Rotation-Agnostic Image Representation Learning for Digital Pathology

(Supplementary File)

Figures

Tables

Algorithms

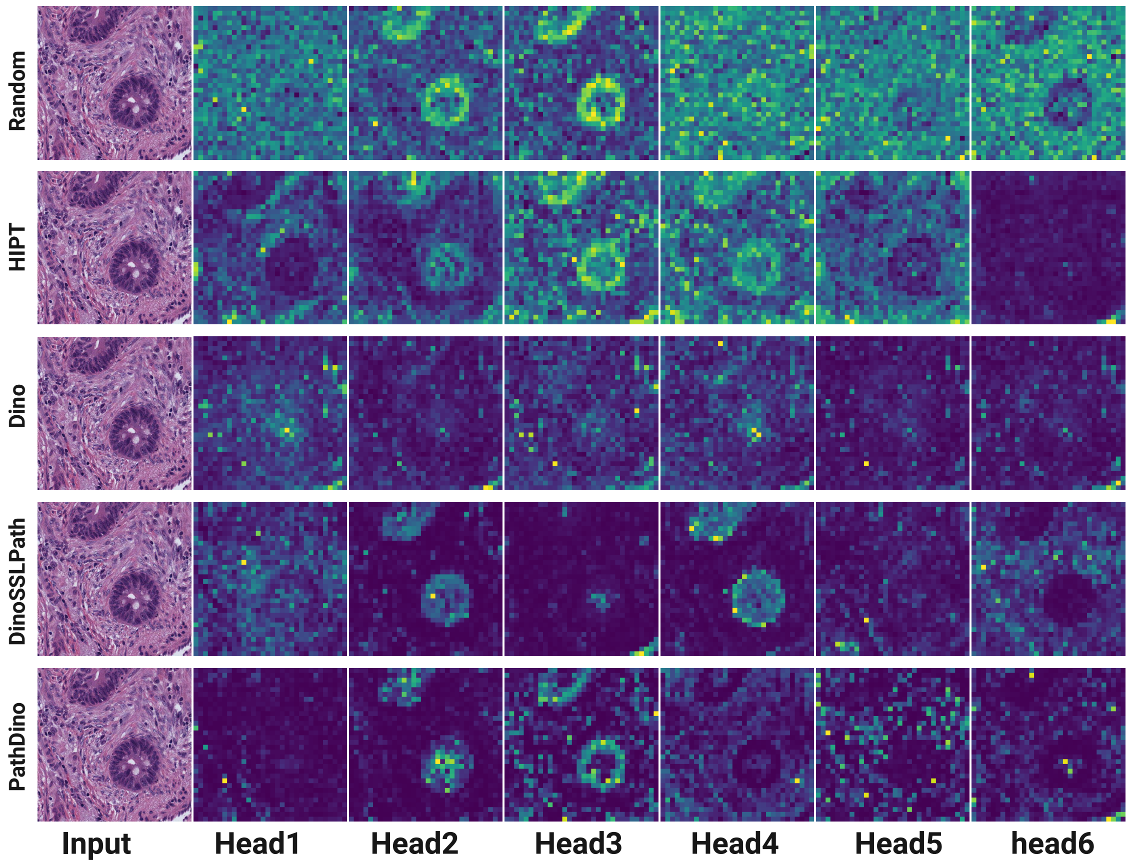

Figure S1. Attention visualization of the proposed PathDino as compared to other ViT-samll models.

Table S1. Summary of Model Attributes. It is worth noting that DinoV2, CLIP, and DinoSSLPath also trained with CNN-based backbones such as ResNet50, however, we only employ the Transformer-based backbones in our comparisons. Floating Point Operations Per Second (FLOPs) are used to quantify the computational complexity of models. Note that we specified the data size used for the exact pre-trained models in our study rather than the general mentioned in the corresponding papers. For example, DinoSSLPath [1] was trained on 32.6M patches (19M from TCGA and 13.6M from a TULIP, which is a private dataset), however, the publically available pre-trained models are trained only on TCGA. Thus, we report 19M for pretraining DinoSSLPath, SwAV-ResNet50, MoCoV2-ResNet50, and Barlow-Twins-ResNet50. PubMed consists of a total of $15,282,336$, however, the training set is 13.9M.

	Model	Pretrained On	Pretraining Domain	Modality	Training Data Size	Learning Paradigm	Input Dim.	Output Embedding	Model Size	FLOPs
CNN	ResNet50 - [2]	ImageNet-1k	Natural Images	Image	$1,200,000$	Supervised	224×224	2048	$23,527,264$	$4,374,897,664$
	DenseNet121 - [3]	ImageNet-1k	Natural Images	Image	$1,200,000$	Supervised	224×224	1024	$6,870,208$	$2,833,364,480$
	EfficientNet-b3-288 - [4]	ImageNet-1k	Natural Images	Image	$1,200,000$	Supervised	288×288	1536	$10,608,936$	$1,587,788,048$
	EfficientNet-b5 - [4]	ImageNet-1k	Natural Images	Image	$1,200,000$	Supervised	448×448	2048	$28,168,048$	$9,402,729,536$
	ConvNext-B [5]	ImageNet-21k	Natural Images	Image	$14,000,000$	Supervised	224×224	1024	$87,510,272$	$15,353,709,568$
	ConvNext-xlarge - [5]	ImageNet-21k	Natural Images	Image	$14,000,000$	Supervised	224×224	2048	$348,035,584$	$60,918,990,848$
	SwAV-ResNet50 - [1]	TCGA	Histopathology Images	Image	$19,000,000$	Self-supervised	1024×768	2048	$23,508,032$	$64,757,958,656$
	MoCoV2-ResNet50 - [1]	TCGA	Histopathology Images	Image	$19,000,000$	Self-supervised	1024×768	2048	$23,508,032$	$64,757,958,656$
	MuDiPath-ResNet50 - [6]	TCGA	Histopathology Images	Image	$882,800$	Supervised	224×224	2048	$25,557,032$	$4,131,592,192$
	MuDiPath-DenseNet-101 - [6]	TCGA	Histopathology Images	Image	$882,800$	Supervised	224×224	1024	$6,953,856$	$2,895,983,104$
	KimiaNet - [7]	TCGA	Histopathology Images	Image	$240,000$	Supervised	1000×1000	1024	$6,953,856$	$57,471,584,640$
	Barlow-Twins-ResNet50 - [1]	TCGA	Histopathology Images	Image	$19,000,000$	Self-supervised	1024×768	2048	$23,508,032$	$64,757,958,656$
Transformers	ViT-B16 [8]	ImageNet-1K	Natural Images	Image	$1,200,000$	Supervised	224×224	768	$85,646,592$	$16,862,862,336$
	DinoV1-ViT-s16 - [9]	ImageNet-1k	Natural Images	Image	$1,200,000$	Self-supervised	224×224	384	$21,589,632$	$4,248,399,360$
	DinoV1-ViT-b16 - [9]	ImageNet-1k	Natural Images	Image	$1,200,000$	Self-supervised	224×224	768	$85,646,592$	$16,862,862,336$
	DinoV2-ViT-b14 - [10]	Internet	Natural Images	Image	$142,000,000$	Self-Supervised	224×224	768	$85,508,352$	$21,963,549,696$
	CLIP - ViT-B/16 - [11]	Internet	Natural Image-Text	Image-Text	$400,000,000$	Contrastive Learning	224×224	512	$85,646,592$	$16,862,862,336$
	BiomedCLIP - [12]	PMC-15M	Medical (PubMed)	Image-Text	$13,900,00$	Contrastive Learning	224×224	512	$85,646,592$	$16,862,862,336$
	HIPT-ViT-s16 [13]	TCGA	Histopathology Images	Image	$104,000,000$	Self-supervised	256×256	384	$21,589,632$	$5,542,417,920$
	PLIP [14]	OpenPath	Histopathology (Twitter)	Image-Text	$208,414$	Contrastive Learning	224×224	512	$85,646,592$	$16,862,862,336$
	iBOT-Path [15]	TCGA	Histopathology Images	Image	$40,000,000$	Self-supervised	224×224	768	$85,646,592$	$16,862,862,336$
	DinoSSLPathology-8 [1]	TCGA	Histopathology Images	Image	$19,000,000$	Self-supervised	224×224	384	$21,368,448$	$16,756,372,992$
	PathDino-224 (ours)	TCGA	Histopathology Images	Image	$2,118,068$	Self-supervised	224×224	384	$9,168,384$	$1,804,061,184$
	PathDino-512 (ours)	TCGA	Histopathology Images	Image	$6,087,558$	Self-supervised	512×512	384	$9,168,384$	$9,387,852,288$

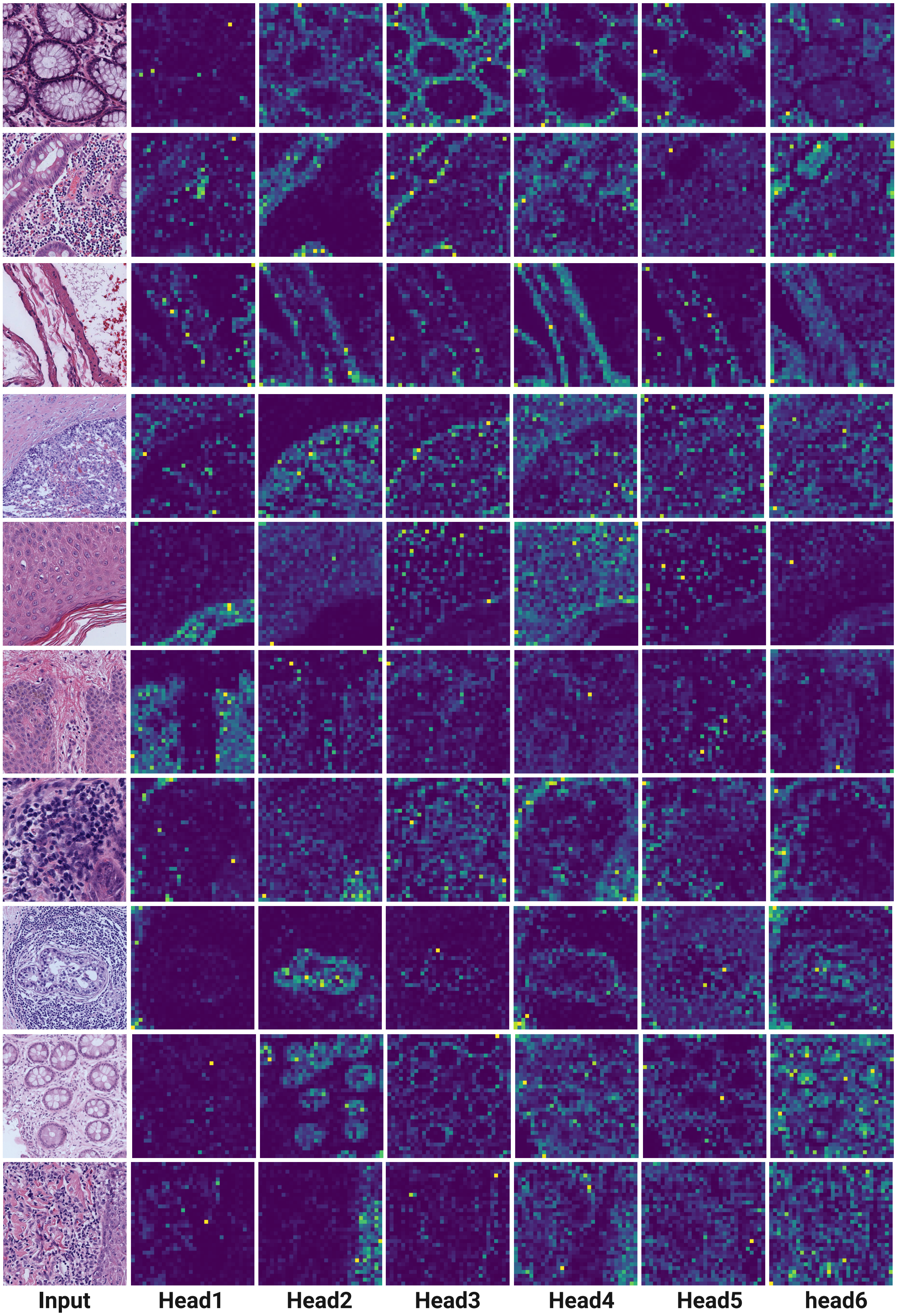

Figure S2. Attention visualization of the proposed PathDino transformer heads.

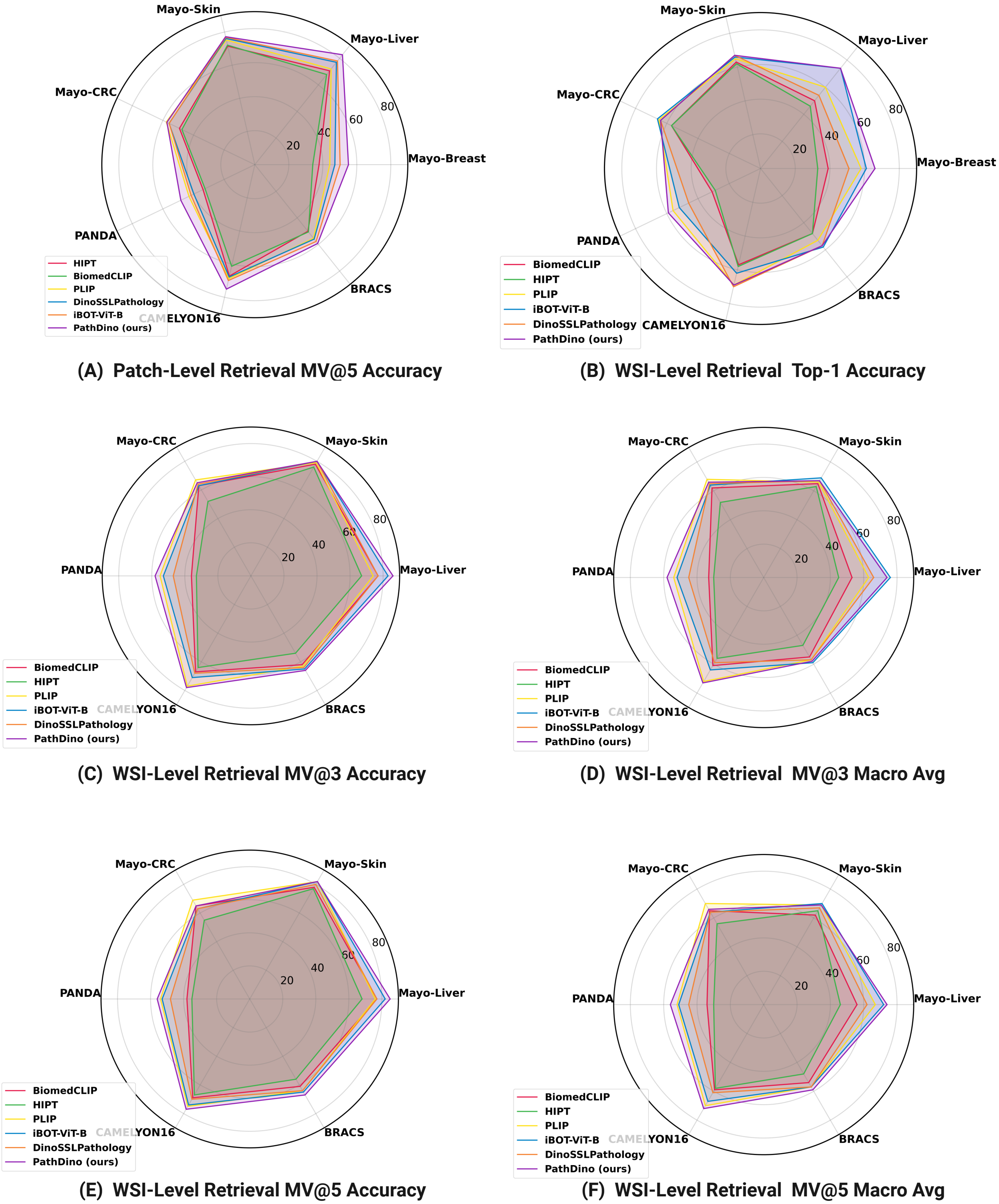

Figure S3. Performance of selected Transformer-based histopathology feature extractors including HIPT, BiomedCLIP, PLIP, DinoSSLPath, iBOT, and PathDino. The performance is represented as the macro average of the F_1 score for the MV@5 in the patch-level retrieval settings.

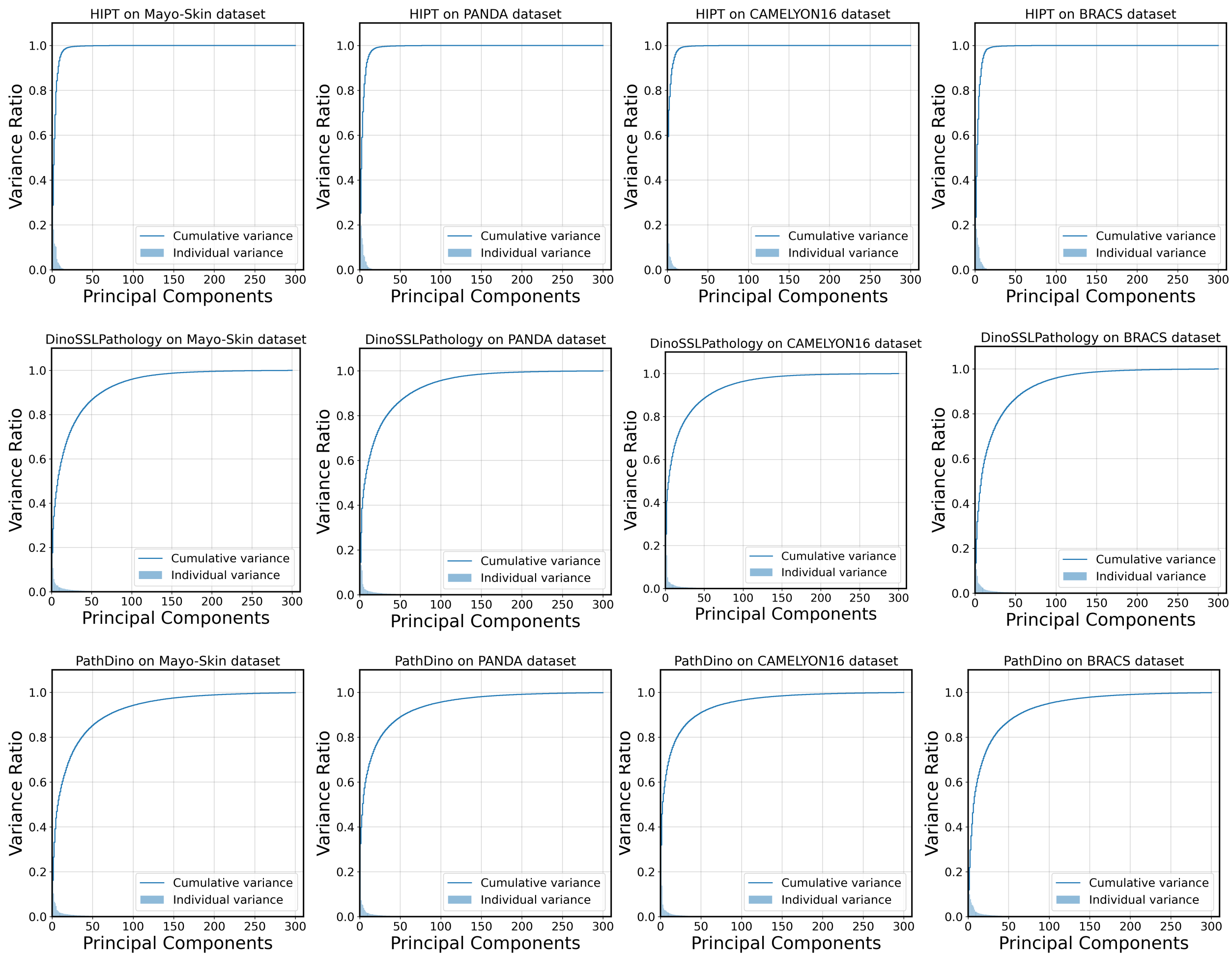

Figure S4. Embedding variance analysis of three selected Transformer-based histopathology feature extractors which have output vector size of 384 including HIPT, DinoSSLPath, and our PathDino.

Table S2. (accuracy) Patient matching outcomes across four internal and three public datasets for classification, subtyping, and grading tasks, utilizing the Yottixel framework and leave-one-patient-out method.

	Dataset		DinoV2 [10]		CLIP [11]		BiomedCLIP [12]		PLIP [14]		KimiaNet [7]		DinoSSLPath [1]		PathDino	
			Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS
Internal Data	Private-Breast	Top 1	36	**53**	45	**49**	47	**47**	55	**58**	**56**	51	**59**	58	63	**68**
	Private-Liver	Top 1	71	**71**	63	**67**	70	**74**	70	**73**	76	**78**	**75**	74	81	**83**
		MV@3	73	**74**	69	**69**	75	**77**	**76**	74	79	**80**	77	**77**	86	**86**
		MV@5	70	**72**	**72**	71	74	**77**	73	**76**	**80**	78	**81**	77	**87**	85
	Private-Skin	Top 1	59	**72**	68	**75**	68	**75**	72	**75**	**78**	75	**79**	78	**79**	78
		MV@3	66	**77**	73	**79**	73	**78**	77	**79**	81	**81**	80	**80**	**81**	80
		MV@5	68	**87**	77	**80**	76	**78**	80	**82**	82	**82**	80	**80**	81	**82**
	Private-CRC	Top 1	52	**56**	**57**	56	55	**58**	60	**64**	60	**64**	60	**63**	57	**63**
		MV@3	52	**59**	**59**	58	60	**63**	61	**67**	60	**65**	61	**64**	60	**65**
		MV@5	55	**59**	56	**60**	59	**65**	62	**69**	60	**62**	63	**63**	61	**65**
Public Data	PANDA [16]	Top 1	**35**	31	35	**36**	33	**34**	53	**56**	**58**	57	**48**	47	**59**	58
		MV@3	**33**	32	37	**38**	36	**36**	53	**55**	**58**	56	**50**	47	58	**58**
		MV@5	35	**35**	39	**40**	38	**38**	53	**54**	**56**	54	**51**	48	**58**	56
	CAMELYON16 [17]	Top 1	57	**61**	66	**67**	60	**61**	70	**73**	75	**76**	62	**74**	**76**	73
		MV@3	58	**58**	66	**67**	58	**67**	71	**77**	72	**78**	**74**	68	77	**78**
		MV@5	57	**63**	64	**69**	64	**69**	70	**75**	79	**81**	**75**	70	**78**	77
	BRACS [18]	Top 1	**53**	51	53	**58**	**56**	55	**62**	60	**66**	62	**62**	61	**65**	64
		MV@3	**55**	53	58	**60**	58	**62**	**64**	63	**66**	**64**	61	**64**	65	**66**
		MV@5	**58**	56	59	**61**	59	**61**	**66**	64	**67**	66	62	**64**	66	**67**

Table S3. (Macro Avg) Patient matching outcomes across four internal and three public datasets for retrieval, subtyping, and grading tasks, utilizing the FPS vs Yottixel framework and leave-one-patient-out method on Macro Average F-1 score.

	Dataset		DinoV2 [10]		CLIP [11]		BiomedCLIP [12]		PLIP [14]		KimiaNet [7]		DinoSSLPath [1]		PathDino	
			Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS	Yottixel	FPS
Internal Data	Private-Breast	Top 1	24	**46**	33	**39**	39	**39**	45	**58**	**56**	47	**55**	51	60	**66**
	Private-Liver	Top 1	**61**	49	**52**	46	**59**	50	58	**60**	62	**66**	**65**	54	**76**	74
		MV@3	**59**	42	**50**	47	**68**	53	**68**	63	**67**	62	**69**	66	**83**	74
		MV@5	**54**	53	**56**	52	**64**	56	59	**67**	**65**	61	**74**	62	**81**	74
	Private-Skin	Top 1	54	**65**	62	**66**	61	**63**	63	**65**	**70**	65	**71**	67	**68**	67
		MV@3	58	**66**	65	**66**	62	**65**	65	**66**	70	**71**	**68**	67	**69**	67
		MV@5	58	**66**	67	**67**	**65**	62	67	**69**	**70**	69	66	**67**	68	**69**
	Private-CRC	Top 1	51	**56**	54	**56**	54	**57**	**59**	65	60	**64**	60	**64**	58	**64**
		MV@3	49	**59**	55	**58**	58	**62**	60	**68**	61	**66**	61	**65**	60	**66**
		MV@5	51	**59**	50	**60**	57	**65**	61	**70**	60	**63**	63	**64**	60	**66**
Public Data	PANDA [16]	Top 1	**31**	28	32	**34**	31	**31**	53	**56**	**59**	58	**47**	46	59	**59**
		MV@3	**30**	29	33	**35**	32	**33**	52	**54**	**58**	56	**48**	45	58	**58**
		MV@5	**31**	30	35	**35**	33	**34**	51	**52**	**55**	54	**48**	45	**57**	56
	CAMELYON16 [17]	Top 1	56	**57**	**65**	61	**59**	57	68	**69**	72	**74**	57	**70**	**73**	69
		MV@3	**56**	50	**64**	58	55	**61**	68	**72**	68	**74**	**69**	59	**74**	73
		MV@5	54	**54**	**62**	61	**61**	59	65	**70**	75	**77**	**70**	61	**73**	72
	BRACS [18]	Top 1	**48**	45	47	**52**	48	**48**	**56**	53	**59**	57	54	**57**	**59**	57
		MV@3	**48**	46	51	**54**	49	**55**	**59**	57	**59**	57	53	**57**	58	**58**
		MV@5	**50**	49	53	**55**	49	**54**	**60**	57	58	**59**	53	**57**	57	**59**

Table S4. Quantitative Assessment via 5-Fold Cross-Validation: Accuracy in Patch-Level Classification in Histopathological Image Analysis.

			Internal Datasets				Public Datasets						
			Private-Breast	Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]	DigestPath [19]	Kather [20]	PanNuke [21]	WSSS4LUAD [22]
Pret. on Natural Data	CNN-based	ResNet50 [2]	62.86±2.68	71.77±3.06	74.88±1.23	53.84±2.63	37.22±0.76	66.30±1.88	54.97±0.21	92.18±2.79	98.82±0.36	77.75±1.80	81.79±0.65
		DenseNet121 [3]	57.75±1.44	72.24±2.06	72.71±5.13	51.22±2.38	34.66±1.61	65.79±4.86	47.92±4.48	91.85±3.12	98.06±0.24	73.27±1.92	82.35±0.92
		EfficientNet-b3-288 [4]	57.99±2.12	74.00±1.13	74.76±0.80	57.52±1.34	37.19±0.59	63.88±2.11	56.51±0.63	91.25±2.58	98.04±0.33	72.49±2.60	83.78±0.45
		EfficientNet-b5 [4]	75.31±0.74	78.94±4.34	78.67±2.78	63.73±3.06	40.00±0.81	70.16±5.25	61.48±0.76	92.79±2.68	99.03±0.20	80.29±1.51	84.44±3.05
		ConvNext-b-224 [5]	73.24±1.54	78.02±1.20	77.09±0.91	57.33±2.93	38.69±0.97	69.46±2.11	57.87±1.59	93.88±1.98	99.35±0.11	82.68±1.56	85.18±0.58
		ConvNext-xlarge [5]	76.88±1.23	79.22±0.99	79.96±0.86	60.53±1.98	38.91±1.95	70.13±3.64	57.85±1.10	94.84±2.16	99.72±0.12	87.53±1.76	87.54±0.93
	Transformer	ViT-b16-224 [8]	53.27±3.03	76.17±1.82	75.79±1.15	51.58±4.03	35.08±1.06	64.99±3.41	52.03±1.13	90.76±2.56	99.04±0.36	79.09±1.82	78.26±8.70
		DinoV1-ViT-s16 [9]	66.85±2.62	76.67±2.60	70.49±6.51	54.00±2.24	33.24±1.40	69.66±3.69	52.91±2.00	95.47±1.59	99.35±0.22	82.08±2.19	83.87±2.88
		DinoV1-ViT-b16 [9]	67.54±1.67	79.86±1.83	73.11±3.10	57.93±2.38	31.65±3.64	69.17±4.00	54.70±4.22	94.75±1.99	99.35±0.18	85.25±2.45	86.63±1.48
		DinoV2-ViT-b14 [10]	59.91±2.90	74.97±5.76	71.59±5.64	55.18±2.15	34.51±2.55	64.60±6.80	50.75±4.31	89.88±6.30	98.22±0.72	70.74±3.08	73.19±8.17
		CLIP - ViT-B/16 [11]	55.04±2.89	80.55±1.29	78.65±1.87	59.43±5.62	40.44±0.77	71.16±3.47	59.46±2.21	88.75±3.24	98.65±0.29	72.03±2.77	82.95±1.30
Pret. on Histopathology Data	CNN-based	Barlow-Twins-ResNet50 [1]	76.15±3.96	85.92±6.54	82.80±1.46	**71.54±5.05**	**45.50±1.87**	76.77±5.48	**66.58±2.36**	89.11±6.83	99.53±0.25	78.14±3.66	89.16±0.96
		MoCoV2-ResNet50 [1]	79.59±1.81	84.37±0.62	79.77±0.49	69.40±0.54	44.10±0.35	74.94±3.01	63.83±0.57	95.88±1.63	96.60±0.75	61.68±2.81	83.07±1.08
		MuDiPath-ResNet50 [6]	53.08±6.68	77.32±1.80	74.38±2.08	55.29±2.59	37.40±1.00	65.32±6.19	50.87±3.07	90.49±1.25	98.97±0.21	79.24±1.32	83.47±1.01
		MuDiPath-DenseNet-101 [6]	54.06±6.78	74.58±1.43	73.97±2.35	54.07±3.09	34.32±2.26	64.65±3.33	51.13±5.84	91.86±1.75	98.80±0.26	75.87±1.28	81.39±2.15
		KimiaNet [7]	77.03±4.13	**89.25±1.46**	**83.45±2.00**	69.58±3.74	38.74±5.74	79.48±3.34	64.75±3.70	94.58±2.40	99.26±0.21	85.42±1.68	82.48±2.41
	Transformer	BiomedCLIP [12]	50.76±3.93	73.50±1.34	75.62±0.59	56.13±2.23	37.41±0.31	67.44±1.62	56.53±0.83	93.79±1.48	96.09±0.67	64.14±2.01	84.39±1.11
		HIPT-ViT-s16 [13]	53.07±6.57	74.93±3.75	74.36±2.37	50.53±2.93	33.70±3.27	65.63±5.92	52.05±5.54	88.27±4.73	97.58±0.50	65.52±3.86	83.49±1.10
		TransPath [23]	20.12±10.81	61.89±4.59	65.22±1.48	42.68±2.57	29.95±1.46	56.80±10.84	52.51±4.60	71.78±10.28	78.82±4.92	30.49±11.68	73.65±1.93
		PLIP [14]	56.66±2.72	77.03±1.45	78.81±0.56	64.14±2.18	42.16±0.26	74.16±1.04	61.18±0.42	93.63±2.33	92.52±1.29	60.86±2.26	77.77±1.25
		iBOT-Path [15]	86.92±1.97	86.66±0.60	81.51±0.73	68.33±0.59	40.22±1.03	74.96±1.71	61.66±2.85	**96.73±1.13**	99.83±0.12	98.13±0.57	90.04±0.97
		DinoSSLPathology-8 [1]	81.06±1.31	80.45±4.27	78.56±0.85	59.28±3.96	36.82±0.53	74.21±2.95	58.48±2.02	95.53±1.27	99.74±0.06	92.50±1.35	89.31±0.68
		PathDino-224 (ours)	79.24±3.84	81.85±0.37	77.32±1.36	60.83±1.98	33.30±1.38	70.03±5.43	53.73±3.72	95.39±2.25	99.72±0.14	89.63±1.30	88.34±1.61
		PathDino-512 (ours)	**89.92±2.65**	89.16±2.29	82.01±0.90	70.68±2.36	38.47±2.66	**81.42±1.04**	63.05±4.44	96.52±1.82	99.67±0.08	91.76±2.05	88.83±0.67

Table S5. Quantitative Assessment via 5-Fold Cross-Validation: Macro-Averaged F-1 Score for Patch-Level Classification in Histopathological Image Analysis.

			Internal Datasets				Public Datasets						
			Private-Breast	Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]	DigestPath [19]	Kather [20]	PanNuke [21]	WSSS4LUAD [22]
Pret. on Natural Data	CNN-based	ResNet50 [2]	55.80 ±2.19	54.77 ±4.96	61.31 ±0.76	53.69 ±2.98	27.90 ±1.17	64.10 ±1.81	45.96 ±1.80	90.13±2.86	98.54 ±0.49	61.37±2.16	65.30 ±6.95
		DenseNet121 [3]	51.84 ±3.32	54.63 ±5.78	58.42 ±1.49	50.54 ±2.78	26.14 ±1.47	62.19 ±3.57	40.48 ±1.90	90.31±4.10	97.43 ±0.34	54.56±1.68	69.54 ±8.54
		EfficientNet-b3-288 [4]	51.17 ±3.01	61.43 ±2.78	63.73 ±0.68	57.58 ±1.35	27.60 ±0.81	61.04 ±1.50	47.85 ±0.88	89.35±3.74	97.63 ±0.24	53.60±2.94	70.04 ±8.42
		EfficientNet-b5 [4]	70.41 ±2.86	72.91 ±4.75	68.18 ±2.76	63.12 ±3.57	30.56 ±2.27	66.52 ±3.83	53.19 ±0.86	91.20±3.88	98.86 ±0.30	65.45±2.07	71.69 ±9.74
		ConvNext-b-224 [5]	65.02 ±4.39	67.42 ±1.35	67.65 ±0.88	56.41 ±3.70	27.65 ±0.76	65.96 ±1.76	48.52 ±1.80	92.69±2.50	99.24 ±0.09	69.52±2.37	68.24 ±6.55
		ConvNext-xlarge [5]	70.72 ±1.75	70.88 ±5.03	69.23 ±2.60	60.48 ±1.81	30.31 ±1.12	68.01 ±2.61	49.44 ±3.14	93.72±2.82	99.69 ±0.15	76.40±1.80	74.33 ±8.44
	Transformer	ViT-b16-224 [8]	43.44 ±2.92	66.79 ±2.21	62.84 ±1.52	50.37 ±5.81	25.88 ±2.93	63.11 ±2.14	43.64 ±2.95	82.81±7.25	98.79 ±0.46	64.73±2.94	58.76 ±19.83
		DinoV1-ViT-s16 [9]	60.57 ±2.90	65.24 ±7.94	58.52 ±2.51	51.82 ±4.24	23.62 ±2.60	66.32 ±3.93	39.70 ±4.32	94.06±2.12	99.16 ±0.23	68.40±2.28	67.65 ±7.25
		DinoV1-ViT-b16 [9]	62.80 ±3.65	71.58 ±2.77	62.23 ±2.33	57.44 ±2.85	24.35 ±2.21	65.89 ±4.27	47.74 ±2.11	93.18±2.88	99.18 ±0.20	72.70±2.52	73.69 ±8.64
		DinoV2-ViT-b14 [10]	51.16 ±4.28	63.08 ±6.31	60.34 ±3.60	53.99 ±3.26	25.20 ±3.77	60.70 ±6.66	41.46 ±4.62	86.34±9.14	97.61 ±1.17	50.50±3.99	69.20 ±8.31
		CLIP - ViT-B/16 [11]	48.56 ±1.72	61.76 ±8.24	66.91 ±2.59	57.13 ±8.27	40.44 ±0.77	63.52 ±7.99	51.87 ±3.84	86.58 ±3.12	98.36 ±0.32	50.72 ±3.62	68.04 ±8.32
Pret. on Histopathology Data	CNN-based	Barlow-Twins-ResNet50 [1]	70.94 ±6.64	77.81 ±4.48	71.87 ±4.22	70.55 ±5.79	**37.77 ±3.03**	70.64 ±9.57	59.54 ±2.34	86.53±8.46	99.47 ±0.25	63.77±4.32	72.56 ±8.62
		SwAV-ResNet50 [1]	78.23 ±5.79	84.78 ±5.13	78.91 ±1.06	**74.65 ±2.62**	40.39 ±1.85	78.35 ±5.18	58.17 ±4.98	95.87±1.34	98.51 ±0.42	62.85±3.56	71.43 ±7.35
		MoCoV2-ResNet50 [1]	73.11 ±2.02	66.87 ±6.62	65.51 ±0.93	69.76 ±0.56	33.52 ±0.41	69.47 ±6.53	52.28 ±1.76	94.81±2.11	95.61 ±0.84	34.22±2.72	65.55 ±6.66
		MuDiPath-ResNet50 [6]	45.79 ±7.27	61.64 ±5.17	64.42 ±0.77	54.42 ±3.25	28.44 ± 1.92	63.42 ±4.56	46.83 ±4.10	87.30±4.20	98.77 ±0.28	65.47±2.10	69.75 ±8.82
		MuDiPath-DenseNet-101 [6]	52.62 ±4.34	58.31 ±3.31	59.57 ±3.28	52.94 ±4.08	26.62 ±2.21	62.14 ±3.29	42.14 ±3.97	89.34±2.90	98.47 ±0.33	59.79±2.45	66.46 ±9.08
		KimiaNet [7]	76.01 ±1.76	85.77 ±4.22	76.78 ±1.95	69.56 ±4.00	34.09 ±5.32	77.90 ±2.77	**58.35 ±1.35**	93.22±3.66	99.14 ±0.25	72.57±3.21	65.93 ±10.00
	Transformer	BiomedCLIP [12]	38.82 ±1.64	48.44 ±1.15	56.62 ±0.61	55.89 ±2.57	25.97 ±0.34	58.19 ±4.99	41.89 ±0.93	92.07±2.33	94.89 ±0.84	37.81±2.12	69.98 ±8.12
		HIPT-ViT-s16 [13]	43.08 ±6.27	59.31 ±5.08	59.47 ±2.85	48.69 ±4.62	25.65 ±1.38	61.56 ±4.25	42.02 ±8.82	84.66±7.17	96.81 ±0.69	42.17±3.80	64.26 ±7.50
		TransPath [23]	10.17 ±5.14	39.78 ±5.56	43.79 ±2.94	34.81 ±2.74	14.69 ±2.14	39.23 ±8.06	38.49 ±0.96	58.28±12.80	72.46±4.98	12.16±2.73	50.16 ±10.75
		PLIP [14]	46.07 ±3.20	50.78 ±1.48	62.48 ±1.11	64.11 ±2.70	31.53 ± 0.47	69.67 ±1.45	46.72 ±1.05	92.07±2.91	90.90 ±1.63	27.77±2.54	61.51 ±7.19
		iBOT-Path [15]	85.12 ±1.74	84.37 ± 1.31	**73.09 ± 0.39**	68.38 ± 0.52	32.95 ±0.86	73.76 ± 1.67	56.52 ± 1.96	95.67±2.17	99.81 ± 0.17	95.76±1.78	73.31 ± 5.91
		DinoSSLPathology-8 [1]	77.59 ±2.17	74.25 ±4.56	66.98 ±1.00	58.17 ±4.77	28.75 ±2.31	70.61 ±1.81	46.42 ±5.59	94.50±1.91	99.68 ±0.11	86.17±2.61	76.30 ±9.60
		PathDino-224 (ours)	78.06 ±4.03	74.34 ±4.98	64.89 ±2.14	60.65 ±2.23	27.74 ±2.44	69.26 ±4.94	46.58 ±3.78	94.03±3.06	99.66 ± 0.19	81.03±2.51	74.47 ± 9.05
		PathDino-512 (ours)	**88.57 ±3.08**	**86.35 ±5.33**	71.36 ±1.64	70.47 ±2.47	32.08 ±2.57	**79.61 ±1.00**	52.59 ±3.21	**95.82±2.26**	99.65 ±0.11	84.79±3.14	72.69 ±7.60

Table S6. Internal histopathology image datasets. Four different datasets were collected at a hospital for four sites including Liver, Skin, Breast, and colon sites.

Dataset	#Class	#WSI	#Patches	Diagnosis
CRC	3	209	4,619	Cancer Adjacent polyp Non-recurrent polyp Recurrent polyp
Liver	3	324	2,976	Alcoholic Steatohepatitis Non-alcoholic Steatohepatitis Normal tissue
Skin	4	660	8,390	Well differentiated Moderately differentiated Poorly differentiated Normal tissue
Breast	16	73	1,141	Adenoid Cystic Carcinoma Adenomyoepthelioma Ductal Carcinoma In Situ Ductal Carcinoma In Situ, Columnar Cell Lesions Including Flat Epithelial Atypia, Atypical Ductal Hyperplasia Intraductal Papilloma, Columnar Cell Lesions Invasive Breast Carcinoma of No Special Type Invasive lobular carcinoma Lobular Carcinoma In Situ + Atypical Lobular Hyperplasia Lobular Carcinoma In Situ, Flat Epithelial Atypia, Atypical Lobular Hyperplasia Malignant Adenomyoepithelioma Metaplastic Carcinoma Microglandular Adenosis Microinvasive carcinoma Mucinous Cystadenocarcinoma Radial scar complex sclerosing lesion Normal tissue

Table S7. Public histopathology image datasets including PANDA, CAMELYON16, BRACS, DigestPath, Kather, PanNuke, and WSSS4LUAD. Number of images represents the total images (patches) used in the evaluation regardless of their training/testing split, since we used the leave-one-out evaluation method for the search task and k-fold cross-validation for the classification task.

Dataset	Analysis Scale	#Class	#WSI	#Image	Diagnosis
PANDA	WSI/Patch	6	10,349	87,451	background (non tissue) or unknown benign tissue (stroma and epithelium combined) cancerous tissue, not specified cancerous epithelium (Gleason 3) cancerous epithelium (Gleason 4) cancerous epithelium (Gleason 5)
CAMELYON16	WSI/Patch	2	128	2,864	Tumor Normal
BRACS	WSI/Patch	3	523	10,984	Benign Atypical Malignant
DigestPath	Patch-Level	2	-	1,103	Benign Malignant
Kather	Patch-Level	9	-	7,180	ADIPOSE BACKGROUND DEBRIS LYMPHO MUCUS Smooth Muscle Normal Colon Mucosa Cancer-Associated Stroma Colorectal Adenocarcinoma Epithelium
PanNuke	Patch-Level	19	-	2,656 2,523 2,722	Breast Colon Bile-duct Esophagus Uterus Lung Cervix Head Neck Skin Adrenal_gland kidney Stomach Prostate testis Liver Thyroid Pancreatic Overin Bladder
WSSS4LUAD	Patch-Level	7	-	10,091	Normal Stroma Stroma-Normal Tumor Tumor-Normal Tumor-Stroma Tumor-Stroma-Normal

Table S8. WSI-level Top-1 Accuracy using the proposed FPS patching method and minimum of medium proposed in Yottixel [24]

			Internal Datasets				Public Datasets		
			Private-Breast	Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.48	0.67	0.73	0.58	0.32	0.54	0.53
		DenseNet121 [3]	0.48	0.64	0.69	0.49	0.3	0.67	0.52
		EfficientNet-b3-288 [4]	0.41	0.66	0.73	0.6	0.32	0.59	0.55
		EfficientNet-b5 [4]	0.51	0.71	0.71	0.63	0.37	0.57	0.54
		ConvNext-b-224 [5]	0.56	0.75	0.74	0.6	0.34	0.62	0.58
		ConvNext-xlarge [5]	0.56	0.76	0.74	0.58	0.35	0.61	0.58
	Transformer	ViT-b16-224 [8]	0.41	0.7	0.72	0.5	0.31	0.6	0.54
		DinoV1-ViT-s16 [9]	0.48	0.71	0.74	0.55	0.36	0.67	0.6
		DinoV1-ViT-b16 [9]	0.55	0.72	0.73	0.61	0.37	0.63	0.59
		DinoV2-ViT-b14 [10]	0.53	0.71	0.72	0.56	0.31	0.61	0.51
		CLIP - ViT-B/16 [11]	0.49	0.67	0.75	0.56	0.36	0.67	0.58
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.58	0.77	0.77	0.64	0.61	0.67	0.61
		MoCoV2-ResNet50 [1]	0.62	0.79	0.74	0.64	0.62	0.67	0.6
		MuDiPath-ResNet50 [6]	0.44	0.7	0.72	0.58	0.35	0.63	0.51
		MuDiPath-DenseNet-101 [6]	0.51	0.68	0.74	0.66	0.36	0.65	0.56
		KimiaNet [7]	0.51	0.78	0.75	0.64	0.57	0.76	0.62
	Transformer	BiomedCLIP - [12]	0.47	0.74	0.75	0.58	0.34	0.61	0.55
		HIPT-ViT-s16 [13]	0.44	0.68	0.73	0.57	0.32	0.62	0.52
		PLIP [14]	0.58	0.73	0.75	0.64	0.56	0.73	0.6
		iBOT-Path [15]	0.64	0.79	0.76	0.65	0.53	0.67	0.64
		DinoSSLPathology-8 [1]	0.58	0.74	0.78	0.63	0.47	0.74	0.61
		PathDino-224 (ours)	0.53	0.75	0.74	0.61	0.46	0.72	0.61
		PathDino-512 (ours)	0.68	0.83	0.78	0.63	0.58	0.73	0.64

Table S9. WSI-level Top-1 Macro avg F-1 score

			Internal Datasets				Public Datasets		
			Private-Breast	Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.43	0.49	0.63	0.58	0.29	0.48	0.48
		DenseNet121 [3]	0.37	0.44	0.61	0.48	0.27	0.65	0.47
		EfficientNet-b3-288 [4]	0.35	0.49	0.64	0.6	0.29	0.55	0.5
		EfficientNet-b5 [4]	0.41	0.52	0.6	0.63	0.35	0.54	0.49
		ConvNext-b-224 [5]	0.47	0.61	0.64	0.6	0.31	0.6	0.53
		ConvNext-xlarge [5]	0.51	0.52	0.64	0.58	0.33	0.57	0.52
	Transformer	ViT-b16-224 [8]	0.3	0.51	0.62	0.5	0.28	0.54	0.49
		DinoV1-ViT-s16 [9]	0.41	0.52	0.64	0.55	0.34	0.63	0.55
		DinoV1-ViT-b16 [9]	0.49	0.59	0.62	0.61	0.35	0.6	0.54
		DinoV2-ViT-b14 [10]	0.46	0.49	0.65	0.56	0.28	0.57	0.45
		CLIP - ViT-B/16 [11]	0.39	0.46	0.66	0.56	0.34	0.61	0.52
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.49	0.56	0.65	0.65	0.62	0.63	0.56
		MoCoV2-ResNet50 [1]	0.49	0.63	0.63	0.64	0.64	0.64	0.53
		MuDiPath-ResNet50 [6]	0.41	0.57	0.61	0.58	0.33	0.59	0.45
		MuDiPath-DenseNet-101 [6]	0.47	0.5	0.63	0.66	0.33	0.63	0.5
		KimiaNet [7]	0.47	0.66	0.65	0.64	0.58	0.74	0.57
	Transformer	BiomedCLIP - [12]	0.39	0.5	0.63	0.57	0.31	0.57	0.48
		HIPT-ViT-s16 [13]	0.33	0.46	0.62	0.57	0.29	0.58	0.48
		PLIP [14]	0.58	0.6	0.65	0.65	0.56	0.69	0.53
		iBOT-Path [15]	0.61	0.74	0.66	0.66	0.52	0.62	0.58
		DinoSSLPathology-8 [1]	0.51	0.54	0.67	0.64	0.46	0.7	0.57
		PathDino-224 (ours)	0.56	0.6	0.63	0.61	0.45	0.69	0.56
		PathDino-512 (ours)	0.66	0.74	0.67	0.64	0.59	0.69	0.57

Table S10. WSI-level MV@3 Accuracy

			Internal Datasets			Public Datasets		
			Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.72	0.76	0.64	0.34	0.6	0.58
		DenseNet121 [3]	0.72	0.74	0.51	0.32	0.67	0.54
		EfficientNet-b3-288 [4]	0.69	0.76	0.61	0.33	0.6	0.58
		EfficientNet-b5 [4]	0.71	0.75	0.64	0.38	0.62	0.57
		ConvNext-b-224 [5]	0.75	0.78	0.62	0.37	0.68	0.6
		ConvNext-xlarge [5]	0.76	0.79	0.61	0.37	0.65	0.62
	Transformer	ViT-b16-224 [8]	0.71	0.76	0.55	0.33	0.67	0.57
		DinoV1-ViT-s16 [9]	0.74	0.77	0.61	0.38	0.67	0.59
		DinoV1-ViT-b16 [9]	0.77	0.77	0.62	0.39	0.69	0.63
		DinoV2-ViT-b14 [10]	0.74	0.77	0.59	0.32	0.58	0.53
		CLIP - ViT-B/16 [11]	0.69	0.79	0.58	0.38	0.67	0.6
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.79	0.77	0.66	0.58	0.7	0.64
		MoCoV2-ResNet50 [1]	0.83	0.78	0.65	0.6	0.69	0.61
		MuDiPath-ResNet50 [6]	0.73	0.78	0.58	0.37	0.67	0.56
		MuDiPath-DenseNet-101 [6]	0.72	0.77	0.63	0.37	0.68	0.58
		KimiaNet [7]6	0.8	0.81	0.65	0.56	0.78	0.64
	Transformer	BiomedCLIP - [12]	0.77	0.78	0.63	0.36	0.67	0.62
		HIPT-ViT-s16 [13]	0.67	0.76	0.52	0.33	0.64	0.54
		PLIP [14]	0.74	0.79	0.67	0.55	0.77	0.63
		iBOT-Path [15]	0.83	0.8	0.63	0.53	0.71	0.65
		DinoSSLPathology-8 [1]	0.77	0.8	0.64	0.47	0.68	0.64
		PathDino-224 (ours)	0.79	0.8	0.59	0.48	0.74	0.61
		PathDino-512 (ours)	0.86	0.8	0.65	0.58	0.78	0.66

Table S11. WSI-level MV@3 Macro Avg F-1 score

			Internal Datasets			Public Datasets		
			Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.49	0.65	0.64	0.3	0.52	0.53
		DenseNet121 [3]	0.49	0.62	0.5	0.28	0.62	0.48
		EfficientNet-b3-288 [4]	0.5	0.65	0.6	0.3	0.52	0.52
		EfficientNet-b5 [4]	0.55	0.61	0.64	0.35	0.56	0.53
		ConvNext-b-224 [5]	0.6	0.67	0.61	0.33	0.63	0.53
		ConvNext-xlarge [5]	0.56	0.66	0.61	0.34	0.6	0.54
	Transformer	ViT-b16-224 [8]	0.48	0.64	0.54	0.29	0.61	0.51
		DinoV1-ViT-s16 [9]	0.54	0.65	0.61	0.34	0.6	0.53
		DinoV1-ViT-b16 [9]	0.66	0.65	0.62	0.35	0.64	0.58
		DinoV2-ViT-b14 [10]	0.54	0.66	0.59	0.29	0.5	0.46
		CLIP - ViT-B/16 [11]	0.47	0.66	0.58	0.35	0.58	0.54
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.63	0.62	0.67	0.59	0.64	0.57
		MoCoV2-ResNet50 [1]	0.71	0.64	0.65	0.6	0.65	0.54
		MuDiPath-ResNet50 [6]	0.5	0.64	0.58	0.33	0.62	0.49
		MuDiPath-DenseNet-101 [6]	0.53	0.63	0.63	0.34	0.64	0.52
		KimiaNet [7]	0.62	0.71	0.66	0.56	0.74	0.57
	Transformer	BiomedCLIP - [12]	0.53	0.65	0.62	0.33	0.61	0.55
		HIPT-ViT-s16 [13]	0.45	0.63	0.52	0.3	0.56	0.47
		PLIP [14]	0.63	0.66	0.68	0.54	0.72	0.57
		iBOT-Path [15]	0.76	0.69	0.64	0.52	0.64	0.59
		DinoSSLPathology-8 [1]	0.66	0.67	0.65	0.45	0.59	0.57
		PathDino-224 (ours)	0.6	0.68	0.59	0.46	0.69	0.53
		PathDino-512 (ours)	0.74	0.67	0.66	0.58	0.73	0.58

Table S12. WSI-level MV@5 Accuracy

			Internal Datasets			Public Datasets		
			Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.71	0.78	0.63	0.36	0.64	0.6
		DenseNet121 [3]	0.7	0.78	0.49	0.34	0.7	0.58
		EfficientNet-b3-288 [4]	0.68	0.78	0.59	0.36	0.61	0.57
		EfficientNet-b5 [4]	0.7	0.77	0.64	0.4	0.66	0.61
		ConvNext-b-224 [5]	0.73	0.79	0.58	0.38	0.71	0.62
		ConvNext-xlarge [5]	0.78	0.8	0.63	0.39	0.67	0.64
	Transformer	ViT-b16-224 [8]	0.72	0.78	0.54	0.35	0.69	0.58
		DinoV1-ViT-s16 [9]	0.76	0.78	0.6	0.39	0.67	0.63
		DinoV1-ViT-b16 [9]	0.78	0.79	0.62	0.41	0.71	0.64
		DinoV2-ViT-b14 [10]	0.72	0.78	0.59	0.35	0.63	0.56
		CLIP - ViT-B/16 [11]	0.71	0.8	0.6	0.4	0.69	0.61
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.79	0.79	0.64	0.57	0.74	0.66
		MoCoV2-ResNet50 [1]	0.82	0.78	0.65	0.58	0.71	0.64
		MuDiPath-ResNet50 [6]	0.73	0.79	0.57	0.39	0.7	0.56
		MuDiPath-DenseNet-101 [6]	0.72	0.78	0.66	0.39	0.71	0.6
		KimiaNet [7]	0.78	0.82	0.62	0.54	0.81	0.66
	Transformer	BiomedCLIP - [12]	0.77	0.78	0.65	0.38	0.69	0.61
		HIPT-ViT-s16 [13]	0.68	0.77	0.55	0.35	0.67	0.56
		PLIP [14]	0.76	0.82	0.69	0.54	0.75	0.64
		iBOT-Path [15]	0.82	0.82	0.63	0.53	0.74	0.65
		DinoSSLPathology-8 [1]	0.77	0.8	0.63	0.48	0.7	0.64
		PathDino-224 (ours)	0.76	0.8	0.57	0.48	0.71	0.64
		PathDino-512 (ours)	0.85	0.82	0.65	0.56	0.77	0.67

Table S13. WSI-level MV@5 Macro Avg F-1 score

			Internal Datasets			Public Datasets		
			Private-Liver	Private-Skin	Private-CRC	PANDA [16]	CAMELYON16 [17]	BRACS [18]
Pret. on Natural	CNN-based	ResNet50 [2]	0.48	0.66	0.63	0.32	0.55	0.54
		DenseNet121 [3]	0.47	0.65	0.49	0.3	0.64	0.52
		EfficientNet-b3-288 [4]	0.5	0.65	0.59	0.31	0.49	0.51
		EfficientNet-b5 [4]	0.47	0.63	0.64	0.36	0.58	0.56
		ConvNext-b-224 [5]	0.5	0.65	0.57	0.33	0.65	0.55
		ConvNext-xlarge [5]	0.53	0.66	0.63	0.35	0.57	0.56
	Transformer	ViT-b16-224 [8]	0.49	0.65	0.53	0.31	0.59	0.5
		DinoV1-ViT-s16 [9]	0.52	0.65	0.61	0.36	0.57	0.56
		DinoV1-ViT-b16 [9]	0.57	0.66	0.61	0.37	0.65	0.57
		DinoV2-ViT-b14 [10]	0.53	0.66	0.59	0.3	0.54	0.49
		CLIP - ViT-B/16 [11]	0.52	0.67	0.6	0.35	0.61	0.55
Pret. on Histopathology	CNN-based	Barlow-Twins-ResNet50 [1]	0.67	0.63	0.65	0.56	0.69	0.57
		MoCoV2-ResNet50 [1]	0.66	0.61	0.66	0.57	0.65	0.56
		MuDiPath-ResNet50 [6]	0.49	0.64	0.57	0.35	0.62	0.49
		MuDiPath-DenseNet-101 [6]	0.49	0.63	0.66	0.35	0.66	0.52
		KimiaNet [7]	0.61	0.69	0.63	0.54	0.77	0.59
	Transformer	BiomedCLIP - [12]	0.56	0.62	0.65	0.34	0.59	0.54
		HIPT-ViT-s16 [13]	0.46	0.65	0.56	0.3	0.58	0.48
		PLIP [14]	0.67	0.69	0.7	0.52	0.7	0.57
		iBOT-Path [15]	0.72	0.70	0.64	0.51	0.67	0.57
		DinoSSLPathology-8 [1]	0.62	0.67	0.64	0.45	0.61	0.57
		PathDino-224 (ours)	0.65	0.67	0.58	0.45	0.64	0.56
		PathDino-512 (ours)	0.74	0.69	0.66	0.56	0.72	0.59

Algorithm S1 HistoRotate, Image Augmentation in a Self-Supervised Manner (DINO Framework)

Require: Input image I, Global crop scales $[a, b]$, Local crop scales $[c, d]$, Number of local crops n, Set of discrete angles Θ

function EXACTROTATION(I, Θ)
    $\theta \leftarrow$ random.choice(Θ)
    $I' \leftarrow$ rotate(I, θ)
    return I'
function HISTOROTATE($I, [a, b], [c, d], n, \Theta$)
    Initialize empty list $crops$
    if size$(I)[0] = 1024$ **then**
        $crops$.append(global-transfo1-1024(I)) ▷ Include Random 360° Rotation
        $crops$.append(global-transfo2-1024(I)) ▷ Include Random 360° Rotation
        for $i = 1, n$ **do**
            $crops$.append(local-transfo(I)) ▷ Always Include Random 360° Rotation
    else
        $crops$.append(global-transfo1_512(I)) ▷ Include Random Rotation from Θ = {90, 180, 270, 360}
        $crops$.append(global-transfo2_512(I)) ▷ Include Random Rotation from Θ = {90, 180, 270, 360}
        for $i = 1, n$ **do**
            $crops$.append(local-transfo(I)) ▷ Always Include Random 360° Rotation
    return $crops$

References

[1] Mingu Kang, Heon Song, Seonwook Park, Donggeun Yoo, and Sérgio Pereira. Benchmarking self-supervised learning on diverse pathology datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3344–3354, June 2023. 2, 5, 6, 7, 9, 10, 11

[2] 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. 2, 6, 7, 9, 10, 11

[3] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 2, 6, 7, 9, 10, 11

[4] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2, 6, 7, 9, 10, 11

[5] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 2, 6, 7, 9, 10, 11

[6] Romain Mormont, Pierre Geurts, and Raphaël Marée. Multi-task pre-training of deep neural networks for digital pathology. *IEEE journal of biomedical and health informatics*, 25(2):412–421, 2020. 2, 6, 7, 9, 10, 11

[7] Abtin Riasatian, Morteza Babaie, Danial Maleki, Shivam Kalra, Mojtaba Valipour, Sobhan Hemati, Manit Zaveri, et al. Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Medical Image Analysis*, 70:102032, 2021. 2, 5, 6, 7, 9, 10, 11

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2, 6, 7, 9, 10, 11

[9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 2, 6, 7, 9, 10, 11

[10] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2, 5, 6, 7, 9, 10, 11

[11] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 2, 5, 6, 7, 9, 10, 11

[12] Sheng Zhang, Yanbo Xu, Naoto Usuyama, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, et al. Large-scale domain-specific pretraining for biomedical vision-language processing. *arXiv preprint arXiv:2303.00915*, 2023. 2, 5, 6, 7, 9, 10, 11

[13] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16144–16155, 2022. 2, 6, 7, 9, 10, 11

[14] Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature Medicine*, pages 1–10, 2023. 2, 5, 6, 7, 9, 10, 11

[15] Alexandre Filiot, Ridouane Ghermi, Antoine Olivier, Paul Jacob, Lucas Fidon, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. Scaling self-supervised learning for histopathology with masked image modeling. *medRxiv*, pages 2023–07, 2023. 2, 6, 7, 9, 10, 11

[16] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1):154–163, 2022. 5, 6, 7, 9, 10, 11

[17] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017. 5, 6, 7, 9, 10, 11

[18] Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, Maria Gabrani, Florinda Feroce, and Maria Frucci. BRACS: A Dataset for BReAst Carcinoma Subtyping in H&E Histology Images. *Database*, 2022:baac093, 10 2022. 5, 6, 7, 9, 10, 11

[19] Qian Da, Xiaodi Huang, Zhongyu Li, Yanfei Zuo, Chenbin Zhang, Jingxin Liu, Wen Chen, Jiahui Li, Dou Xu, Zhiqiang Hu, et al. Digestpath: A benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system. *Medical Image Analysis*, 80:102485, 2022. 6, 7

[20] Jakob Nikolas Kather, Frank Gerrit Zöllner, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Cleo-Aron Weis. Collection of textures in colorectal cancer histology. *Zenodo*, may 2016. 6, 7

[21] Jevgenij Gamper, Navid Alemi Koohbanani, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020. 6, 7

[22] Chu Han, Xipeng Pan, Lixu Yan, Huan Lin, Bingbing Li, Su Yao, Shanshan Lv, Zhenwei Shi, Jinhai Mai, Jiatai Lin, et al. Wsss4luad: Grand challenge on weakly-supervised tissue semantic segmentation for lung adenocarcinoma. *arXiv preprint arXiv:2204.06455*, 2022. 6, 7

[23] Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Junzhou Huang, Wei Yang, and Xiao Han. Transpath: Transformer-based self-supervised learning for histopathological image classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 186–195. Springer, 2021. 6, 7

[24] Shivam Kalra, Hamid R Tizhoosh, Charles Choi, Sultaan Shah, Phedias Diamandis, Clinton JV Campbell, and Liron Pantanowitz. Yottixel–an image search engine for large archives of histopathology whole slide images. *Medical Image Analysis*, 65:101757, 2020. 9