上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Adapt Foundational Segmentation Model with Heterogeneous Searching Space

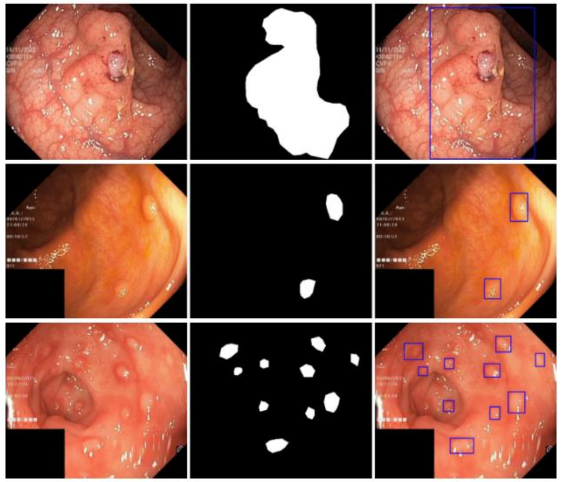
Li Yi, Jie Hu, Songan Zhang, GUANNAN JIANG

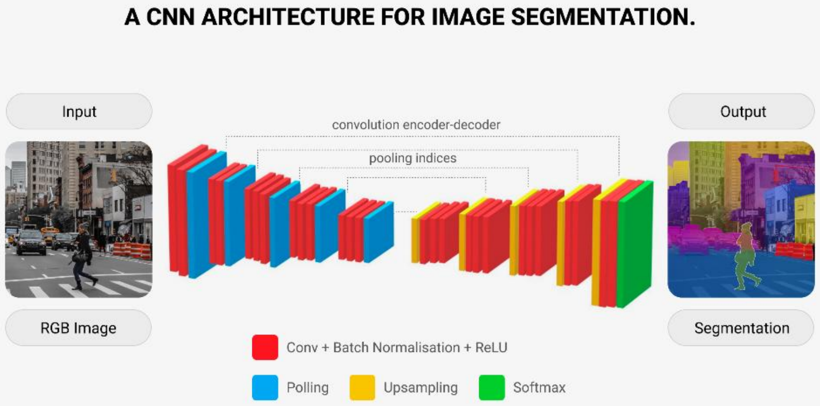
2025/10/3

饮 水 思 源 • 爱 国 荣 校

Background and Significance

- **Foundational Segmentation Model** will be employed in various domain
- Owing to scarce labeled data and infrequent usage, **finetuning** Foundational Segmentation Model for atypical domain is **time-consuming and ineffective**
- It's significant to develop **a domain adaptation strategy** to facilitate Foundational Segmentation Model's performance **in all domains** without finetuning



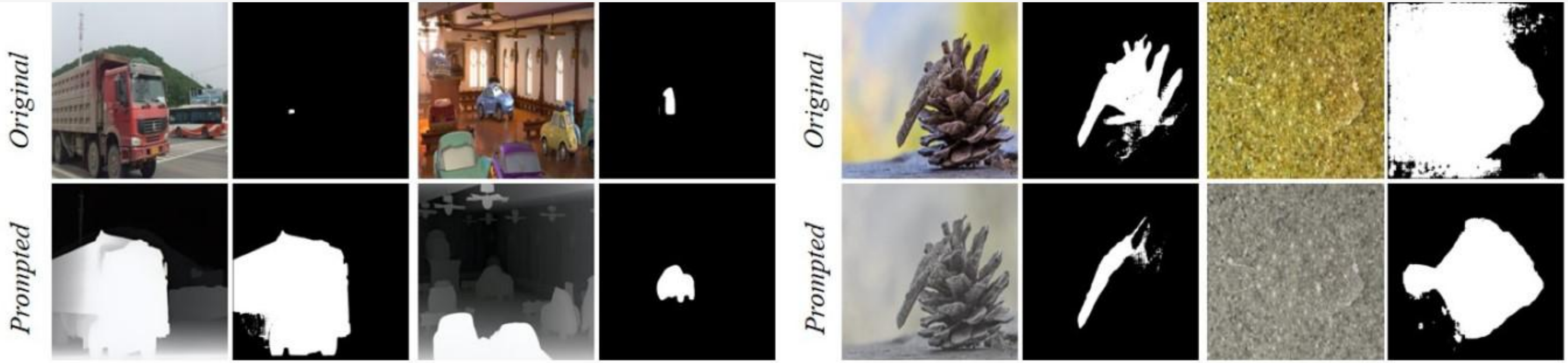


Dataset

- **Common Field: NJU2k, VT1k**
- **Thermal Infrared: VT1k-T**
- **Depth Image: NJU2k-Depth**
- **Camouflage Objects: CAMO, COD10k, NC4k**
- **Endoscopic Image: Kvasir-SEG**
- **Ultrasound Image: BUSI**
- **Industrial Data: KoletorSDDV2, MTSD**

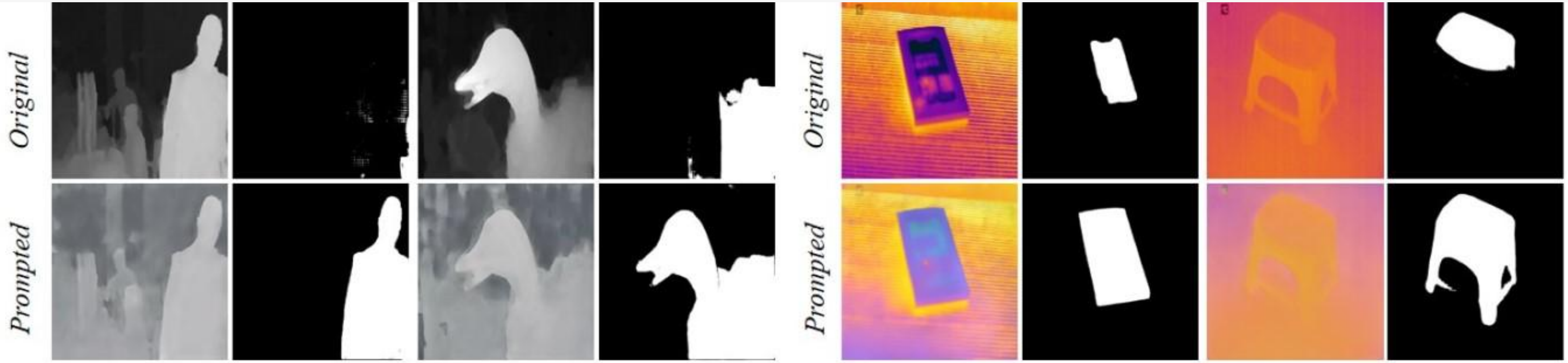
Feasibility Certification

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



(a) Common Scenes: NJU2k, SAM-H, + 38.14 mIoU
Prompt: DepthAnything: -, BrightnessUp: 4, ImageDemoireing: 6

(b) Camouflage Images: CAMO, SAM-H, + 4.60 mIoU
Prompt: SaturationDown: 8, BilateralFilter: 9, ContrastDown: 3



(c) Depth Images: NJU2k, SAM-H, + 6.61 mIoU
Prompt: ImageDemoireing: -, AlbedoMap: -, SaturationDown: 7

(d) Thermal Images: VT1k, SAM-H, + 4.74 mIoU
Prompt: AlbedoMap: -, BoxFilter: 7, ImageDemoireing: -

Heterogeneous Searching Space

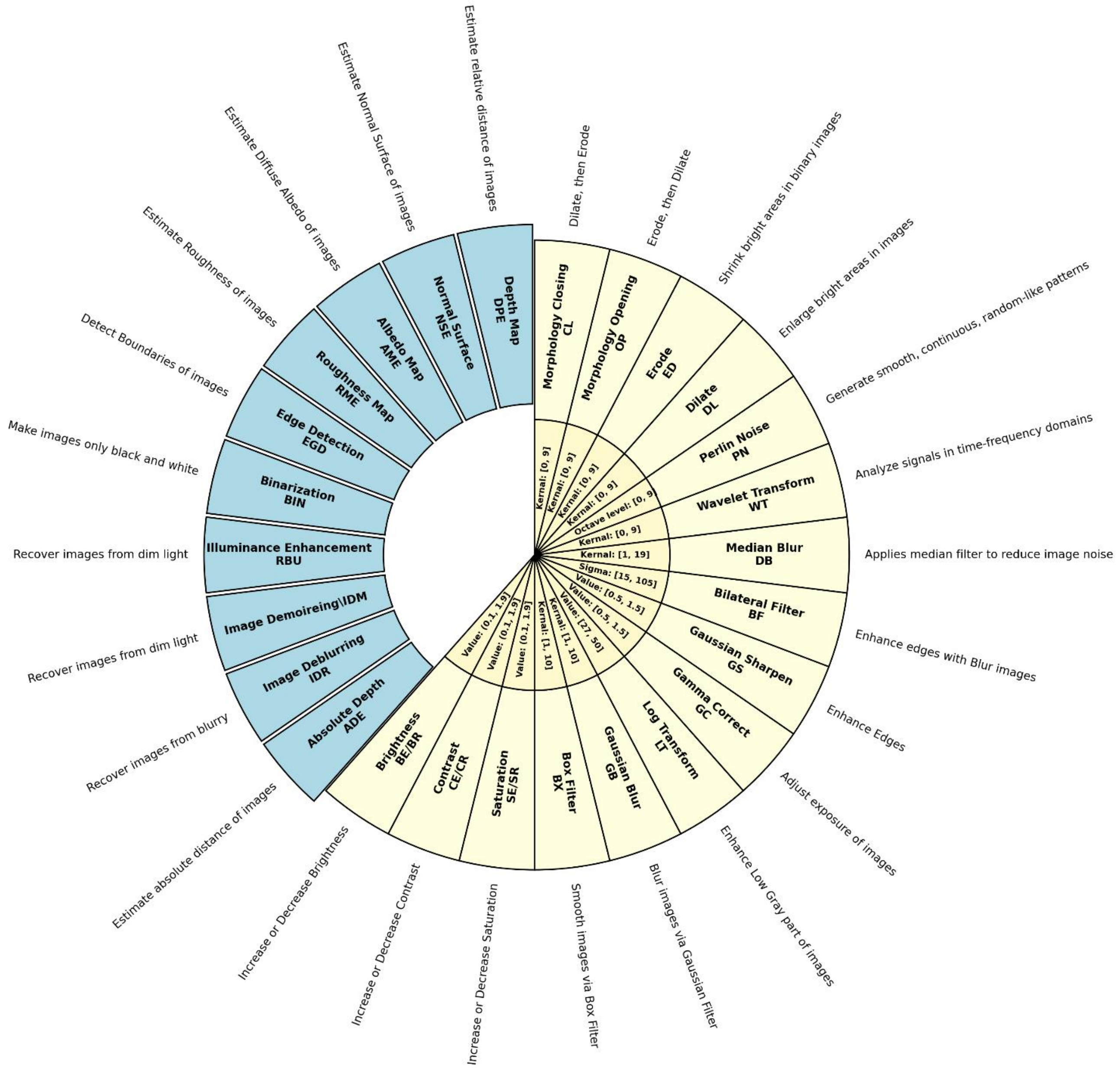
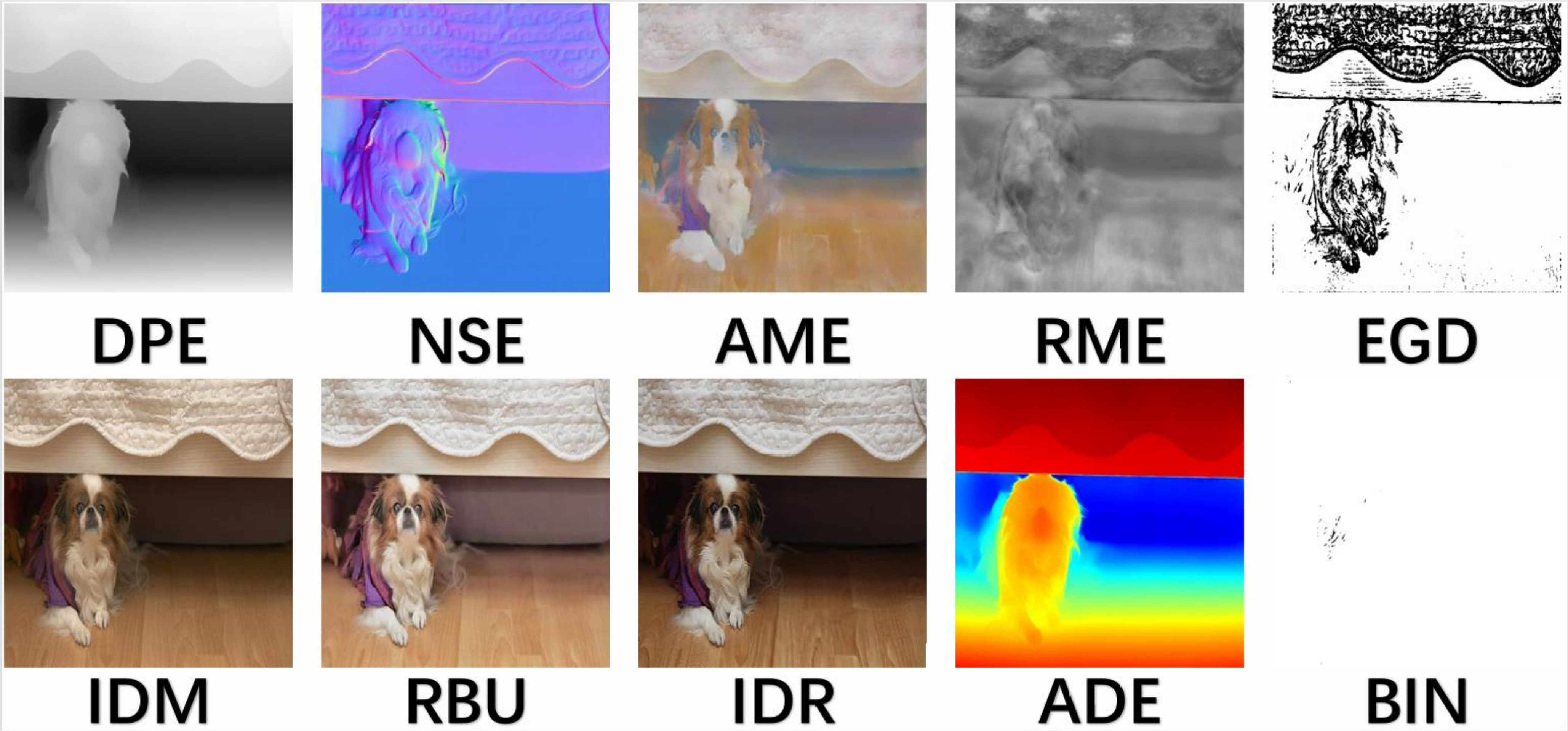


	Image Processing		Description
Learning-based Methods	Depth Map Estimation (DPE) [7]		Estimate relative distance of images
	Normal Surface Estimation (NSE) [8]		Estimate Normal Surface of images
	Albedo Map Estimation (AME) [4]		Estimate Diffuse Albedo of images
	Roughness Map Estimation (RME) [4]		Estimate Roughness of images
	Edge Detection (EGD) [5]		Detect Boundaries of images
	Binarization (BIN) [6]		Make images only black and white
	Illuminance Enhancement (RBU) [2]		Recover images from dim light
	Image Demoiréing (IDM) [9]		Remove moiré pattern
	Image Deblurring (IDR) [3]		Recover images from blurry
	Absolute Depth Estimation (ADE) [1]		Estimate absolute distance of images
Rule-based Methods	Brightness (BE/BR)	Value: (0.1, 1.9]	Increase or Decrease Brightness
	Contrast (CE/CR)	Value: (0.1, 1.9]	Increase or Decrease Contrast
	Saturation (SE/SR)	Value: (0.1, 1.9]	Increase or Decrease Saturation
	Box Filter (BX)	Kernal: [1, 10]	Smooth images via Box Filter
	Gaussian Blur (GB)	Kernal: [1, 10]	Blur images via Gaussian Filter
	Log Transform (LT)	Value: [27, 50]	Enhance Low Gray part of images
	Gamma Correct (GC)	Value: [0.5, 1.5]	Adjust exposure of images
	Gaussian Sharpen (GS)	Value: [0.5, 1.5]	Enhance Edges
	Bilateral Filter (BF)	Sigma: [15, 105]	Enhance edges with Blur images
	Median Filter (DB)	Kernal: [1, 19]	Enhance edges with Shape images
	Wavelet Transformation (WT)	LH: [0, 3] HH: [4, 6] HL: [7, 9]	Analyze signals in time-frequency domains
	Perlin Noise (PN)	Octave level: [0, 9]	Generate smooth, continuous, random-like patterns
	Dilate (DL)	Kernal: [0, 9]	Enlarge bright areas in images
	Erode (ED)	Kernal: [0, 9]	Shrink bright areas in binary images
	Morphology Opening (OP)	Kernal: [0, 9]	Erode, then Dilate
	Morphology Closing (CL)	Kernal: [0, 9]	Dilate, then Erode

Table 1: This table shows the composition of heterogeneous Action Space, including 22 rule-based methods and 10 learning-based methods, and their corresponding abbreviation

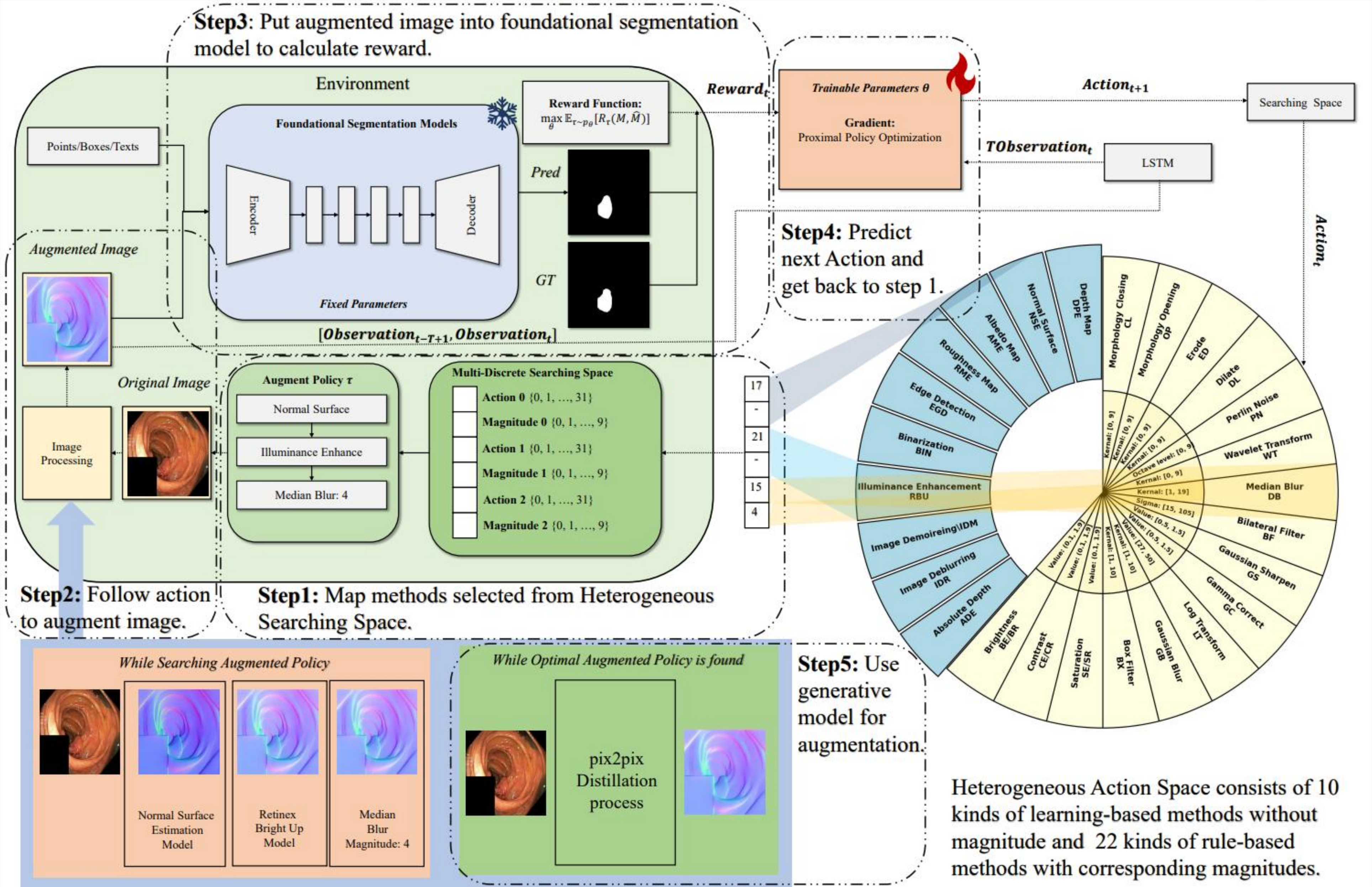
- **Mainly use SOTA in certain field to become learning-based methods**

Heterogeneous Searching Space



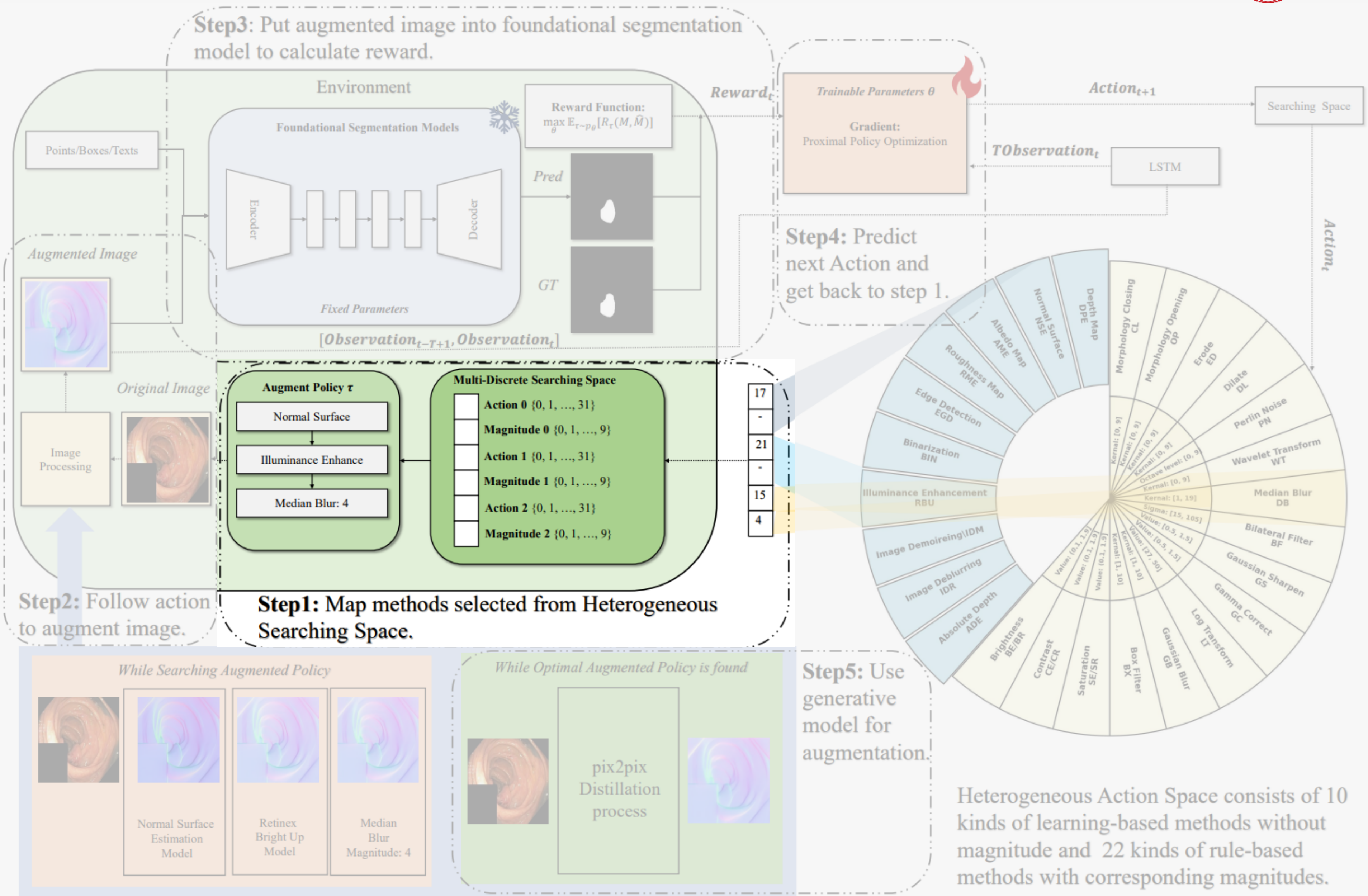
Paradigm

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY



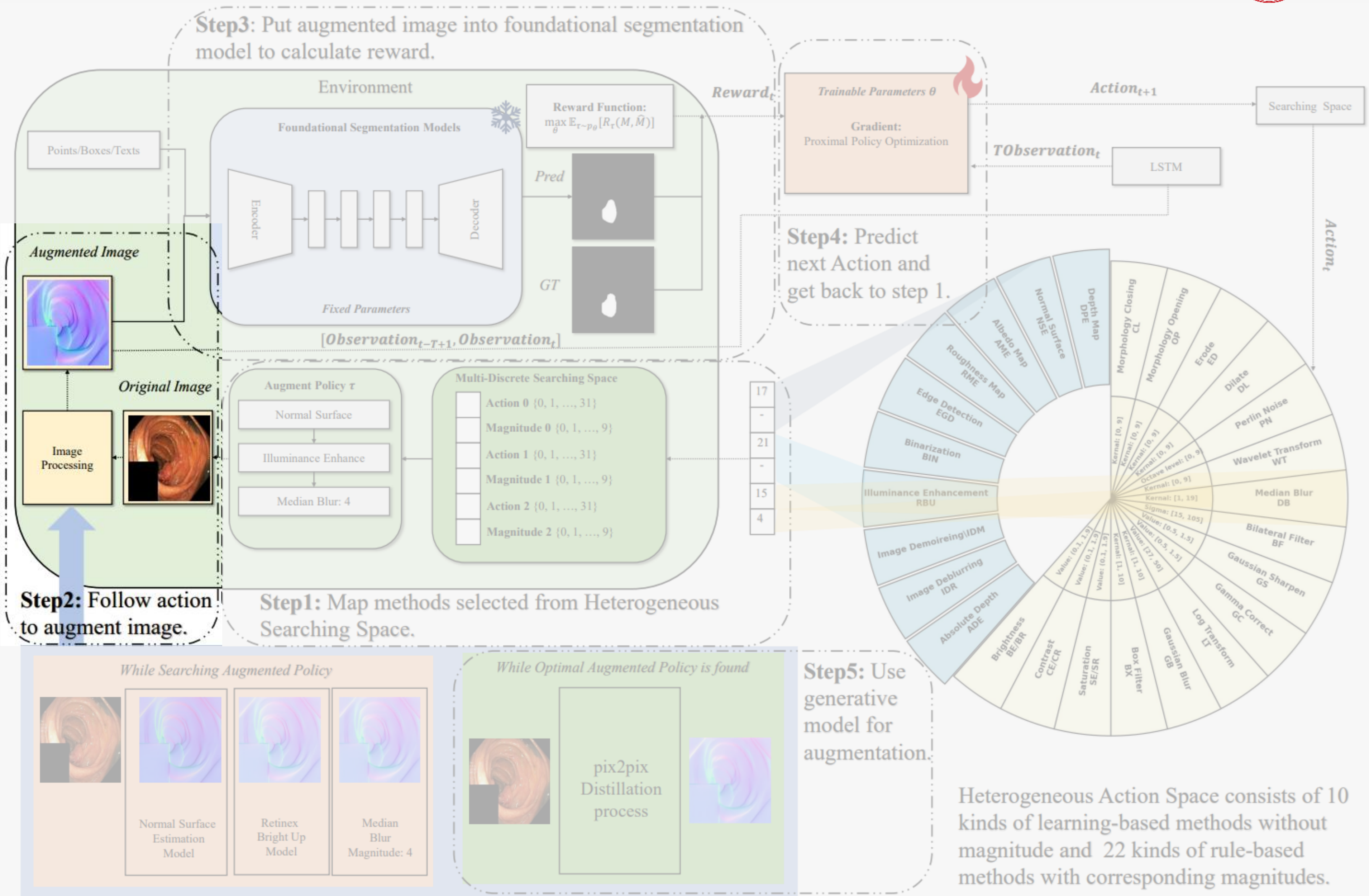
Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

Paradigm



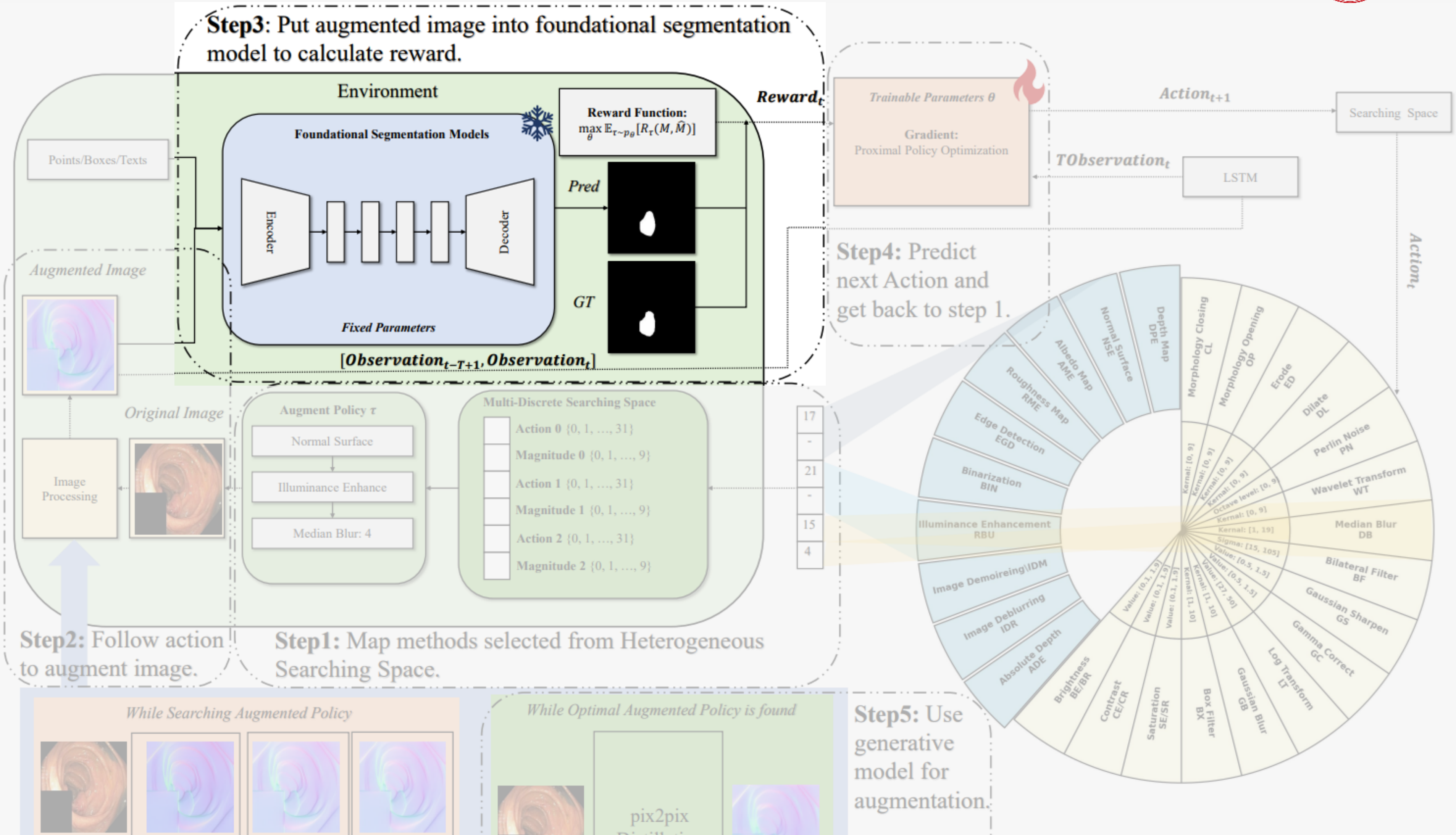
Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

Paradigm



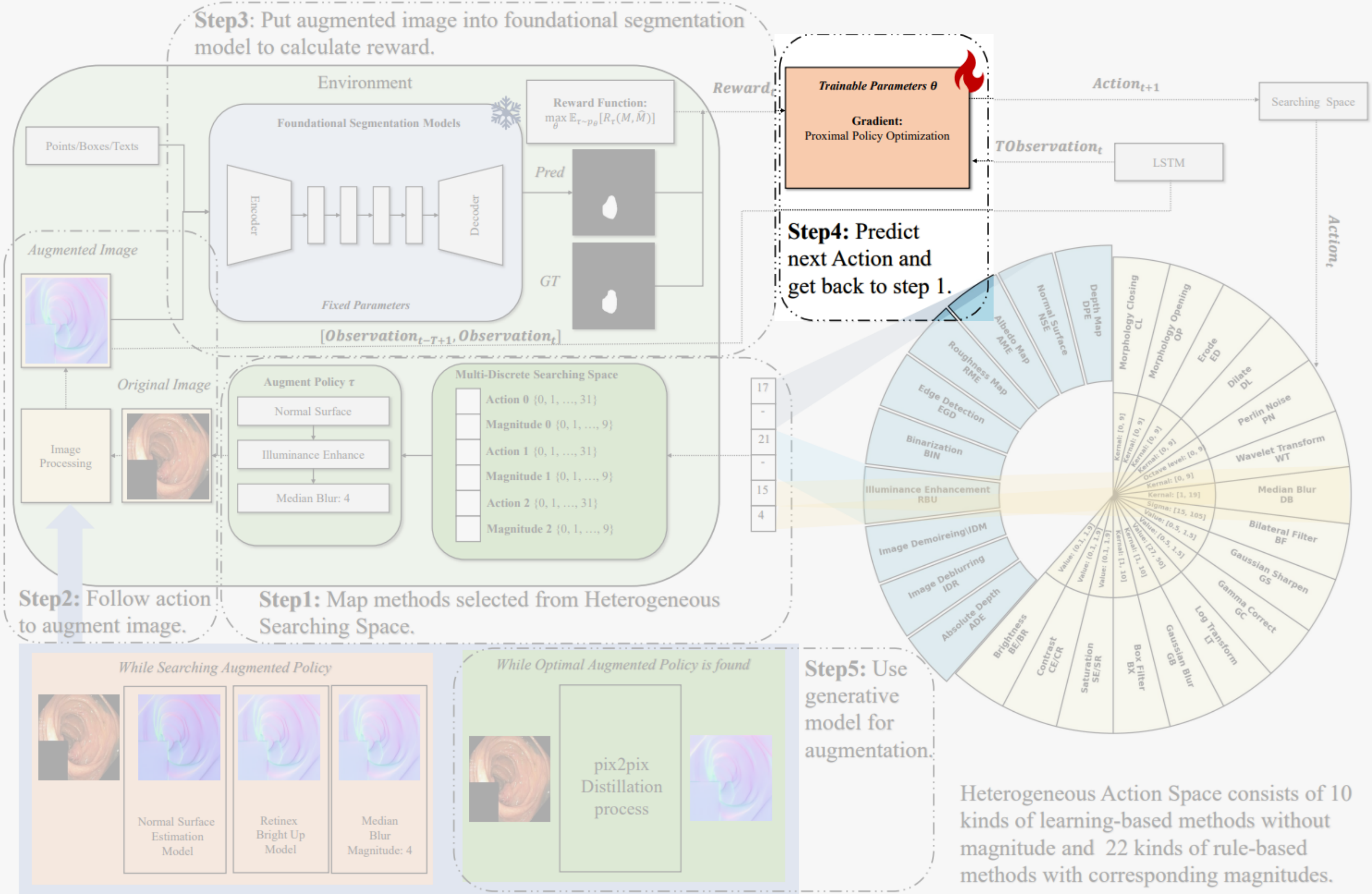
Step3: Put augmented image into foundational segmentation model to calculate reward.
Environment
Foundational Segmentation Models
Reward Function:
Points/Boxes/Texts
Encoder
Decoder
Pred
GT
Fixed Parameters
Reward
Trainable Parameters θ
Gradient:
Proximal Policy Optimization
Searching Space
LSTM
Step4: Predict next Action and get back to step 1.
Augmented Image
Original Image
Image Processing
Augment Policy τ
Normal Surface
Illuminance Enhance
Median Blur: 4
Multi-Discrete Searching Space
Action 0 {0, 1, ..., 31}
Magnitude 0 {0, 1, ..., 9}
Action 1 {0, 1, ..., 31}
Magnitude 1 {0, 1, ..., 9}
Action 2 {0, 1, ..., 31}
Magnitude 2 {0, 1, ..., 9}
Step2: Follow action to augment image.
Step1: Map methods selected from Heterogeneous Searching Space.
While Searching Augmented Policy
Normal Surface Estimation Model
Retinex Bright Up Model
Median Blur Magnitude: 4
While Optimal Augmented Policy is found
pix2pix Distillation process
Step5: Use generative model for augmentation.
Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

Paradigm



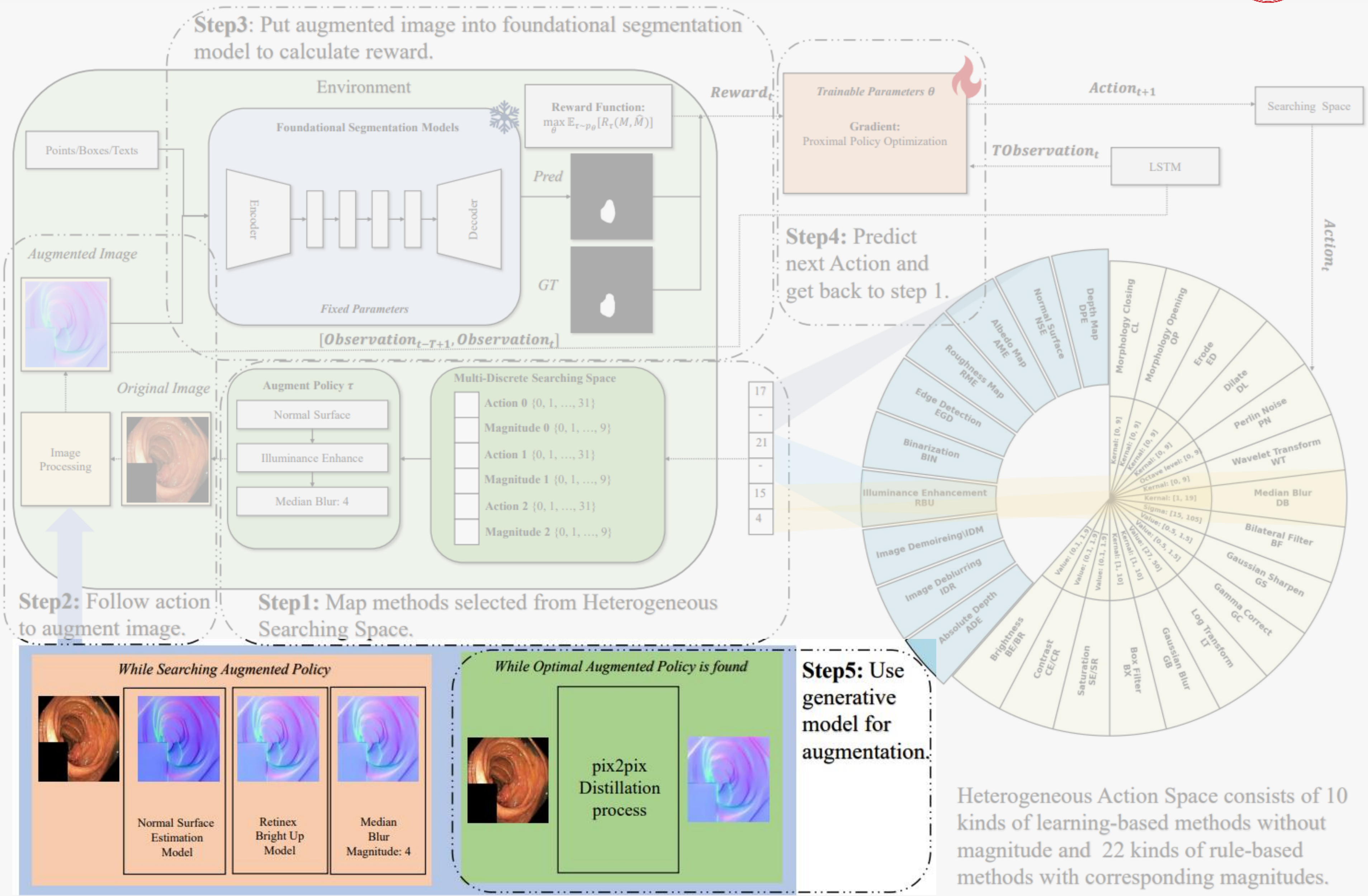
Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

Paradigm



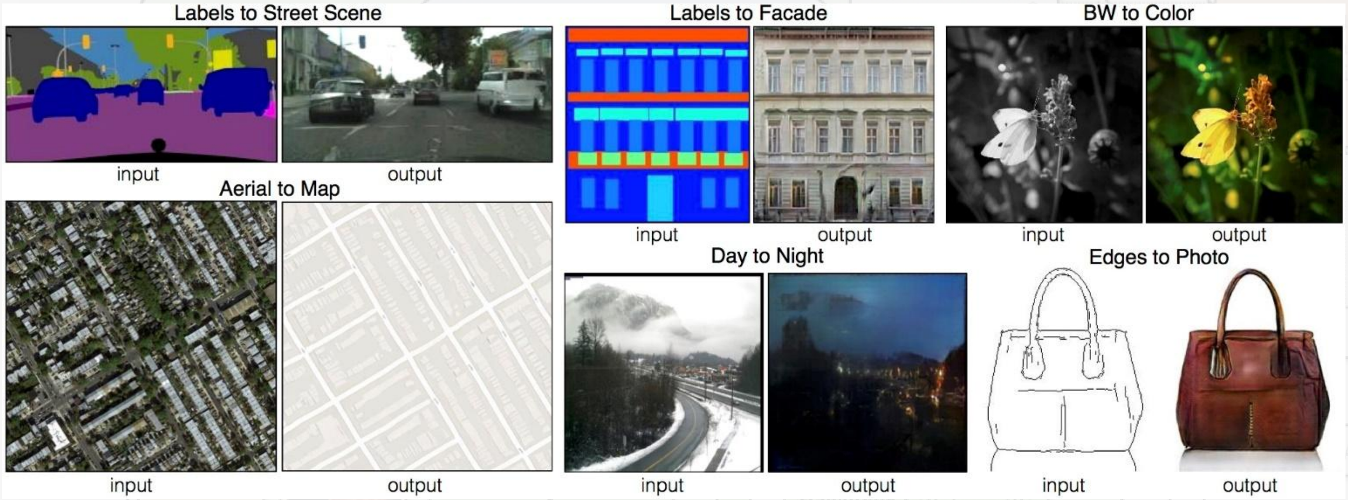
Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

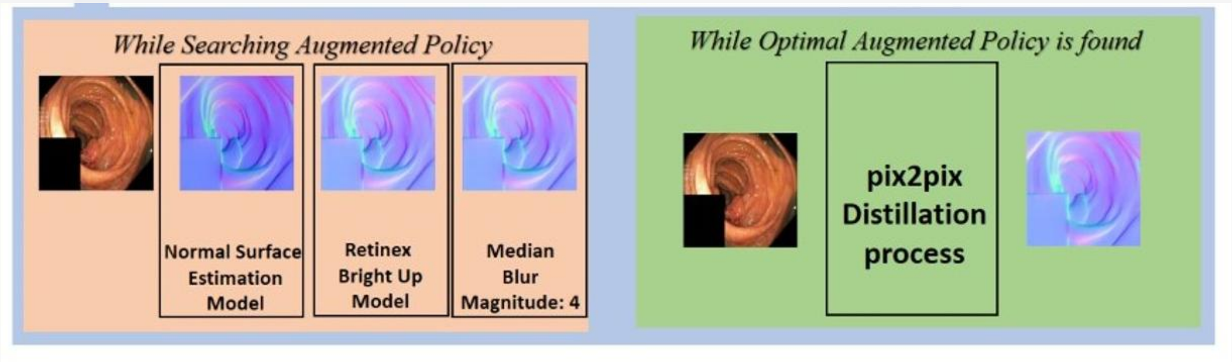
Paradigm



Heterogeneous Action Space consists of 10 kinds of learning-based methods without magnitude and 22 kinds of rule-based methods with corresponding magnitudes.

Distillation

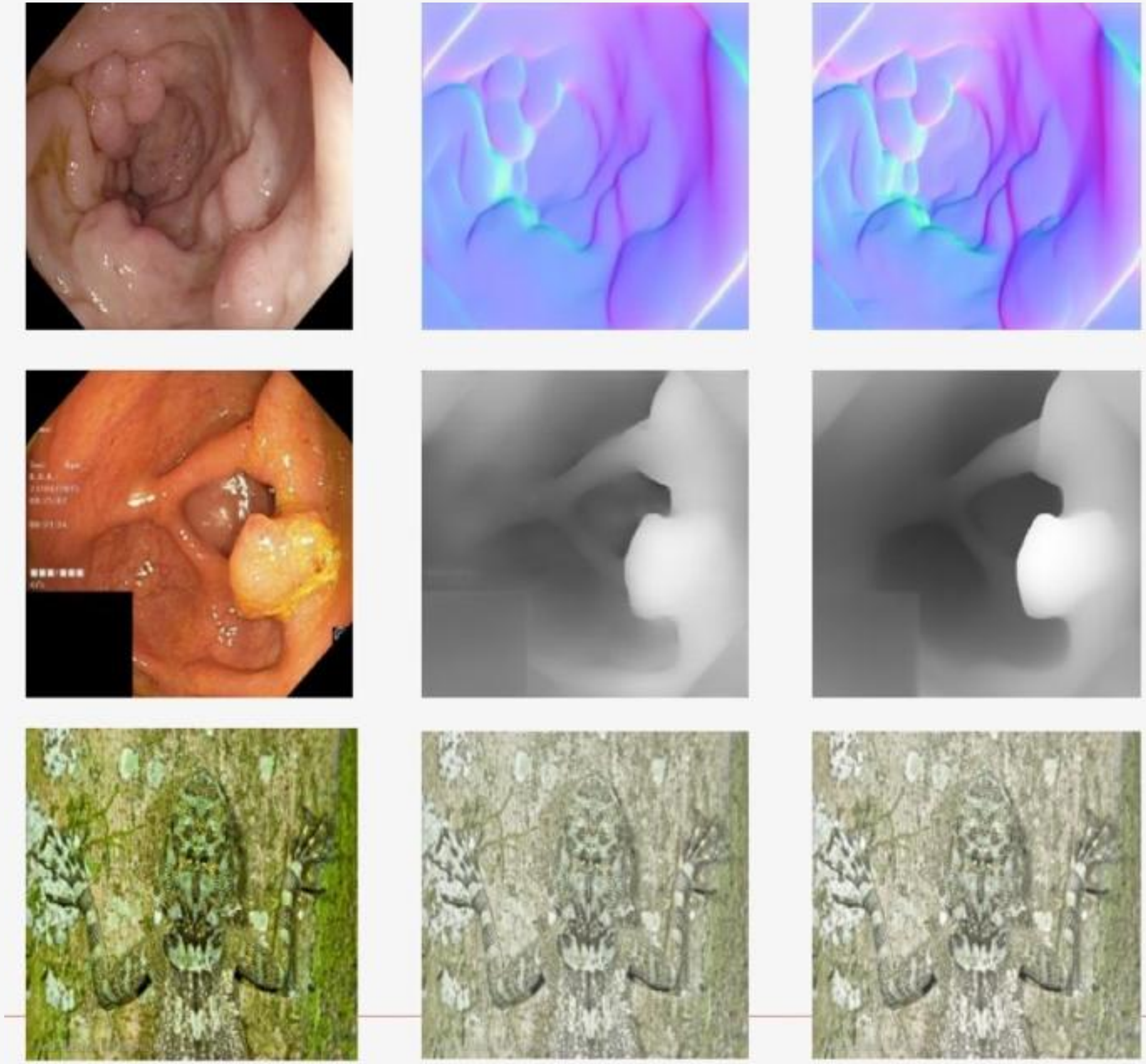




- **During Validation Process:** After finding augmented policy, using original images and augmented images as paired images to train a generator

- **During Inferencing Process:** Using Generative Model instead to accelerate augmenting process

Distillation



Main Result (5-shot)

Backbone	Dataset Type	Dataset Name	SAM [17]	SAM-FT [11]	SAM-IA-32 (Ours)	Augmented Policy
ViT-B	Common Scenes	NJU2K	41.35	44.98	**72.61**	DPE-IDM-SG:3
		VT1K	47.11	49.89	**65.26**	DB:9-DPE-IDM
	Camouflage Objects	CAMO	37.92	38.40	**62.70**	NSE-IDB-DPE
		COD10K	49.78	49.83	**59.95**	DPE-IDB-GC:1
		NC4K	49.09	49.63	**60.48**	DPE-GB:1-IDB
	Depth Map	NJU2K Depth	55.83	57.23	**63.44**	BE:4-RBU-CR:6
	Thermal Map	VT1K-T	57.38	57.64	**59.72**	SL:8-NSE-DPE
	Medical Images	Kvasir	62.13	60.95	**70.56**	NSE-BF:3-GB:7
	Ultrasound	BUSI	52.11	56.70	**59.77**	MB:7-GB:4-BE:7
	Industrial Data	MTSD	54.84	54.91	**56.01**	GB:2-GB:2-CR:3
		KolektorSDD2	46.57	**52.09**	51.77	SG:8-CE:8-GB:6
ViT-H	Common Scenes	NJU2K	42.50	45.17	**78.55**	NSE-DPE-AME
		VT1K	49.69	51.02	**63.41**	DB:1-BF:8-DPE
	Camouflage Objects	CAMO	49.00	49.43	**53.60**	SR:8-BF:9-CR:3
		COD10K	58.39	58.50	**60.67**	NSE-RBU-DPE
		NC4K	57.23	57.45	**66.31**	DPE-SP:9-GC:7
	Depth Map	NJU2K Depth	61.37	63.47	**67.92**	BR:4-IDM-RBU
	Thermal Map	VT1K-T	56.58	57.44	**59.48**	NSE-GB:7-GB:4
	Medical Images	Kvasir	63.29	63.81	**71.92**	GB:0-NSE-GB:7
	Ultrasound	BUSI	54.54	**56.59**	56.42	GC:9-SG:0-SR:0
	Industrial Data	MTSD	59.24	**59.31**	58.21	BE:1-RBU-SE:6
		KolektorSDD2	48.04	**54.24**	54.13	CE:9-SG:7-MB:4

Table 1. 5-shot adaptation results with SAM ViT-B, ViT-H model. Our model brings significant improvement in most of domains.

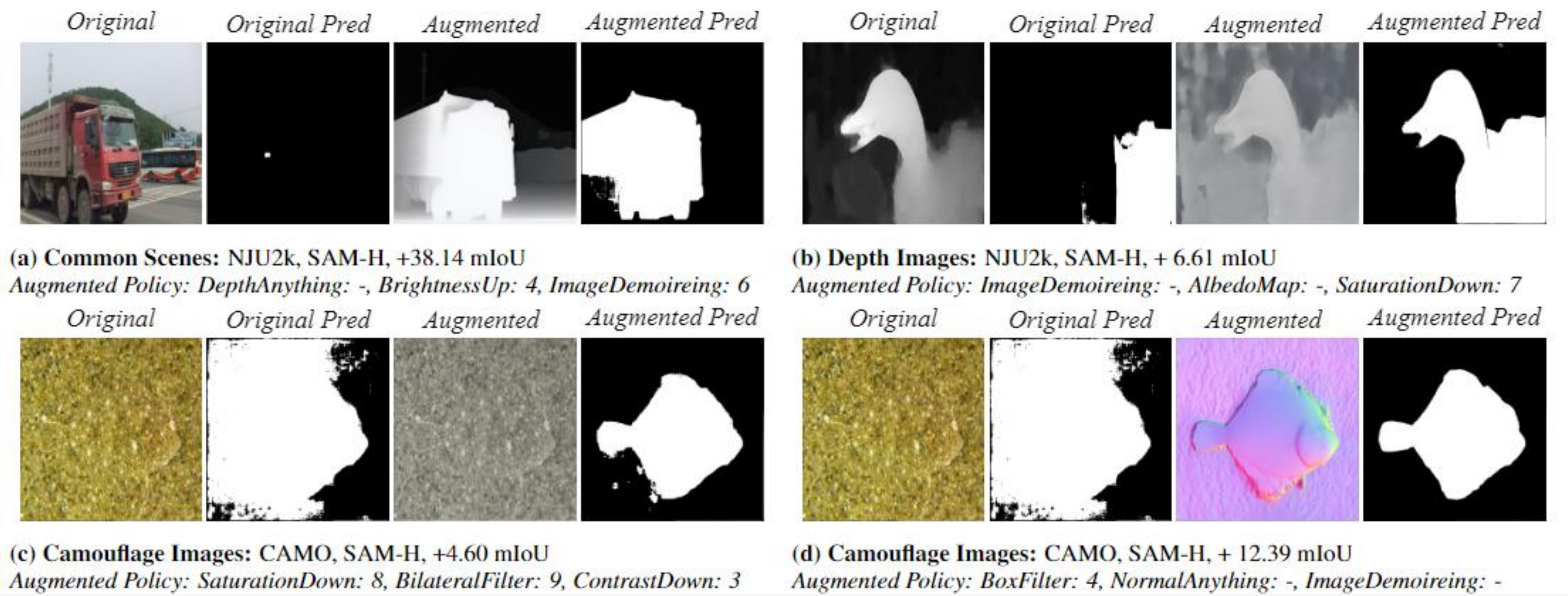
Main Result (10-shot)

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

Backbone	Dataset Type	Dataset Name	SAM [17]	SAM-FT [11]	SAM-IA-32 (Ours)	Augmented Policy
ViT-B	Common Scenes	NJU2K	41.35	45.24	**71.06**	DPE-BE:6-CE:6
		VT1K	47.11	50.37	**64.95**	BE:4-DPE-IDM
	Camouflage Objects	CAMO	37.92	38.71	**62.76**	NSE-SE:0-DPE
		COD10K	49.78	49.89	**59.13**	SE:0-DPE-GC:0
		NC4K	49.09	49.81	**64.37**	DPE-SP:6-AME
	Depth Map	NJU2K Depth	55.83	60.24	**63.44**	GC:4-CE:7-IDM
	Thermal Map	VT1K-T	57.38	58.46	**58.66**	SE:1-IDB-DPE
	Medical Images	Kvasir	62.13	64.44	**71.34**	DB:4-NSE-DB:0
	Ultrasound	BUSI	52.11	57.68	**62.21**	DB:8-GB:6-BF:9
	Industrial Data	MTSD	54.84	**55.04**	53.14	RBU-IDM-SR:1
		KolektorSDD2	46.57	**53.72**	51.62	SL:8-CE:7-GB:5
ViT-H	Common Scenes	NJU2K	42.50	45.89	**80.69**	DPE-BE:4-IDM
		VT1K	49.69	51.33	**63.80**	NSE-CE:1-DPE
	Camouflage Objects	CAMO	49.00	50.43	**64.34**	BX:4-NSE-IDM
		COD10K	58.39	58.61	**62.50**	CR:3-DPE-GB:6
		NC4K	57.23	57.64	**66.37**	NSE-DPE-IDM
	Depth Map	NJU2K Depth	61.37	64.26	**67.98**	IDM-AME-SR:7
	Thermal Map	VT1K-T	56.58	58.67	**61.32**	AME-BX:7-IDM
	Medical Images	Kvasir	63.29	65.21	**71.81**	NSE-RBU-DB:4
	Ultrasound	BUSI	54.54	57.29	**63.86**	ED:6-CE:3-GB:6
	Industrial Data	MTSD	59.24	**59.51**	58.33	SE:4-GC:3-RBU
		KolektorSDD2	48.04	55.67	**55.83**	SG:7-CE:9-GB:8

Table 2. 10-shot training experiment with SAM ViT-B, ViT-H model. Our model brings significant improvement in most of domains.

Qualitative Result



(a) Common Scenes: NJU2k, SAM-H, +38.14 mIoU
Augmented Policy: DepthAnything: -, BrightnessUp: 4, ImageDemoireing: 6

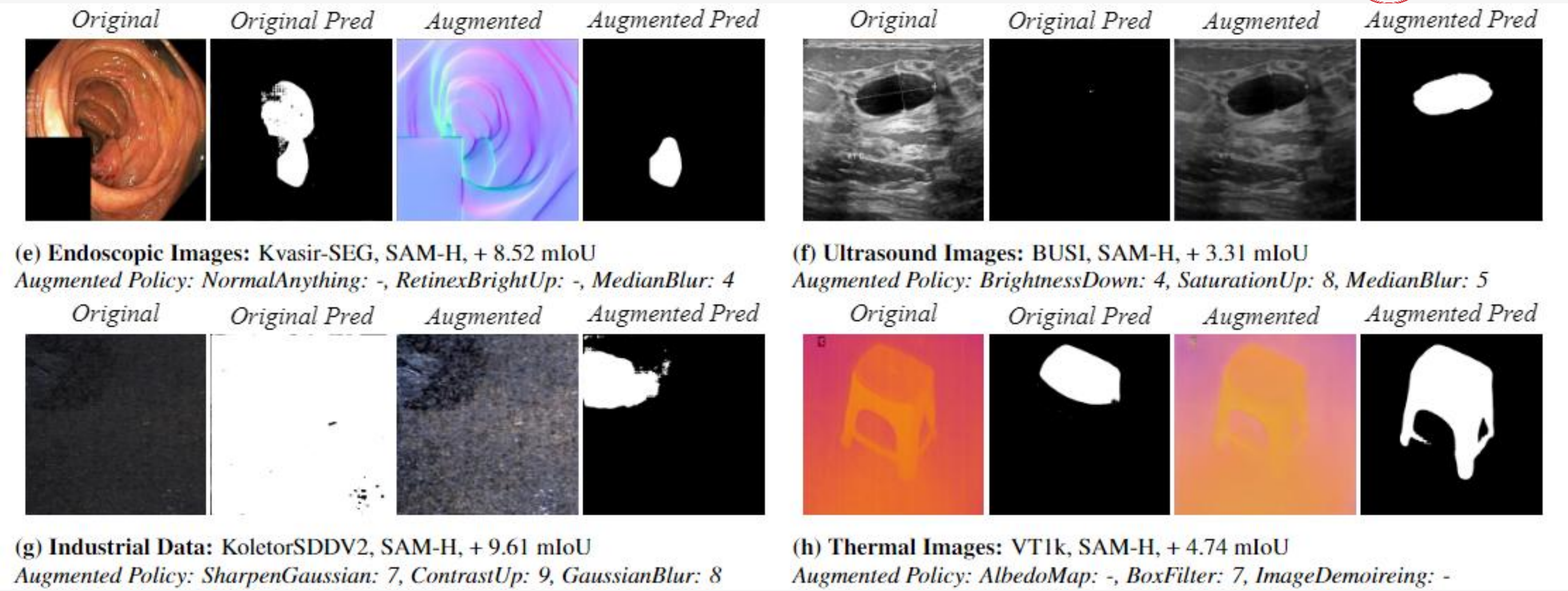
(b) Depth Images: NJU2k, SAM-H, + 6.61 mIoU
Augmented Policy: ImageDemoireing: -, AlbedoMap: -, SaturationDown: 7

(c) Camouflage Images: CAMO, SAM-H, +4.60 mIoU
Augmented Policy: SaturationDown: 8, BilateralFilter: 9, ContrastDown: 3

(d) Camouflage Images: CAMO, SAM-H, + 12.39 mIoU
Augmented Policy: BoxFilter: 4, NormalAnything: -, ImageDemoireing: -

- For common fields, depth estimation helps a lot. (relative depth)
- Learning-based method brings higher upper limit for camouflage objects.

Qualitative Result



(e) Endoscopic Images: Kvasir-SEG, SAM-H, + 8.52 mIoU
Augmented Policy: NormalAnything: -, RetinexBrightUp: -, MedianBlur: 4

(f) Ultrasound Images: BUSI, SAM-H, + 3.31 mIoU
Augmented Policy: BrightnessDown: 4, SaturationUp: 8, MedianBlur: 5

(g) Industrial Data: KoletorSDDV2, SAM-H, + 9.61 mIoU
Augmented Policy: SharpenGaussian: 7, ContrastUp: 9, GaussianBlur: 8

(h) Thermal Images: VT1k, SAM-H, + 4.74 mIoU
Augmented Policy: AlbedoMap: -, BoxFilter: 7, ImageDemoireing: -

- Normal surface estimation makes tumor and polyp more obvious in medical field.
- Rule-based method performs better in industrial detection by hiding unrelated part.

Distillation Result

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Dataset	Augmented Policy	Original		Distill	
		Time	mIoU	Time	mIoU
KvasirSEG	NSE-RBU-DB:4	29245.58	73.48	13.21	68.57
CAMO	SR:8-BF:9-SR:3	8.03	50.00	2.89	50.28
CAMO	SR:6-CR:2-RBU	685.56	50.42	1.96	50.34
VT1k T	AME-GB:7-IDM	122.42	61.30	2.76	61.47
NJU2k Depth	IDM-AME-SR:7	344.88	67.66	8.64	67.31
CAMO	BF:4-NME-IDM	3071.27	65.05	3.68	56.79
BUSI	ED:6-CE:3-GB:6	7.22	63.41	5.32	62.16

Table 3. Performance metrics for various image datasets and policies. All with vitH 10-shot.

- **Save a lot of time when countering time consuming methods (e.g. Normal Surface Estimation), while drops in performance.**

Ablation Study (Reward Design)

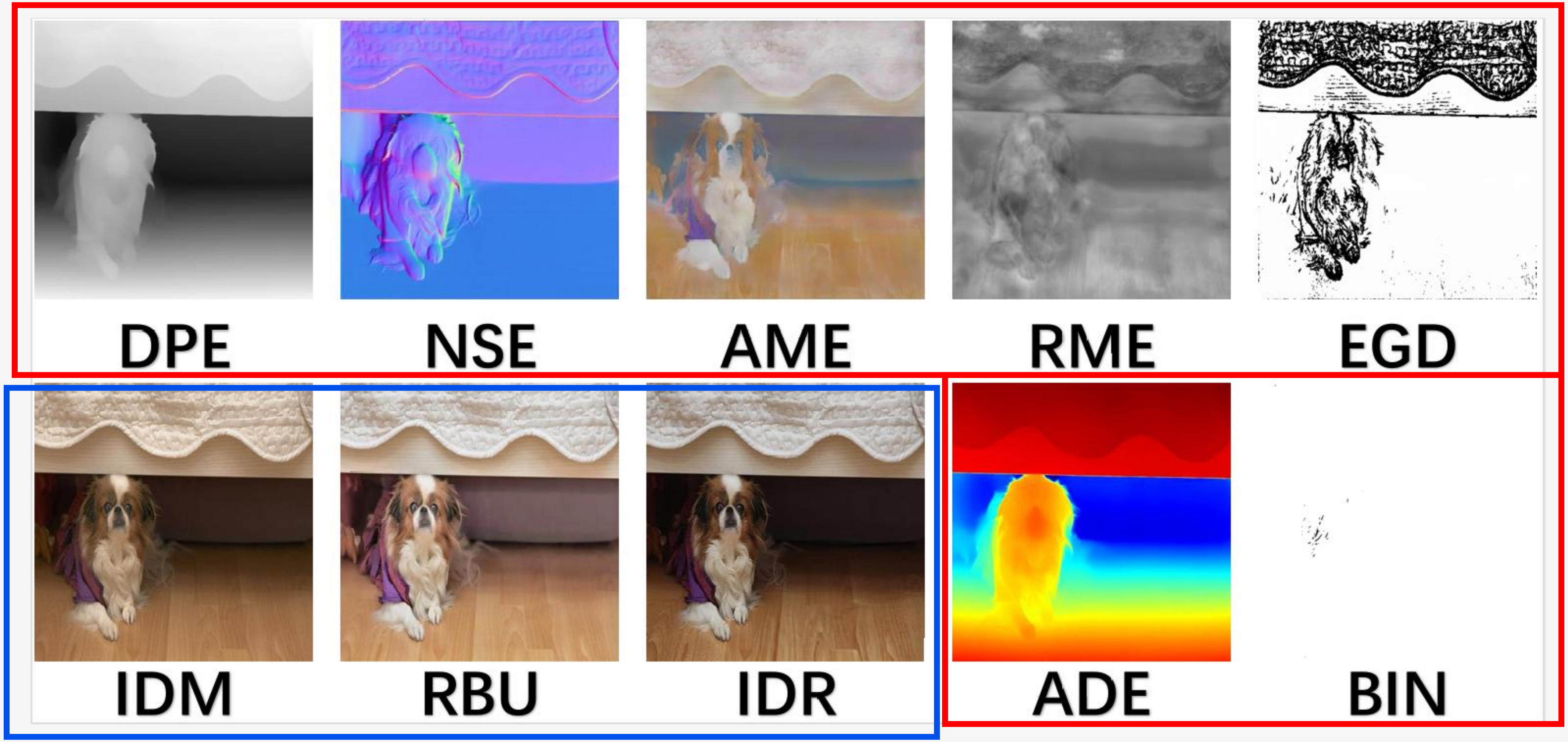
Dataset	Reward Function	Best Policy	mIoU	Improvement
NJU2k	Single	BF:8-SR:6-SR:6	47.95	+5.45
NJU2k	SingleResidual	DPE-LG:6-AME	80.43	+37.93
NJU2k	Three	DPE-BE:4-IDM	80.69	+38.19
NJU2k	All	DPE- BE:6-IDM	78.78	+36.28
CAMO	Single	SE:3-SR:0-GC:2	50.95	-2.37
CAMO	SingleResidual	BF:4-NSE-BF:6	63.44	+10.12
CAMO	Three	BF:4-NSE-IDM	64.34	+11.02
CAMO	All	BF:4-NSE-BF:1	63.41	+10.09

Table 4. Performance comparison of different reward functions. All with vitH 10-shot.

- **It's one-sided to judge an augmented policy's performance on single image.**
- **Judge it on all of the training images will be time-consuming and easily lead to overfitting.**
- **Decide reward on residual or one with other two randomly picked samples will be effective.**

Ablation Study (Searching Space Design)

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY



Ablation Study (Searching Space Design)

Dataset	Mild-25	Radical-29	Complete-32
CAMO Augmented Policy	51.14 SR:7-SR:1-SR:1	64.29 BX:7-GB:9-DPE	64.34 BX:4-NSE-IDM
VT1k T Augmented Policy	59.19 IDB-SE:3-BR:9	59.43 SR:6-DB:3-RME	61.32 AME-BX:7-IDM
NJU2k Augmented Policy	52.28 IDM-SR:3-DB:0	79.61 DPE-LG:2-AME	80.69 DPE-BE:4-IDM
NJU2k Depth Augmented Policy	65.36 CE:7-CE:2-GB:6	67.70 SE:2-AME-BF:6	67.92 IDM-AME-SR:7

Table 5. Performance comparison of different searching spaces. All with vitH 10-shot.

- **Radical methods will bring great enhancement in performance for some domains.**
- **Mild methods will improve on basis of radical methods**
- **Both two kinds of method is helpful to final result.**

Ablation Study (Random vs Reinforcement)

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Dataset	Random 1000	Random 10000	Reinforcement Learning
CAMO Augmented Policy	61.23 DL:6-DPE-SE:3	57.31 ADE-GB:7-GC:6	**64.34** BX:4-NSE-IDM
VT1k T Augmented Policy	59.45 LG:0-SL:7-IDM	59.91 AME-CR:8-IDB	**61.32** AME-BX:7-IDM
NJU2k Augmented Policy	71.63 DPE-IDM-IDM	65.72 SR:3-IDM-RBU	**80.69** DPE-BE:4-IDM
NJU2k Depth Augmented Policy	66.73 SL:2-ED:2-AME	66.73 SL:2-ED:2-AME	**67.92** IDM-AME-SR:7

Table 6. Performance comparison of completely random step and record best and reinforcement learning. All with vitH 10-shot.

- **Randomly selection will not always lead to better results when more steps are taken.**
- **Use reinforcement learning brings much more stable improvement and helps in faster convergence.**

Thank you for listening!