**0.08**
	Air Bubble	57.78	53.44 $\pm$0.01	58.54 $\pm$0.17	56.79 $\pm$0.03	58.39 $\pm$0.03	**63.25** $\pm$**0.07**
	Defocus Blur	50.02	35.88 $\pm$0.07	42.69 $\pm$0.09	48.67 $\pm$0.01	47.31 $\pm$0.05	**60.20** $\pm$**0.09**
	Motion Blur	28.94	15.63 $\pm$0.02	26.47 $\pm$0.07	27.40 $\pm$0.06	40.79 $\pm$0.06	**44.05** $\pm$**0.25**
	Gaussian Noise	32.12	16.01 $\pm$0.03	25.86 $\pm$0.03	30.64 $\pm$0.02	56.23 $\pm$0.03	**62.09** $\pm$**0.11**
	Shot Noise	28.83	22.11 $\pm$0.12	19.37 $\pm$0.05	25.78 $\pm$0.01	48.69 $\pm$0.08	**55.65** $\pm$**0.10**
	Brightness	37.27	22.43 $\pm$0.00	35.19 $\pm$0.11	37.20 $\pm$0.01	**57.39** $\pm$**0.03**	52.24 $\pm$0.06
	Contrast	23.47	15.14 $\pm$0.00	19.03 $\pm$0.02	23.95 $\pm$0.00	39.22 $\pm$0.05	**41.18** $\pm$**0.04**
	Mean	39.89	28.30	37.20	38.42	51.18	**56.13**

Table 2. Comparison of different adaptation methods on the NCT-100k dataset (8 classes) using Quilt [4] as the base VLM.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Lung	82.87	70.34 $\pm$0.18	88.00 $\pm$0.18	83.03 $\pm$0.01	-	**89.41** $\pm$**0.15**

LC25K-Lung-C		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
	Stain-Light	73.79	45.57 $\pm$0.75	72.14 $\pm$0.31	73.85 $\pm$0.01	-	**78.70** $\pm$**0.13**
	Stain-Heavy	57.88	33.33 $\pm$0.00	56.92 $\pm$0.18	59.55 $\pm$0.00	-	**75.57** $\pm$**0.11**
	Dust	73.20	53.29 $\pm$0.16	71.85 $\pm$0.67	71.69 $\pm$0.09	-	**88.89** $\pm$**0.02**
	Air Bubble	76.75	62.57 $\pm$0.01	68.75 $\pm$0.21	77.12 $\pm$0.03	-	**84.61** $\pm$**0.31**
	Defocus Blur	73.40	34.47 $\pm$0.05	84.41 $\pm$0.57	70.44 $\pm$0.00	-	**91.19** $\pm$**0.01**
	Motion Blur	77.89	64.18 $\pm$0.27	82.30 $\pm$0.14	77.13 $\pm$0.02	-	**86.75** $\pm$**0.62**
	Gaussian Noise	80.54	70.64 $\pm$0.46	**85.86** $\pm$**0.63**	78.99 $\pm$0.08	-	76.43 $\pm$0.20
	Shot Noise	78.47	62.49 $\pm$0.21	**83.85** $\pm$**0.57**	77.31 $\pm$0.03	-	71.35 $\pm$0.56
	Brightness	77.19	66.63 $\pm$0.06	71.50 $\pm$0.03	76.57 $\pm$0.00	-	**91.61** $\pm$**0.20**
	Contrast	56.32	33.33 $\pm$0.00	63.24 $\pm$0.39	57.95 $\pm$0.00	-	**42.91** $\pm$**0.04**
	Mean	72.54	52.65	74.08	72.06	-	**78.80**

Table 3. Comparison of adaptation methods on LC25K-Lung (3 classes) using Quilt [4] as the base VLM. CLIPArTT is not applicable due to fewer than three classes.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Colon	94.41	89.28 $\pm$0.54	98.70 $\pm$0.04	94.50 $\pm$0.03	-	**99.21** $\pm$**0.03**

LC25K-C		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
	Stain-Light	76.99	50.00 $\pm$0.00	82.77 $\pm$0.54	77.00 $\pm$0.00	-	**96.84** $\pm$**0.07**
	Stain-Heavy	61.06	50.00 $\pm$0.00	53.37 $\pm$0.54	63.51 $\pm$0.01	-	**94.88** $\pm$**0.04**
	Dust	90.42	58.76 $\pm$1.14	97.42 $\pm$0.60	90.08 $\pm$0.10	-	**98.04** $\pm$**0.15**
	Air Bubble	89.24	50.95 $\pm$0.11	95.82 $\pm$0.47	89.22 $\pm$0.36	-	**96.38** $\pm$**0.02**
	Gaussian Noise	75.78	50.00 $\pm$0.00	82.52 $\pm$0.16	76.56 $\pm$0.04	-	**99.00** $\pm$**0.02**
	Shot Noise	60.23	50.00 $\pm$0.00	52.22 $\pm$0.32	59.54 $\pm$0.10	-	**97.83** $\pm$**0.02**
	Defocus Blur	81.89	50.00 $\pm$0.00	94.22 $\pm$0.03	83.54 $\pm$0.00	-	**97.70** $\pm$**0.15**
	Motion Blur	83.46	56.04 $\pm$0.19	**87.36** $\pm$**0.03**	80.85 $\pm$0.25	-	61.63 $\pm$0.13
	Brightness	95.50	91.12 $\pm$0.33	98.99 $\pm$0.00	95.70 $\pm$0.00	-	**99.15** $\pm$**0.02**
	Contrast	66.70	50.00 $\pm$0.00	53.86 $\pm$0.31	59.84 $\pm$0.01	-	**80.22** $\pm$**0.00**
	Mean	78.13	55.69	79.86	77.58	-	**92.17**

Table 4. Comparison of adaptation methods on LC25K-Colon (2 classes) using Quilt [4] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-All	79.28	71.39 ±0.12	**87.13** ±0.17	79.22 ±0.01	80.47 ±0.01	86.97 ±0.01
LC25K-C Stain-Light	60.17	55.97 ±0.16	61.49 ±0.43	59.32 ±0.00	**73.04** ±0.05	71.25 ±0.07
Stain-Heavy	46.36	24.53 ±0.17	44.72 ±0.11	45.48 ±0.01	62.49 ±0.04	**76.58** ±0.07
Dust	68.65	59.92 ±0.06	68.42 ±0.32	68.18 ±0.01	78.85 ±0.03	**87.58** ±0.01
Air Bubble	72.05	64.32 ±0.06	77.97 ±0.51	72.15 ±0.15	77.15 ±0.03	**80.82** ±0.06
Gaussian Noise	57.59	47.81 ±0.02	57.25 ±0.32	57.51 ±0.16	**81.22** ±0.06	80.85 ±0.06
Shot Noise	48.82	21.01 ±0.04	46.19 ±0.08	46.97 ±0.09	**77.03** ±0.05	74.68 ±0.17
Defocus Blur	55.65	20.12 ±0.01	56.07 ±0.09	54.24 ±0.00	48.99 ±0.03	**70.31** ±0.15
Motion Blur	56.26	20.77 ±0.01	53.40 ±0.52	55.15 ±0.18	51.71 ±0.04	**59.46** ±0.48
Brightness	77.77	68.05 ±0.07	79.69 ±0.00	77.40 ±0.00	82.26 ±0.06	**89.30** ±0.06
Contrast	28.01	20.05 ±0.00	20.88 ±0.00	**30.03** ±0.02	24.81 ±0.03	25.93 ±0.04
Mean	57.13	40.26	56.61	56.64	65.76	**71.68**

Table 5. Comparison of adaptation methods on LC25K-All (5 classes) using Quilt [4] as the base VLM.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Skin	44.22	24.31 ±0.13	40.09 ±0.06	45.16 ±0.00	46.42 ±0.14	**50.62** ±0.24
Skin-C Stain-Light	20.29	9.64 ±0.05	11.75 ±0.05	21.09 ±0.00	23.96 ±0.10	**33.28** ±0.20
Stain-Heavy	13.68	3.28 ±0.01	9.59 ±0.18	13.72 ±0.00	20.63 ±0.14	**28.22** ±0.06
Dust	40.07	21.34 ±0.19	40.18 ±0.09	42.31 ±0.15	40.73 ±0.07	**46.62** ±0.05
Air Bubble	30.63	9.26 ±0.04	23.13 ±0.21	30.43 ±0.02	38.40 ±0.20	**44.29** ±0.05
Gaussian Noise	14.15	2.57 ±0.01	5.59 ±0.00	14.02 ±0.01	26.77 ±0.06	**33.94** ±0.09
Shot Noise	11.04	2.52 ±0.01	4.30 ±0.12	9.57 ±0.07	20.99 ±0.14	**25.95** ±0.06
Defocus Blur	25.96	20.51 ±0.01	17.97 ±0.19	27.05 ±0.01	**37.17** ±0.03	36.30 ±0.09
Motion Blur	26.44	6.90 ±0.13	23.63 ±0.03	26.43 ±0.13	31.81 ±0.09	**36.40** ±0.06
Brightness	35.18	9.42 ±0.10	34.03 ±0.63	34.89 ±0.00	**45.44** ±0.16	42.55 ±0.13
Contrast	4.64	2.39 ±0.00	2.84 ±0.02	5.17 ±0.00	9.10 ±0.07	**10.58** ±0.04
Mean	22.21	8.78	17.30	22.47	29.50	**33.81**

Table 6. Comparison of adaptation methods on Skin (16 classes) using Quilt [4] as the base VLM.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Renal	49.76	43.19 ±0.05	**50.77** ±0.08	50.29 ±0.01	43.28 ±0.12	46.14 ±0.35
Renal-C Stain-Light	26.78	17.86 ±0.45	19.83 ±0.13	28.50 ±0.01	28.02 ±0.25	**41.72** ±0.13
Stain-Heavy	15.27	1.98 ±0.02	2.73 ±0.03	14.47 ±0.02	23.48 ±0.07	**44.35** ±0.37
Dust	45.74	42.34 ±0.00	**48.91** ±0.10	45.63 ±0.10	39.02 ±0.02	47.98 ±0.31
Air Bubble	38.91	42.37 ±0.01	**43.31** ±0.04	39.19 ±0.06	28.06 ±0.02	36.01 ±0.00
Gaussian Noise	36.84	41.76 ±0.02	41.64 ±0.05	36.69 ±0.02	33.43 ±0.01	**45.63** ±0.21
Shot Noise	34.23	41.86 ±0.01	41.14 ±0.08	33.37 ±0.05	30.25 ±0.17	**41.89** ±0.01
Defocus Blur	25.70	16.14 ±0.50	**31.02** ±0.75	24.54 ±0.00	32.87 ±0.23	25.68 ±0.01
Motion Blur	44.37	42.02 ±0.01	43.19 ±0.14	**44.43** ±0.06	30.40 ±0.00	36.25 ±0.19
Brightness	24.81	12.21 ±0.27	21.41 ±0.08	23.31 ±0.00	**32.14** ±0.01	31.76 ±0.12
Contrast	11.99	10.91 ±0.01	10.78 ±0.10	11.85 ±0.00	16.05 ±0.00	**31.63** ±0.08
Mean	30.46	26.95	30.40	30.20	29.37	**38.29**

Table 7. Comparison of adaptation methods on Renal (5 classes) using Quilt [4] as the base VLM.

VLM: Quilt Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
MHIST	62.95	63.15 $\pm$0.00	63.10 $\pm$0.56	61.51 $\pm$0.00	-	**64.02** $\pm$**0.26**
MHIST-C						
Stain-Light	47.19	36.85 $\pm$0.00	36.85 $\pm$0.00	47.49 $\pm$0.00	-	**62.90** $\pm$**0.87**
Stain-Heavy	42.78	36.85 $\pm$0.00	36.85 $\pm$0.00	41.56 $\pm$0.00	-	**60.85** $\pm$**0.05**
Dust	61.92	63.15 $\pm$0.00	62.69 $\pm$0.46	61.67 $\pm$0.15	-	**64.12** $\pm$**0.97**
Air Bubble	62.64	63.15 $\pm$0.00	63.15 $\pm$0.00	62.13 $\pm$0.00	-	**63.97** $\pm$**0.61**
Gaussian Noise	61.11	63.15 $\pm$0.00	**63.56** $\pm$**1.94**	61.98 $\pm$0.26	-	62.74 $\pm$0.31
Shot Noise	61.31	**63.15** $\pm$**0.00**	59.67 $\pm$2.87	60.59 $\pm$0.00	-	62.64 $\pm$0.41
Defocus Blur	63.66	63.15 $\pm$0.00	63.15 $\pm$0.00	**63.66** $\pm$**0.00**	-	59.42 $\pm$0.36
Motion Blur	66.33	63.15 $\pm$0.00	56.86 $\pm$0.46	**67.09** $\pm$**0.46**	-	61.11 $\pm$0.92
Brightness	54.76	36.85 $\pm$0.00	41.25 $\pm$0.51	53.89 $\pm$0.05	-	**63.41** $\pm$**1.48**
Contrast	55.78	**61.05** $\pm$**0.77**	48.87 $\pm$0.77	55.89 $\pm$0.10	-	58.96 $\pm$0.20
Mean	57.75	55.05	53.29	57.60	-	**62.01**

Table 8. Comparison of adaptation methods on MHIST (2 classes) using Quilt [4] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

VLM: PathGen Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
NCT7K	64.74	73.02 $\pm$0.09	73.86 $\pm$0.15	64.42 $\pm$0.01	64.95 $\pm$0.02	**79.95** $\pm$**0.36**
NCT7K-C						
Stain-Light	50.97	60.02 $\pm$0.43	57.67 $\pm$0.01	50.64 $\pm$0.00	61.08 $\pm$0.05	**67.55** $\pm$**0.47**
Stain-Heavy	43.66	47.42 $\pm$0.16	45.26 $\pm$0.21	43.53 $\pm$0.00	56.90 $\pm$0.07	**68.81** $\pm$**0.56**
Dust	65.75	76.05 $\pm$0.11	75.17 $\pm$0.07	65.38 $\pm$0.13	71.83 $\pm$0.28	**82.05** $\pm$**0.44**
Air Bubble	48.79	52.91 $\pm$0.24	56.47 $\pm$0.60	49.09 $\pm$0.17	57.91 $\pm$0.04	**72.01** $\pm$**0.37**
Defocus Blur	45.03	43.64 $\pm$0.28	42.46 $\pm$0.47	45.63 $\pm$0.00	60.37 $\pm$0.11	**71.70** $\pm$**0.53**
Motion Blur	44.99	50.20 $\pm$0.11	45.45 $\pm$0.73	45.12 $\pm$0.34	63.97 $\pm$0.38	**76.37** $\pm$**0.21**
Gaussian Noise	23.09	11.22 $\pm$0.24	16.56 $\pm$0.46	22.49 $\pm$0.08	52.30 $\pm$0.38	**79.73** $\pm$**0.14**
Shot Noise	19.55	10.23 $\pm$0.02	13.76 $\pm$0.12	19.53 $\pm$0.06	42.89 $\pm$0.14	**76.44** $\pm$**0.38**
Brightness	56.53	62.16 $\pm$0.34	57.71 $\pm$0.42	56.13 $\pm$0.01	63.37 $\pm$0.14	**77.16** $\pm$**0.12**
Contrast	36.87	36.87 $\pm$0.03	34.09 $\pm$0.11	36.33 $\pm$0.01	44.52 $\pm$0.10	**68.72** $\pm$**0.11**
Mean	43.52	45.07	44.46	43.39	57.51	**74.05**

Table 9. Comparison of different adaptation methods on the NCT-7k dataset (8 classes) using PathGen [9] as the base VLM.

VLM: PathGen Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
NCT100K	66.41	67.3 $\pm$0.02	68.73 $\pm$0.05	65.73 $\pm$0.0	69.84 $\pm$0.02	**81.73** $\pm$**0.03**
NCT100K-C						
Stain-Light	53.99	56.56 $\pm$0.03	63.4 $\pm$0.12	54.03 $\pm$0.0	66.37 $\pm$0.01	**73.29** $\pm$**0.0**
Stain-Heavy	39.40	25.89 $\pm$0.14	39.63 $\pm$0.15	39.36 $\pm$0.01	49.36 $\pm$0.02	**68.26** $\pm$**0.05**
Defocus Blur	67.46	70.89 $\pm$0.03	74.85 $\pm$0.22	66.93 $\pm$0.01	73.12 $\pm$0.07	**79.86** $\pm$**0.08**
Motion Blur	47.78	35.47 $\pm$0.04	51.29 $\pm$0.15	47.20 $\pm$0.01	65.98 $\pm$0.07	**69.37** $\pm$**0.05**
Brightness	50.66	42.33 $\pm$0.05	47.05 $\pm$0.16	49.29 $\pm$0.0	59.86 $\pm$0.05	**70.33** $\pm$**0.06**
Contrast	48.39	44.32 $\pm$0.09	43.75 $\pm$0.05	48.48 $\pm$0.06	58.15 $\pm$0.08	**69.13** $\pm$**0.07**
Gaussian Noise	29.91	19.62 $\pm$0.07	23.71 $\pm$0.0	29.64 $\pm$0.05	52.02 $\pm$0.07	**72.35** $\pm$**0.11**
Shot Noise	27.29	15.58 $\pm$0.01	20.88 $\pm$0.06	26.85 $\pm$0.02	44.19 $\pm$0.02	**68.77** $\pm$**0.07**
Dust	57.32	61.53 $\pm$0.01	61.83 $\pm$0.06	57.10 $\pm$0.0	65.09 $\pm$0.01	**77.24** $\pm$**0.05**
Air Bubble	36.86	30.26 $\pm$0.03	28.87 $\pm$0.08	35.23 $\pm$0.0	48.09 $\pm$0.01	**68.16** $\pm$**0.06**
Mean	45.91	40.25	45.53	45.41	58.22	**71.68**

Table 10. Comparison of different adaptation methods on the NCT-100k dataset (8 classes) using PathGen [9] as the base VLM.

VLM: PathGen Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Lung	82.79	81.85 ±0.01	77.74 ±0.05	81.41 ±0.0	-	**96.05 ±0.07**
LC25K-Lung-C Stain-Light	79.57	66.81 ±0.07	71.75 ±0.73	79.29 ±0.0	-	**93.84 ±0.06**
Stain-Heavy	68.81	62.84 ±0.22	66.31 ±0.36	68.75 ±0.01	-	**88.61 ±0.07**
Dust	75.79	71.03 ±0.24	72.92 ±0.36	75.92 ±0.26	-	**94.76 ±0.03**
Air Bubble	62.75	64.39 ±0.01	65.70 ±0.24	61.69 ±0.08	-	**91.63 ±0.07**
Defocus Blur	67.42	39.16 ±0.17	54.98 ±0.28	65.13 ±0.0	-	**90.54 ±0.15**
Motion Blur	61.91	35.67 ±0.37	45.56 ±0.07	60.83 ±0.08	-	**90.83 ±0.01**
Gaussian Noise	55.27	33.38 ±0.01	49.51 ±0.28	55.27 ±0.01	-	**91.94 ±0.07**
Shot Noise	56.48	33.41 ±0.01	60.81 ±0.45	58.04 ±0.03	-	**90.82 ±0.14**
Brightness	85.40	88.78 ±0.14	90.06 ±0.31	84.52 ±0.0	-	**94.98 ±0.10**
Contrast	57.46	33.33 ±0.0	61.81 ±0.17	54.91 ±0.03	-	**65.78 ±0.13**
Mean	67.09	52.88	63.94	66.44	-	**89.37**

Table 11. Comparison of adaptation methods on LC25K-Lung (3 classes) using PathGen [9] as the base VLM. CLIPArTT is not applicable due to fewer than three classes.

VLM: PathGen Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Colon	95.92	97.69 ±0.06	**98.98 ±0.01**	95.82 ±0.0	-	98.27 ±0.07
LC25K-Colon-C Stain-Light	89.14	94.94 ±0.05	**97.57 ±0.11**	89.16 ±0.0	-	97.44 ±0.03
Stain-Heavy	83.11	79.81 ±1.01	92.71 ±0.13	82.96 ±0.02	-	**95.07 ±0.02**
Dust	92.93	96.13 ±0.16	**98.78 ±0.01**	92.87 ±0.08	-	96.60 ±0.21
Air Bubble	83.59	83.38 ±0.63	**95.56 ±0.23**	83.58 ±0.08	-	90.66 ±0.13
Defocus Blur	57.57	50.00 ±0.0	50.86 ±0.42	56.50 ±0.0	-	**92.51 ±0.10**
Motion Blur	56.28	50.00 ±0.0	50.03 ±0.03	55.69 ±0.15	-	**94.46 ±0.07**
Gaussian Noise	66.16	50.04 ±0.01	56.03 ±0.42	66.10 ±0.05	-	**98.02 ±0.01**
Shot Noise	54.89	50.00 ±0.0	50.39 ±0.39	55.70 ±0.09	-	**96.84 ±0.06**
Brightness	95.58	97.45 ±0.08	**98.76 ±0.11**	95.52 ±0.0	-	97.57 ±0.12
Contrast	50.34	50.00 ±0.0	50.00 ±0.0	50.09 ±0.0	-	**84.70 ±0.90**
Mean	72.96	70.18	74.07	72.82	-	**94.39**

Table 12. Comparison of adaptation methods on LC25K-Colon (2 classes) using PathGen [9] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

VLM: PathGen Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-ALL	83.25	90.47 ±0.13	83.74 ±0.09	82.38 ±0.01	87.40 ±0.04	**93.27 ±0.01**
LC25K-ALL-C Stain-Light	73.17	80.43 ±0.09	77.24 ±0.35	73.25 ±0.01	81.20 ±0.05	**91.10 ±0.09**
Stain-Heavy	57.83	54.13 ±0.03	62.67 ±0.09	57.83 ±0.0	75.37 ±0.07	**82.59 ±0.16**
Dust	77.90	84.48 ±0.11	80.58 ±0.44	77.67 ±0.19	84.00 ±0.06	**91.30 ±0.07**
Air Bubble	60.99	60.13 ±0.04	65.34 ±0.54	59.91 ±0.07	69.98 ±0.11	**82.70 ±0.13**
Defocus Blur	45.27	38.73 ±0.0	40.18 ±0.07	44.24 ±0.0	54.64 ±0.09	**80.48 ±0.03**
Motion Blur	40.34	24.41 ±0.05	38.64 ±0.12	39.61 ±0.08	64.98 ±0.02	**83.98 ±0.11**
Gaussian Noise	50.92	28.10 ±0.09	48.11 ±0.22	51.08 ±0.08	78.68 ±0.04	**87.64 ±0.10**
Shot Noise	45.25	20.02 ±0.01	46.20 ±0.02	45.92 ±0.07	67.94 ±0.06	**83.99 ±0.18**
Brightness	84.35	89.58 ±0.10	90.50 ±0.51	84.21 ±0.0	85.37 ±0.01	**91.82 ±0.08**
Contrast	20.56	20.00 ±0.0	20.00 ±0.0	20.16 ±0.0	43.00 ±0.07	**45.23 ±0.19**
Mean	55.66	50.00	56.95	55.39	70.52	**82.08**

Table 13. Comparison of adaptation methods on LC25K-All (5 classes) using PathGen [9] as the base VLM.

VLM: PathGen Dataset		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Skin		55.26	56.02 ±0.02	63.52 ±0.06	54.71 ±0.0	63.22 ±0.13	**68.34 ±0.12**
Skin-C	Stain-Light	27.44	13.55 ±0.06	20.91 ±0.47	27.33 ±0.0	49.33 ±0.02	**61.42 ±0.21**
	Stain-Heavy	14.69	2.95 ±0.01	5.92 ±0.02	15.62 ±0.0	33.27 ±0.07	**52.58 ±0.03**
	Dust	52.72	56.19 ±0.07	56.98 ±0.12	52.62 ±0.16	60.01 ±0.09	**66.81 ±0.12**
	Air Bubble	35.77	21.48 ±0.03	37.05 ±0.15	35.25 ±0.14	47.24 ±0.18	**56.69 ±0.14**
	Defocus Blur	26.16	21.16 ±0.10	16.11 ±0.22	25.07 ±0.0	37.34 ±0.06	**48.42 ±0.13**
	Motion Blur	26.94	20.82 ±0.09	24.10 ±0.34	26.44 ±0.0	38.33 ±0.19	**52.20 ±0.02**
	Gaussian Noise	9.92	2.55 ±0.07	4.05 ±0.06	10.14 ±0.08	22.10 ±0.06	**31.26 ±0.02**
	Shot Noise	6.02	1.61 ±0.02	1.76 ±0.02	5.81 ±0.07	16.82 ±0.07	**30.14 ±0.02**
	Brightness	46.97	45.31 ±0.11	52.13 ±0.02	46.67 ±0.0	52.01 ±0.03	**63.23 ±0.16**
	Contrast	15.61	3.56 ±0.05	18.05 ±0.12	16.83 ±0.0	24.27 ±0.0	**43.73 ±0.17**
	Mean	26.22	18.92	23.71	26.18	38.07	**50.65**

Table 14. Comparison of adaptation methods on Skin (16 classes) using PathGen [9] as the base VLM.

VLM: PathGen Dataset		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Renal		48.61	49.66 ±0.02	50.28 ±0.06	48.67 ±0.0	50.42 ±0.01	**56.81 ±0.17**
Renal-C	Stain-Light	42.56	45.01 ±0.04	45.98 ±0.09	41.40 ±0.0	**47.94 ±0.17**	44.97 ±0.06
	Stain-Heavy	40.55	44.75 ±0.01	42.62 ±0.20	40.50 ±0.0	**49.56 ±0.04**	39.45 ±0.11
	Dust	49.33	47.42 ±0.02	47.22 ±0.04	48.76 ±0.07	51.07 ±0.04	**55.18 ±0.15**
	Air Bubble	35.00	7.29 ±0.05	38.46 ±0.15	35.94 ±0.09	**46.26 ±0.04**	44.53 ±0.02
	Defocus Blur	6.54	1.96 ±0.0	2.14 ±0.01	6.36 ±0.0	20.73 ±0.06	**49.45 ±0.29**
	Motion Blur	11.33	2.25 ±0.0	2.61 ±0.06	10.63 ±0.08	21.49 ±0.03	**45.18 ±0.25**
	Gaussian Noise	9.35	1.95 ±0.0	1.98 ±0.01	9.54 ±0.01	28.78 ±0.04	**39.98 ±0.05**
	Shot Noise	3.54	1.95 ±0.0	1.95 ±0.0	3.32 ±0.01	14.26 ±0.11	**38.66 ±0.13**
	Brightness	25.10	2.77 ±0.03	8.98 ±0.07	25.42 ±0.01	42.75 ±0.17	**45.10 ±0.27**
	Contrast	7.13	2.71 ±0.02	3.93 ±0.06	6.32 ±0.0	24.00 ±0.09	**38.32 ±0.22**
	Mean	23.04	15.81	19.59	22.82	34.68	**44.08**

Table 15. Comparison of adaptation methods on Renal (5 classes) using PathGen [9] as the base VLM.

VLM: PathGen Dataset		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
MHIST		57.11	39.00 ±0.31	46.21 ±0.97	57.22 ±0.0	-	**60.49 ±0.31**
MHIST-C	Stain-Light	55.99	36.95 ±0.0	41.10 ±0.05	56.29 ±0.0	-	**62.69 ±0.46**
	Stain-Heavy	56.50	37.26 ±0.0	41.86 ±0.72	56.50 ±0.0	-	**61.98 ±0.05**
	Dust	58.55	45.50 ±1.38	47.80 ±0.72	58.19 ±0.26	-	**59.83 ±0.56**
	Air Bubble	54.45	**63.15 ±0.0**	56.14 ±0.67	52.87 ±0.05	-	44.93 ±0.92
	Defocus Blur	57.32	**63.15 ±0.0**	**63.15 ±0.0**	57.63 ±0.0	-	51.38 ±0.41
	Motion Blur	**45.14**	36.85 ±0.0	42.17 ±1.23	43.19 ±0.20	-	43.60 ±0.10
	Gaussian Noise	53.22	37.31 ±0.05	42.48 ±0.92	52.71 ±0.10	-	**60.18 ±0.20**
	Shot Noise	57.01	40.63 ±1.74	53.02 ±1.64	57.83 ±0.61	-	**60.95 ±0.05**
	Brightness	62.85	**63.15 ±0.0**	62.79 ±0.77	62.69 ±0.05	-	59.67 ±0.31
	Contrast	51.89	36.85 ±0.0	44.93 ±1.64	51.48 ±0.0	-	**56.96 ±0.15**
	Mean	55.29	46.08	49.54	54.94	-	**56.22**

Table 16. Comparison of adaptation methods on MHIST (2 classes) using PathGen [9] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
NCT7K	67.55	68.66 ±0.06	73.59 ±0.29	67.41 ±0.0	58.28 ±0.14	**82.12** ±0.24
NCT7K-C Stain-Light	56.80	58.36 ±0.09	58.81 ±0.13	56.64 ±0.0	56.53 ±0.05	**71.36** ±0.18
Stain-Heavy	28.30	29.88 ±0.39	29.31 ±0.0	28.14 ±0.0	37.83 ±0.12	**68.77** ±0.01
Dust	66.92	67.74 ±0.14	72.92 ±0.03	66.97 ±0.19	60.86 ±0.09	**79.89** ±0.01
Air Bubble	52.50	51.83 ±0.24	58.56 ±0.69	52.06 ±0.02	42.69 ±0.34	**77.39** ±0.02
Defocus Blur	40.27	36.86 ±0.07	42.21 ±0.03	40.23 ±0.0	49.01 ±0.05	**67.86** ±0.04
Motion Blur	40.85	37.11 ±0.33	42.13 ±0.41	40.97 ±0.02	44.80 ±0.09	**77.02** ±0.13
Gaussian Noise	12.38	18.47 ±0.09	8.68 ±0.24	12.50 ±0.09	13.12 ±0.19	**65.05** ±0.09
Shot Noise	4.01	9.75 ±0.13	1.18 ±0.03	3.93 ±0.06	11.21 ±0.02	**55.82** ±0.19
Brightness	49.31	45.29 ±0.08	48.27 ±0.32	49.30 ±0.0	48.51 ±0.05	**81.98** ±0.16
Contrast	22.19	26.65 ±0.03	24.50 ±0.04	22.14 ±0.0	26.30 ±0.15	**33.96** ±0.11
Mean	37.35	38.19	38.66	37.29	39.09	**67.91**

Table 17. Comparison of different adaptation methods on the NCT-7k dataset (8 classes) using CONCH [7] as the base VLM.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
NCT100K	63.79	65.00 ±0.02	71.82 ±0.04	63.73 ±0.0	59.66 ±0.08	**82.76** ±0.01
NCT100K-C Stain-Light	55.71	56.17 ±0.02	60.76 ±0.02	55.62 ±0.0	56.33 ±0.01	**73.41** ±0.02
Stain-Heavy	22.37	18.54 ±0.02	21.21 ±0.04	22.37 ±0.0	31.37 ±0.05	**63.92** ±0.02
Dust	61.15	62.41 ±0.09	68.23 ±0.04	61.56 ±0.01	57.15 ±0.02	**79.90** ±0.03
Air Bubble	49.11	49.63 ±0.05	57.76 ±0.29	49.02 ±0.09	50.25 ±0.06	**71.74** ±0.02
Defocus Blur	27.14	18.79 ±0.03	25.30 ±0.07	26.82 ±0.0	40.16 ±0.03	**72.14** ±0.09
Motion Blur	28.95	21.09 ±0.03	27.96 ±0.04	28.44 ±0.05	38.13 ±0.02	**70.86** ±0.01
Gaussian Noise	11.17	14.25 ±0.05	4.55 ±0.03	11.45 ±0.05	19.72 ±0.0	**57.66** ±0.11
Shot Noise	3.97	12.05 ±0.02	1.06 ±0.07	3.94 ±0.02	13.37 ±0.0	**47.19** ±0.09
Brightness	20.04	50.11 ±0.03	55.42 ±0.03	50.02 ±0.0	47.96 ±0.02	**78.88** ±0.11
Contrast	**50.19**	20.16 ±0.0	20.11 ±0.0	20.20 ±0.0	15.63 ±0.04	19.63 ±0.02
Mean	32.98	32.32	34.24	32.94	37.01	**63.53**

Table 18. Comparison of different adaptation methods on the NCT-100k dataset (8 classes) using CONCH [7] as the base VLM.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Lung	87.95	90.22 ±0.06	92.67 ±0.01	87.45 ±0.01	-	**97.02** ±0.12
LC25K-Lung-C Stain-Light	77.00	78.00 ±0.01	79.78 ±0.16	77.18 ±0.0	-	**94.94** ±0.08
Stain-Heavy	51.31	40.34 ±0.09	49.37 ±0.24	51.31 ±0.0	-	**82.68** ±0.19
Dust	75.19	75.75 ±0.01	78.88 ±0.12	75.52 ±0.01	-	**96.19** ±0.02
Air Bubble	52.51	38.82 ±0.04	48.03 ±0.07	52.73 ±0.11	-	**94.73** ±0.01
Defocus Blur	53.69	42.55 ±0.14	52.66 ±0.07	53.61 ±0.0	-	**93.20** ±0.04
Motion Blur	60.34	54.04 ±0.49	57.32 ±0.0	60.08 ±0.09	-	**93.15** ±0.13
Gaussian Noise	34.33	33.68 ±0.03	33.54 ±0.05	34.32 ±0.02	-	**82.47** ±0.13
Shot Noise	33.43	33.39 ±0.0	33.34 ±0.0	33.43 ±0.0	-	**76.15** ±0.18
Brightness	87.69	90.08 ±0.18	93.08 ±0.10	87.71 ±0.0	-	**96.46** ±0.04
Contrast	**35.83**	33.33 ±0.0	33.53 ±0.09	35.80 ±0.0	-	27.24 ±0.18
Mean	56.13	52.00	55.95	56.17	-	**83.72**

Table 19. Comparison of adaptation methods on LC25K-Lung (3 classes) using CONCH [7] as the base VLM. CLIPArTT is not applicable due to fewer than three classes.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-Colon	95.31	96.32 ±0.07	97.46 ±0.02	95.15 ±0.0	-	**99.18 ±0.08**
LC25K-Colon-C Stain-Light	89.76	92.27 ±0.06	93.40 ±0.03	89.75 ±0.0	-	**99.20 ±0.08**
Stain-Heavy	58.07	53.16 ±0.30	55.89 ±0.14	58.07 ±0.0	-	**96.52 ±0.20**
Dust	93.92	95.73 ±0.10	97.05 ±0.02	93.81 ±0.07	-	**99.02 ±0.04**
Air Bubble	89.04	92.28 ±0.44	93.03 ±0.24	88.92 ±0.20	-	**99.00 ±0.00**
Defocus Blur	59.32	53.89 ±0.19	57.08 ±0.30	59.34 ±0.0	-	**99.19 ±0.06**
Motion Blur	57.64	55.14 ±0.33	56.01 ±0.16	57.36 ±0.31	-	**99.22 ±0.02**
Gaussian Noise	65.46	59.96 ±0.03	64.48 ±0.08	65.16 ±0.14	-	**97.45 ±0.06**
Shot Noise	53.54	51.07 ±0.08	51.80 ±0.09	53.46 ±0.09	-	**90.66 ±0.13**
Brightness	95.16	95.78 ±0.07	96.96 ±0.09	95.10 ±0.0	-	**99.35 ±0.05**
Contrast	50.00	50.00 ±0.0	50.00 ±0.0	50.00 ±0.0	-	**94.57 ±0.16**
Mean	71.19	69.93	71.57	71.10	-	**97.42**

Table 20. Comparison of adaptation methods on LC25K-Colon (2 classes) using CONCH [7] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
LC25K-All	86.58	87.98 ±0.06	90.55 ±0.01	86.17 ±0.0	89.49 ±0.04	**95.81 ±0.08**
LC25K-All-C Stain-Light	69.46	73.28 ±0.04	74.47 ±0.03	69.51 ±0.0	43.98 ±0.07	**88.90 ±0.03**
Stain-Heavy	37.96	25.34 ±0.13	36.37 ±0.07	37.93 ±0.0	29.42 ±0.04	**73.85 ±0.02**
Dust	80.01	83.83 ±0.16	85.02 ±0.15	80.13 ±0.16	81.19 ±0.03	**95.16 ±0.05**
Air Bubble	60.73	59.55 ±0.02	61.20 ±0.08	60.96 ±0.03	56.05 ±0.18	**93.13 ±0.11**
Defocus Blur	50.35	49.89 ±0.22	50.17 ±0.0	50.29 ±0.0	46.68 ±0.13	**89.77 ±0.20**
Motion Blur	53.15	51.13 ±0.09	51.56 ±0.0	52.99 ±0.12	65.31 ±0.01	**91.89 ±0.08**
Gaussian Noise	31.41	24.25 ±0.07	30.01 ±0.02	31.56 ±0.01	25.31 ±0.10	**82.57 ±0.10**
Shot Noise	25.41	20.10 ±0.0	22.46 ±0.10	25.31 ±0.02	27.31 ±0.03	**63.39 ±0.19**
Brightness	83.22	85.23 ±0.06	88.57 ±0.03	83.06 ±0.0	79.61 ±0.0	**94.17 ±0.09**
Contrast	19.98	20.01 ±0.0	20.00 ±0.0	19.97 ±0.0	25.74 ±0.02	**30.98 ±0.31**
Mean	51.17	49.26	51.98	51.17	48.06	**80.38**

Table 21. Comparison of adaptation methods on LC25K-All (5 classes) using CONCH [7] as the base VLM.

VLM: CONCH Dataset	Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Skin	34.56	33.34 ±0.14	35.98 ±0.12	34.11 ±0.0	31.39 ±0.15	**64.32 ±0.01**
Skin-C Stain-Light	26.86	27.29 ±0.01	27.48 ±0.0	26.88 ±0.0	23.28 ±0.07	**48.99 ±0.02**
Stain-Heavy	18.02	15.98 ±0.04	17.76 ±0.13	18.04 ±0.0	15.10 ±0.03	**40.54 ±0.07**
Dust	30.98	29.67 ±0.01	30.24 ±0.07	30.96 ±0.01	32.19 ±0.09	**62.41 ±0.20**
Air Bubble	22.00	20.42 ±0.10	21.69 ±0.03	21.73 ±0.02	25.10 ±0.07	**52.94 ±0.15**
Defocus Blur	23.48	20.22 ±0.07	23.06 ±0.07	23.47 ±0.0	22.20 ±0.03	**54.91 ±0.09**
Motion Blur	20.75	15.59 ±0.31	19.99 ±0.21	20.61 ±0.0	25.37 ±0.09	**59.95 ±0.26**
Gaussian Noise	19.02	21.14 ±0.05	21.24 ±0.02	18.96 ±0.05	20.85 ±0.01	**37.14 ±0.08**
Shot Noise	15.31	10.62 ±0.02	16.29 ±0.16	15.13 ±0.07	15.31 ±0.09	**30.76 ±0.12**
Brightness	33.44	31.23 ±0.02	32.27 ±0.03	33.36 ±0.0	36.16 ±0.16	**62.89 ±0.15**
Contrast	5.31	4.28 ±0.0	5.06 ±0.02	5.30 ±0.0	2.96 ±0.01	**25.59 ±0.30**
Mean	21.52	19.64	21.51	21.44	21.85	**47.61**

Table 22. Comparison of adaptation methods on Skin (16 classes) using CONCH [7] as the base VLM.

VLM: CONCH Dataset		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
Renal		46.16	48.74 ±0.22	51.26 ±0.20	45.65 ±0.0	52.87 ±0.06	**57.22 ±0.22**
Renal-C	Stain-Light	28.20	22.84 ±0.01	27.28 ±0.04	28.36 ±0.0	34.60 ±0.09	**51.95 ±0.06**
	Stain-Heavy	7.11	2.69 ±0.03	3.23 ±0.06	6.85 ±0.0	10.46 ±0.12	**42.67 ±0.01**
	Dust	43.43	42.31 ±0.07	45.87 ±0.07	43.58 ±0.04	46.88 ±0.18	**59.11 ±0.07**
	Air Bubble	35.22	35.25 ±0.17	39.93 ±0.24	35.35 ±0.03	32.86 ±0.0	**53.90 ±0.09**
	Defocus Blur	27.20	26.24 ±0.11	27.54 ±0.10	27.32 ±0.0	29.02 ±0.01	**51.10 ±0.12**
	Motion Blur	15.76	12.31 ±0.11	14.47 ±0.02	15.62 ±0.09	19.12 ±0.27	**44.94 ±0.03**
	Gaussian Noise	3.54	2.12 ±0.0	2.02 ±0.0	3.23 ±0.02	9.14 ±0.03	**49.01 ±0.20**
	Shot Noise	3.26	1.96 ±0.0	2.00 ±0.0	3.17 ±0.01	7.64 ±0.09	**46.63 ±0.09**
	Brightness	19.47	11.27 ±0.04	15.24 ±0.04	19.30 ±0.0	19.82 ±0.05	**53.48 ±0.09**
	Contrast	23.01	24.56 ±0.06	24.47 ±0.14	23.01 ±0.0	16.88 ±0.05	**37.12 ±0.03**
	Mean	20.62	18.16	20.21	20.58	22.64	**48.99**

Table 23. Comparison of adaptation methods on Renal (5 classes) using CONCH [7] as the base VLM.

VLM: CONCH Dataset		Source	TENT	LAME	TPT	CLIPArTT	LATTE (Ours)
MHIST		59.98	62.59 ±0.56	60.39 ±1.02	59.72 ±0.05	-	**63.00 ±0.36**
MHIST-C	Stain-Light	57.11	49.18 ±0.77	55.22 ±1.59	57.11 ±0.0	-	**63.36 ±0.51**
	Stain-Heavy	53.94	46.42 ±2.41	50.46 ±0.51	53.84 ±0.0	-	**56.65 ±0.15**
	Dust	58.24	**63.15 ±0.0**	61.51 ±0.51	57.47 ±0.56	-	62.64 ±1.23
	Air Bubble	60.70	63.15 ±0.0	62.95 ±0.20	62.33 ±0.10	-	**64.59 ±1.33**
	Defocus Blur	51.28	**63.15 ±0.0**	58.96 ±0.41	52.05 ±0.05	-	56.81 ±0.51
	Motion Blur	61.82	**63.15 ±0.0**	63.00 ±0.26	61.72 ±0.51	-	56.91 ±0.61
	Gaussian Noise	61.72	**63.15 ±0.0**	**63.15 ±0.0**	61.67 ±0.05	-	50.31 ±0.77
	Shot Noise	62.03	**63.15 ±0.0**	**63.15 ±0.0**	62.03 ±0.20	-	51.59 ±0.51
	Brightness	63.25	63.15 ±0.0	**64.07 ±0.10**	63.25 ±0.0	-	58.03 ±1.43
	Contrast	48.82	36.85 ±0.0	45.34 ±0.31	**49.13 ±0.0**	-	45.04 ±0.41
	Mean	57.89	57.45	**58.78**	58.06	-	56.59

Table 24. Comparison of adaptation methods on MHIST (2 classes) using CONCH [7] as the base VLM. CLIPArTT is not applicable due to binary classification constraints.

References

[1] Andrew A Borkowski, Marilyn M Bui, L Brannon Thomas, Catherine P Wilson, Lauren A DeLand, and Stephen M Mastorides. Lung and colon cancer histopathological image dataset (lc25000). *arXiv preprint arXiv:1912.12142*, 2019. 1

[2] Malik Boudiaf, Romain Mueller, Ismail Ben Ayed, and Luca Bertinetto. Parameter-free online test-time adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8344–8353, 2022. 2

[3] Otso Brummer, Petri Pölönen, Satu Mustjoki, and Oscar Brück. Integrative analysis of histological textures and lymphocyte infiltration in renal cell carcinoma using deep learning. *bioRxiv*, pages 2022–08, 2022. 1

[4] Wisdom Ikezogwo, Saygin Seyfioglu, Fatemeh Ghezloo, Dylan Geva, Fatwir Sheikh Mohammed, Pavan Kumar Anand, Ranjay Krishna, and Linda Shapiro. Quilt-1m: One million image-text pairs for histopathology. *Advances in neural information processing systems*, 36, 2024. 1, 3, 4, 5

[5] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, 2018. 1

[6] Katharina Kriegsmann, Frithjof Lobers, Christiane Zgorzelski, Jörg Kriegsmann, Charlotte Janßen, Rolf Rüdinger Meliß, Thomas Muley, Ulrich Sack, Georg Steinbuss, and Mark Kriegsmann. Deep learning for the detection of anatomical tissue structures and neoplasms of the skin on scanned histopathological tissue sections. *Frontiers in Oncology*, 12: 1022967, 2022. 1

[7] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024. 1, 2, 8, 9, 10

[8] Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, and Chaowei Xiao. Test-time prompt tuning for zero-shot generalization in vision-language models. In *Advances in Neural Information Processing Systems*, pages 14274–14289. Curran Associates, Inc., 2022. 2

[9] Yuxuan Sun, Yunlong Zhang, Yixuan Si, Chenglu Zhu, Zhongyi Shui, Kai Zhang, Jingxiong Li, Xingheng Lyu, Tao Lin, and Lin Yang. Pathgen-1.6m: 1.6 million pathology image-text pairs generation through multi-agent collaboration, 2024. 1, 2, 5, 6, 7

[10] Gustavo Adolfo Vargas Hakim, David Osowiechi, Mehrdad Noori, Milad Cheraghalikhani, Ali Bahri, Moslem Yazdanpanah, Ismail Ben Ayed, and Christian Desrosiers. Clipartt: Light-weight adaptation of clip to new domains at test time, 2024. 1, 2

[11] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno A. Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 2

[12] Jerry Wei, Arief Suriawinata, Bing Ren, Xiaoying Liu, Mikhail Lisovsky, Louis Vaickus, Charles Brown, Michael Baker, Naofumi Tomita, Lorenzo Torresani, et al. A petri dish for histopathology image analysis. In *Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15–18, 2021, Proceedings*, pages 11–24. Springer, 2021. 1