NEURAL INFORMATION
PROCESSING SYSTEMS

Samba: Severity-aware Recurrent Modeling for Cross-domain Medical Image Grading

Qi Bi, Jingjun Yi, Hao Zheng, Wei Ji, Haolan Zhan, Yawen Huang, Yuexiang Li, Yefeng Zheng

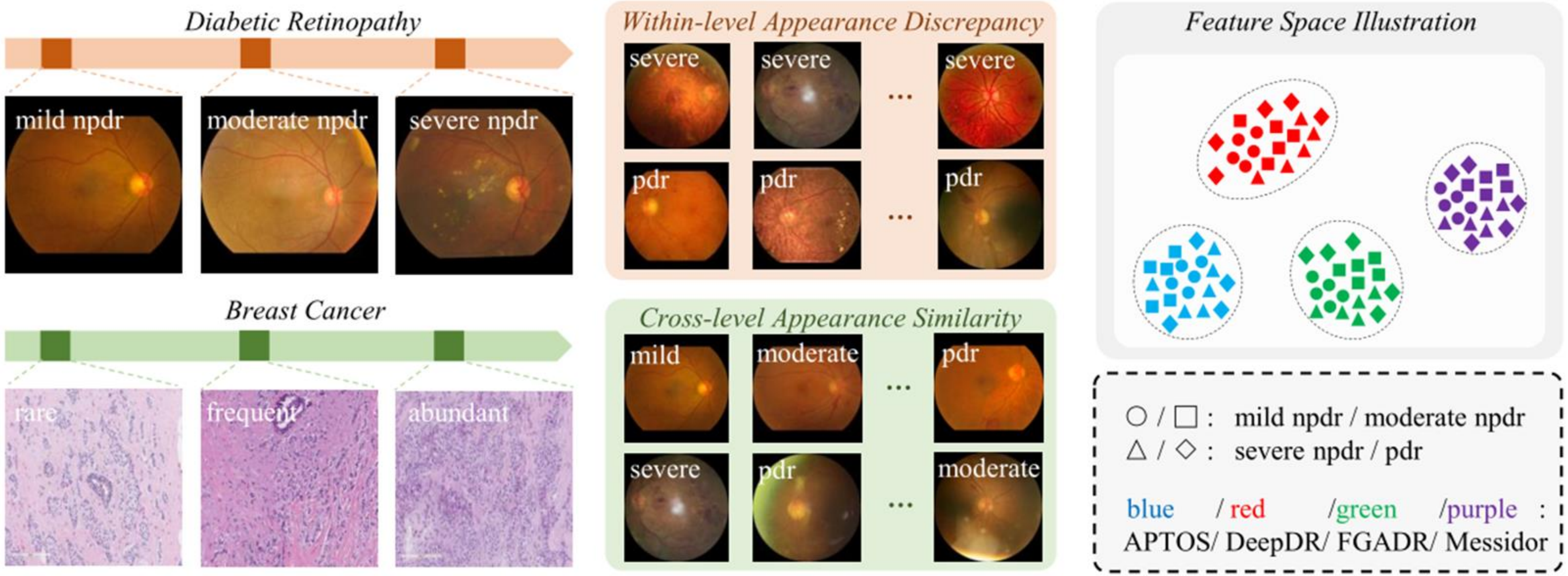
1. Westlake University, China   2. Jarvis Research Center, Tencent Youtu Lab, China

WESTLAKE
UNIVERSITY

腾讯优图

Problem Statement

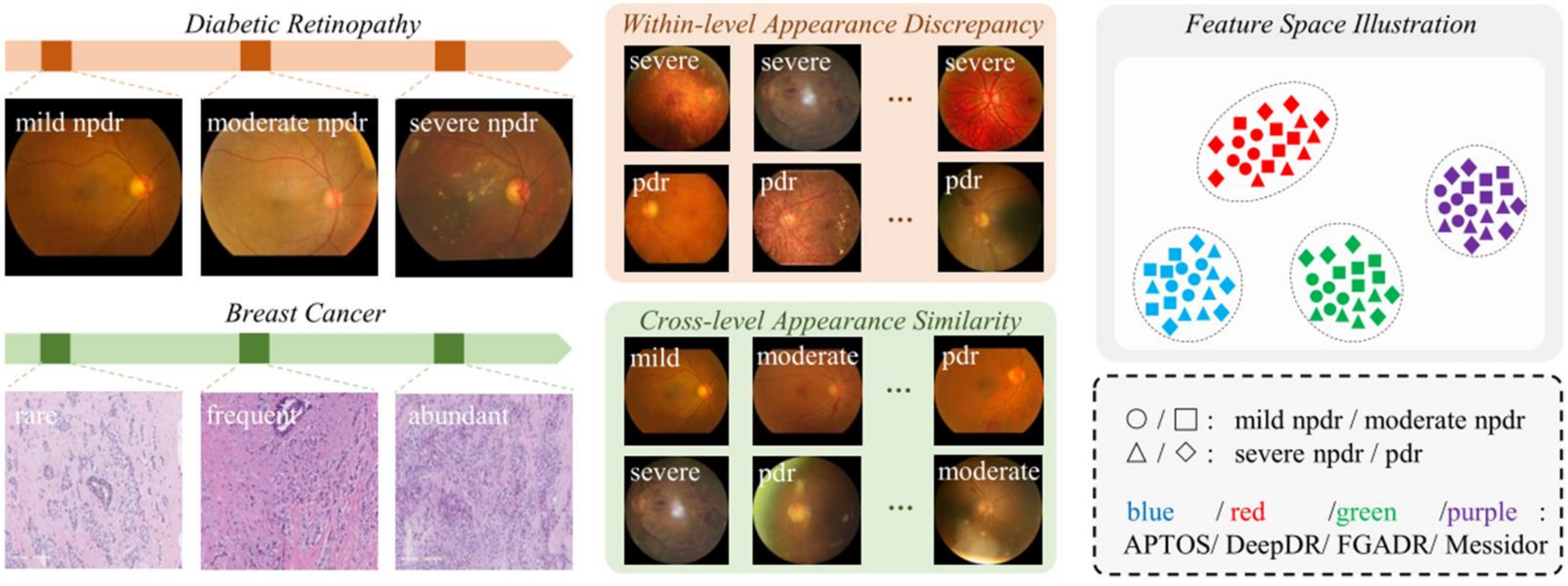
- Disease grading aims to assess the severity level of a disease or a pathological region from a medical Image
- The development of a disease is a continuous progress


Diabetic Retinopathy
mild npdr
moderate npdr
severe npdr
Breast Cancer
rare
frequent
abundant
Within-level Appearance Discrepancy
severe
severe
severe
pdr
pdr
pdr
Cross-level Appearance Similarity
mild
moderate
pdr
severe
pdr
moderate
Feature Space Illustration
○ / □ : mild npdr / moderate npdr
△ / ◇ : severe npdr / pdr
blue / red /green /purple :
APTOS/ DeepDR/ FGADR/ Messidor

(a) Continuous disease development (b) Within-/cross-level ambiguity (c) Cross-domain clustering challenge

Problem Statement

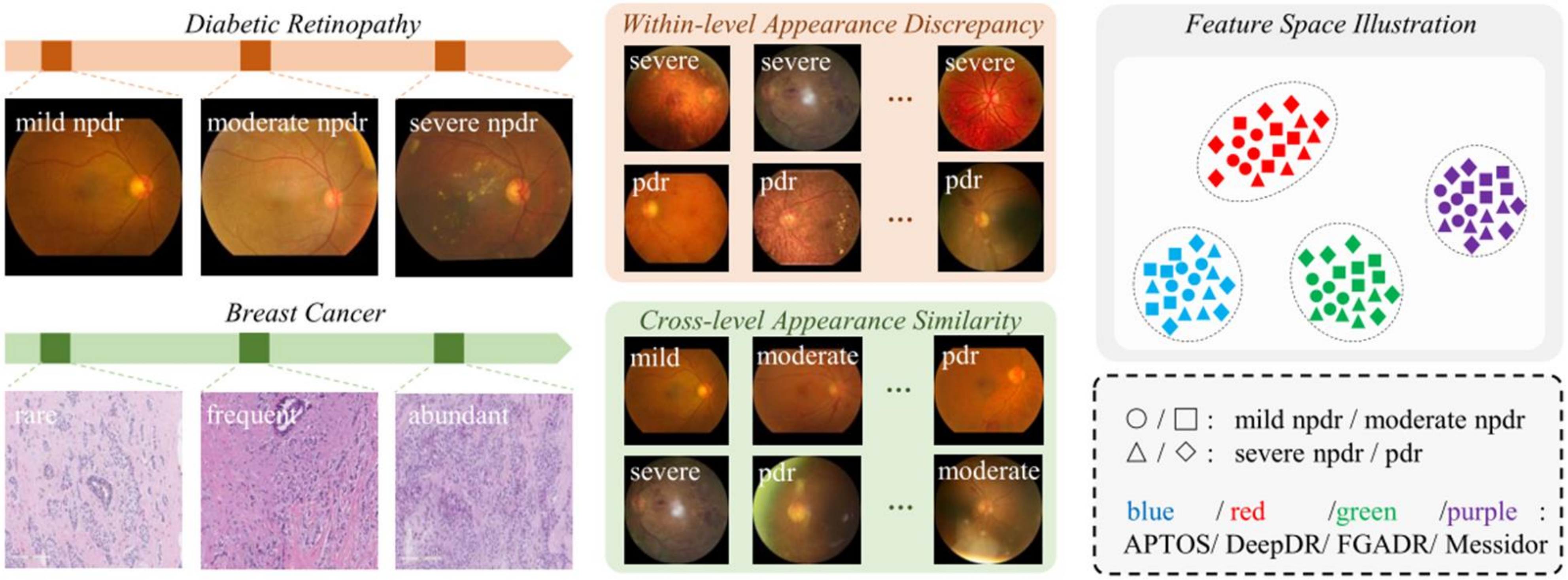
- Assumption: the medical images used for training and inference are independently & identically distributed (i.i.d.)
- Within-level discrepancy & cross-level similarity


Diabetic Retinopathy
mild npdr
moderate npdr
severe npdr
Breast Cancer
rare
frequent
abundant
Within-level Appearance Discrepancy
severe
severe
severe
pdr
pdr
pdr
Cross-level Appearance Similarity
mild
moderate
pdr
severe
pdr
moderate
Feature Space Illustration
○ / □ : mild npdr / moderate npdr
△ / ◇ : severe npdr / pdr
blue / red /green /purple :
APTOS/ DeepDR/ FGADR/ Messidor

(a) Continuous disease development (b) Within-/cross-level ambiguity (c) Cross-domain clustering challenge

Problem Statement

- From ***the same unseen domain***, instead of those from ***the same grade level***, to be clustered in the feature space


Diabetic Retinopathy
mild npdr
moderate npdr
severe npdr
Breast Cancer
rare
frequent
abundant
Within-level Appearance Discrepancy
severe
severe
...
severe
pdr
pdr
...
pdr
Cross-level Appearance Similarity
mild
moderate
...
pdr
severe
pdr
...
moderate
Feature Space Illustration
○ / □ : mild npdr / moderate npdr
△ / ◇ : severe npdr / pdr
blue / red /green /purple :
APTOS/ DeepDR/ FGADR/ Messidor

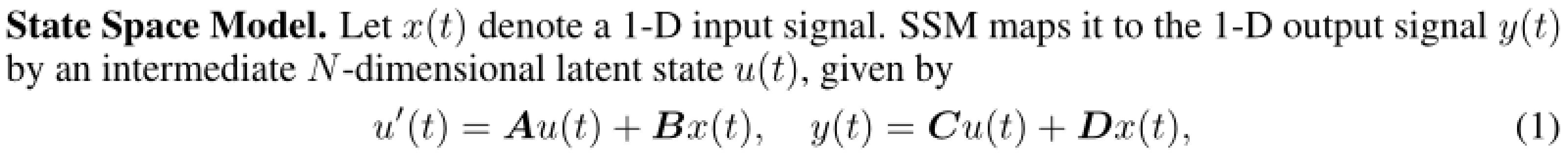
(a) Continuous disease development (b) Within-/cross-level ambiguity (c) Cross-domain clustering challenge

What's New?

- We develop a Severity-aware Recurrent Modeling, dubbed as Samba, for general disease grading within- and cross-domain medical images.
- We propose to encode the image patches in a recurrent manner to capture the decisive lesions and transport the critical information.
- An EM-based state recalibration mechanism is designed to reduce the impacts of cross-domain variants by mapping the feature embeddings into a compact space.
- Extensive experiments on three cross-domain disease grading benchmarks show the effectiveness of the Samba against the baseline.

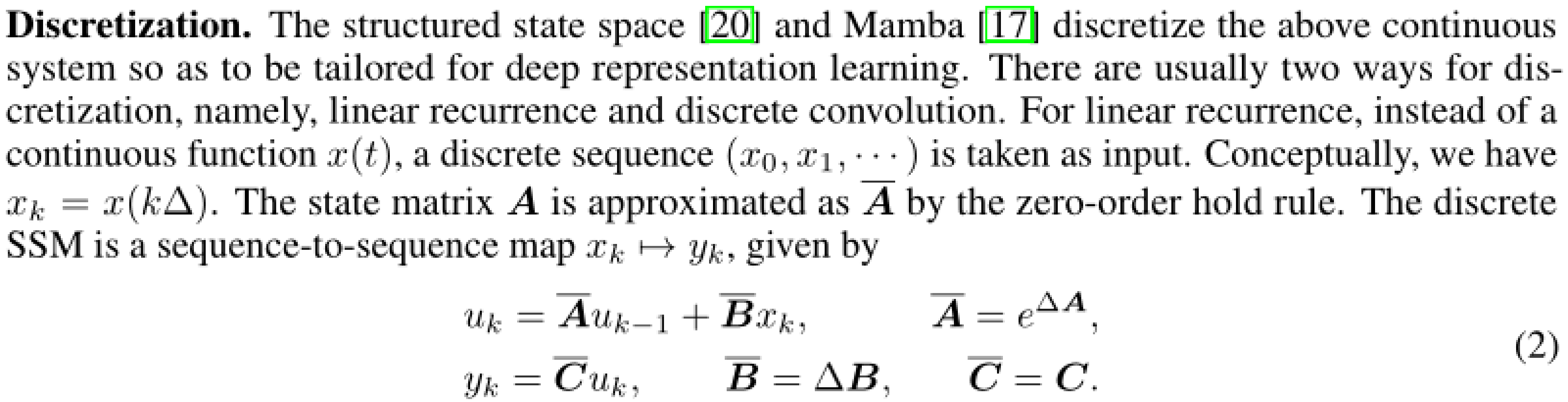
Methodology

- What's Selective State Model (SSM)?



State Space Model. Let $x(t)$ denote a 1-D input signal. SSM maps it to the 1-D output signal $y(t)$ by an intermediate N-dimensional latent state $u(t)$, given by

$$u'(t) = \boldsymbol{A}u(t) + \boldsymbol{B}x(t), \quad y(t) = \boldsymbol{C}u(t) + \boldsymbol{D}x(t), \tag{1}$$



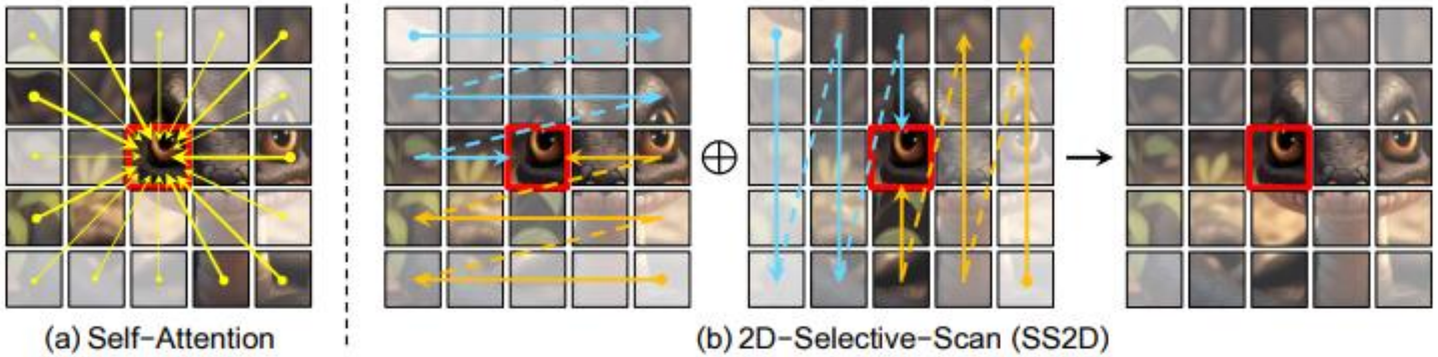
Discretization. The structured state space [20] and Mamba [17] discretize the above continuous system so as to be tailored for deep representation learning. There are usually two ways for discretization, namely, linear recurrence and discrete convolution. For linear recurrence, instead of a continuous function $x(t)$, a discrete sequence $(x_0, x_1, \cdots)$ is taken as input. Conceptually, we have $x_k = x(k\Delta)$. The state matrix $\boldsymbol{A}$ is approximated as $\overline{\boldsymbol{A}}$ by the zero-order hold rule. The discrete SSM is a sequence-to-sequence map $x_k \mapsto y_k$, given by

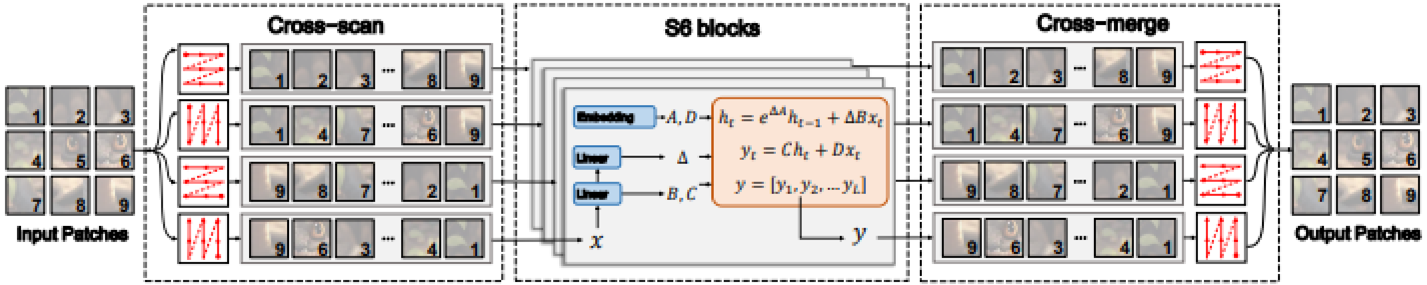
$$\begin{aligned} u_k &= \overline{\boldsymbol{A}}u_{k-1} + \overline{\boldsymbol{B}}x_k, \qquad \overline{\boldsymbol{A}} = e^{\Delta \boldsymbol{A}}, \\ y_k &= \overline{\boldsymbol{C}}u_k, \qquad \overline{\boldsymbol{B}} = \Delta\boldsymbol{B}, \qquad \overline{\boldsymbol{C}} = \boldsymbol{C}. \end{aligned} \tag{2}$$

Methodology

- What's Selective State Model (SSM)?

  Selective scan in Vmamba



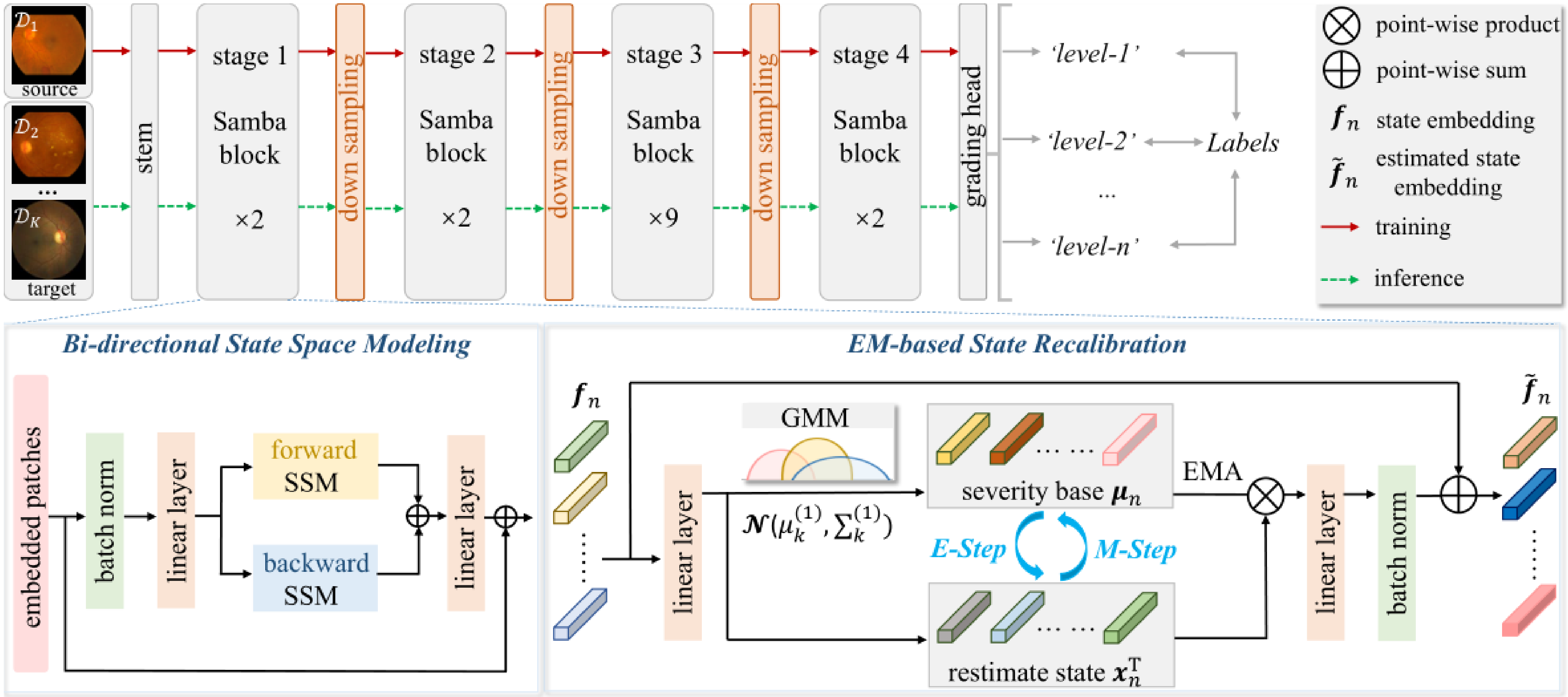


Disclaimer: image from

Vmamba: Visual state space model. arXiv preprint arXiv:2401.10166, 2024.

Methodology

- Framework Overview

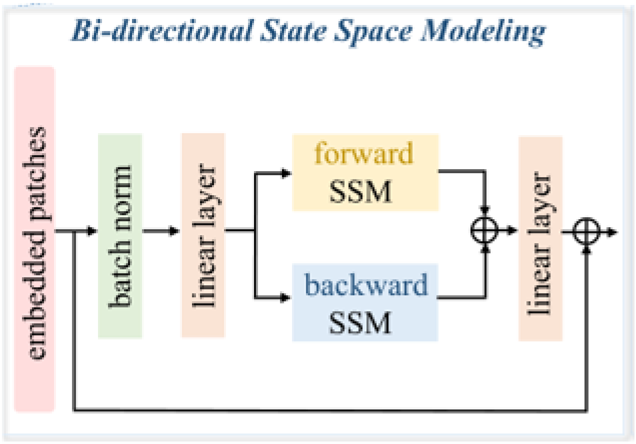


Methodology

- Key Module 1: Bi-directional Design

-- The embeddings are first input to the bidirectional Mamba layers to store and transport the information about decisive lesions.

-- With the guidance of global severity awareness, the update of hidden states can selectively ignore information about low-level lesions, primarily preserving information about the most severe lesions.



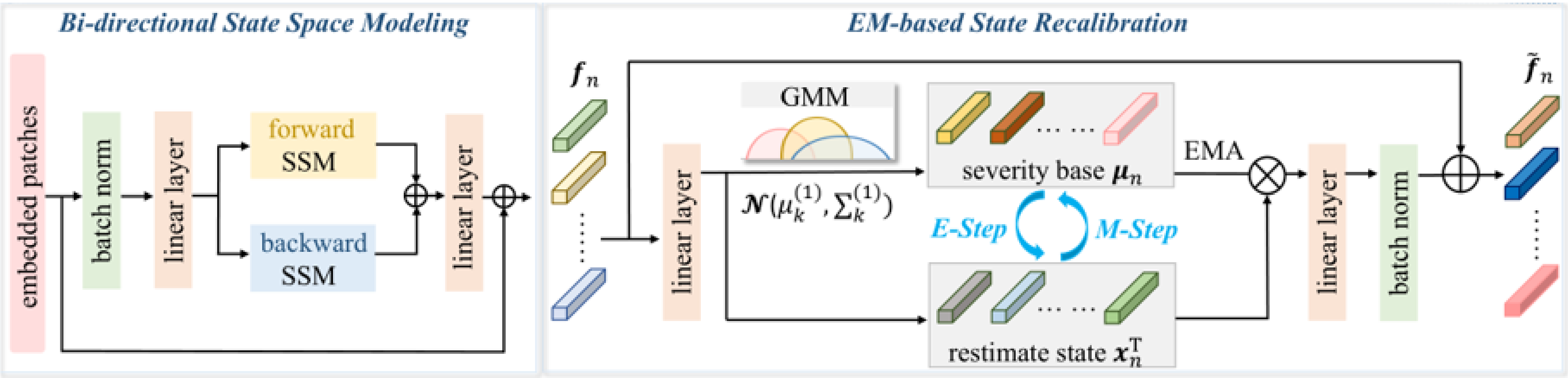
Methodology

- Key Module 2: E-M based State Recalibration

-- To reduce the impact of domain shift, we aim to map the features into a more compact space by feature recalibration.

-- Feature distribution of background and grading-related lesions is modeled by a Gaussian Mixture Model (GMM)

$$p(\boldsymbol{f}_n) = \sum_{k=1}^{K} z_{nk}\mathcal{N}(\boldsymbol{f}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$



Bi-directional State Space Modeling
embedded patches
batch norm
linear layer
forward SSM
backward SSM
linear layer
EM-based State Recalibration
f_n
linear layer
GMM
$\mathcal{N}(\mu_k^{(1)}, \Sigma_k^{(1)})$
severity base $\boldsymbol{\mu}_n$
E-Step
M-Step
restimate state $\boldsymbol{x}_n^{\mathrm{T}}$
EMA
linear layer
batch norm
$\tilde{f}_n$

Methodology

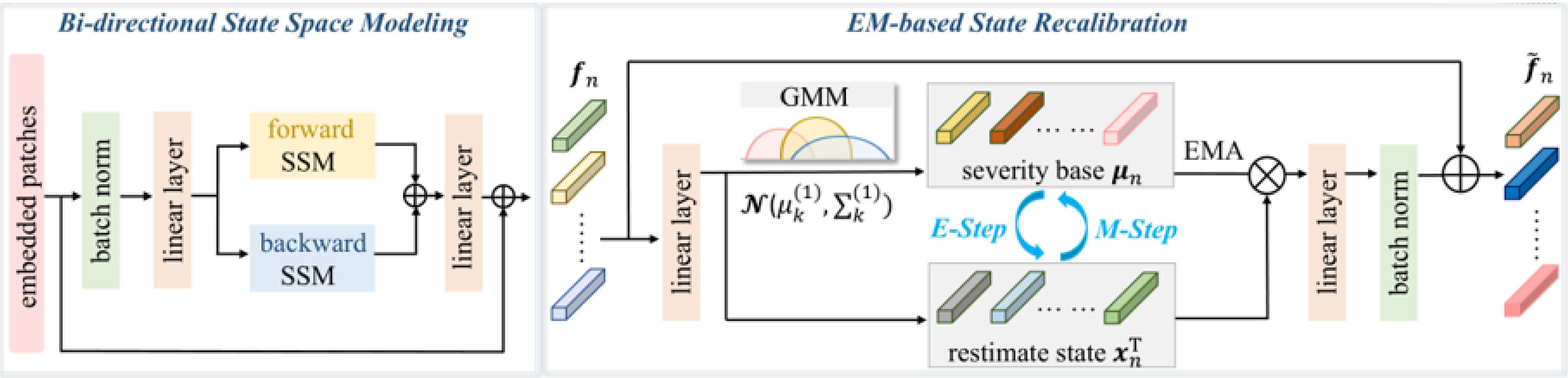
- Key Module 2: E-M based State Recalibration

-- E-step: estimate the severity basis

$$z_{nk} = \frac{\mathcal{K}(\boldsymbol{f}_n, \boldsymbol{\mu}_k)}{\sum_{i=1}^{K} \mathcal{K}(\boldsymbol{f}_n, \boldsymbol{\mu}_i)}$$

-- M-step: severity base likelihood maximization

$$\boldsymbol{\mu}_k^{t+1} = \frac{1}{\sum_{m=1}^{N_p} z_{mk}^t} \sum_{n=1}^{N_p} z_{nk}^t \boldsymbol{f}_n$$



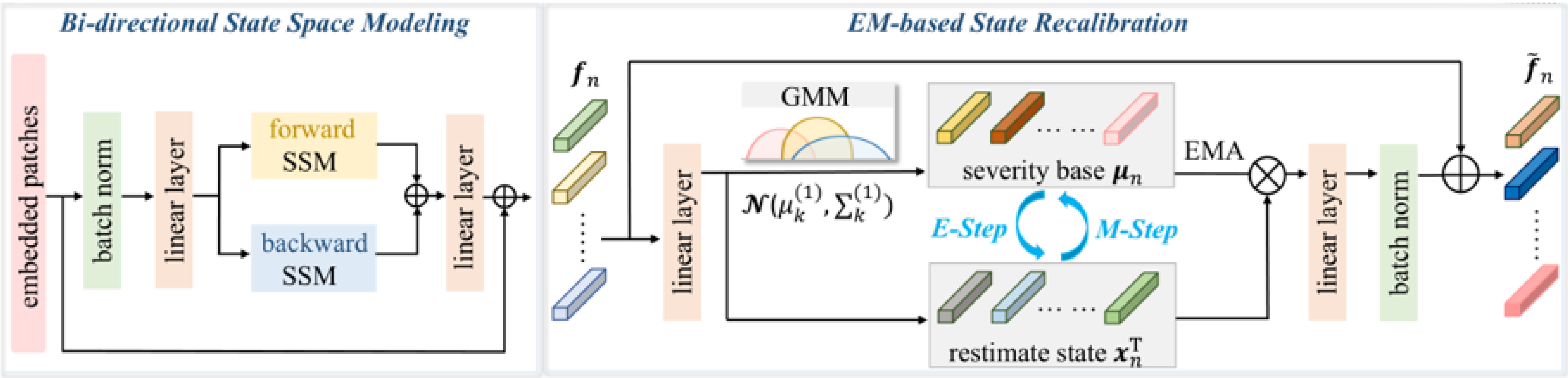
Bi-directional State Space Modeling
embedded patches
batch norm
linear layer
forward SSM
backward SSM
linear layer
EM-based State Recalibration
f_n
linear layer
GMM
$\mathcal{N}(\mu_k^{(1)}, \Sigma_k^{(1)})$
severity base $\boldsymbol{\mu}_n$
E-Step
M-Step
restimate state $\boldsymbol{x}_n^{\mathrm{T}}$
EMA
linear layer
batch norm
$\tilde{f}_n$

Methodology

- Key Module 2: E-M based State Recalibration

-- Moving averaging is adapted to update the bases when training

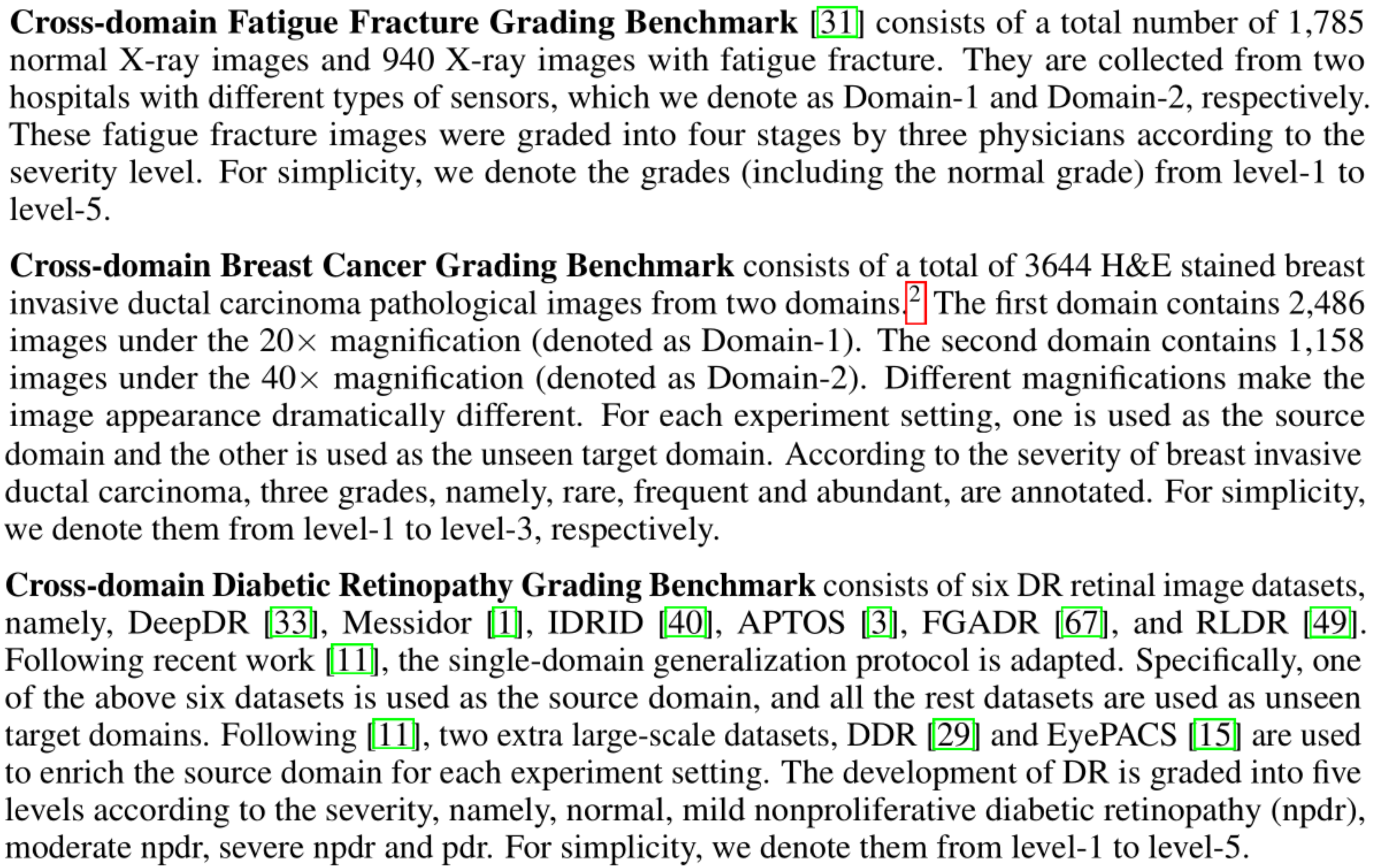
$$\boldsymbol{\mu}^0 \longleftarrow \alpha\boldsymbol{\mu}^0 + (1-\alpha)\overline{\boldsymbol{\mu}}^T$$



Bi-directional State Space Modeling
embedded patches
batch norm
linear layer
forward SSM
backward SSM
linear layer
EM-based State Recalibration
f_n
linear layer
GMM
$\mathcal{N}(\mu_k^{(1)}, \Sigma_k^{(1)})$
severity base $\boldsymbol{\mu}_n$
E-Step
M-Step
restimate state $\boldsymbol{x}_n^T$
EMA
linear layer
batch norm
$\tilde{f}_n$

Experiment

- Datasets



Cross-domain Fatigue Fracture Grading Benchmark [31] consists of a total number of 1,785 normal X-ray images and 940 X-ray images with fatigue fracture. They are collected from two hospitals with different types of sensors, which we denote as Domain-1 and Domain-2, respectively. These fatigue fracture images were graded into four stages by three physicians according to the severity level. For simplicity, we denote the grades (including the normal grade) from level-1 to level-5.

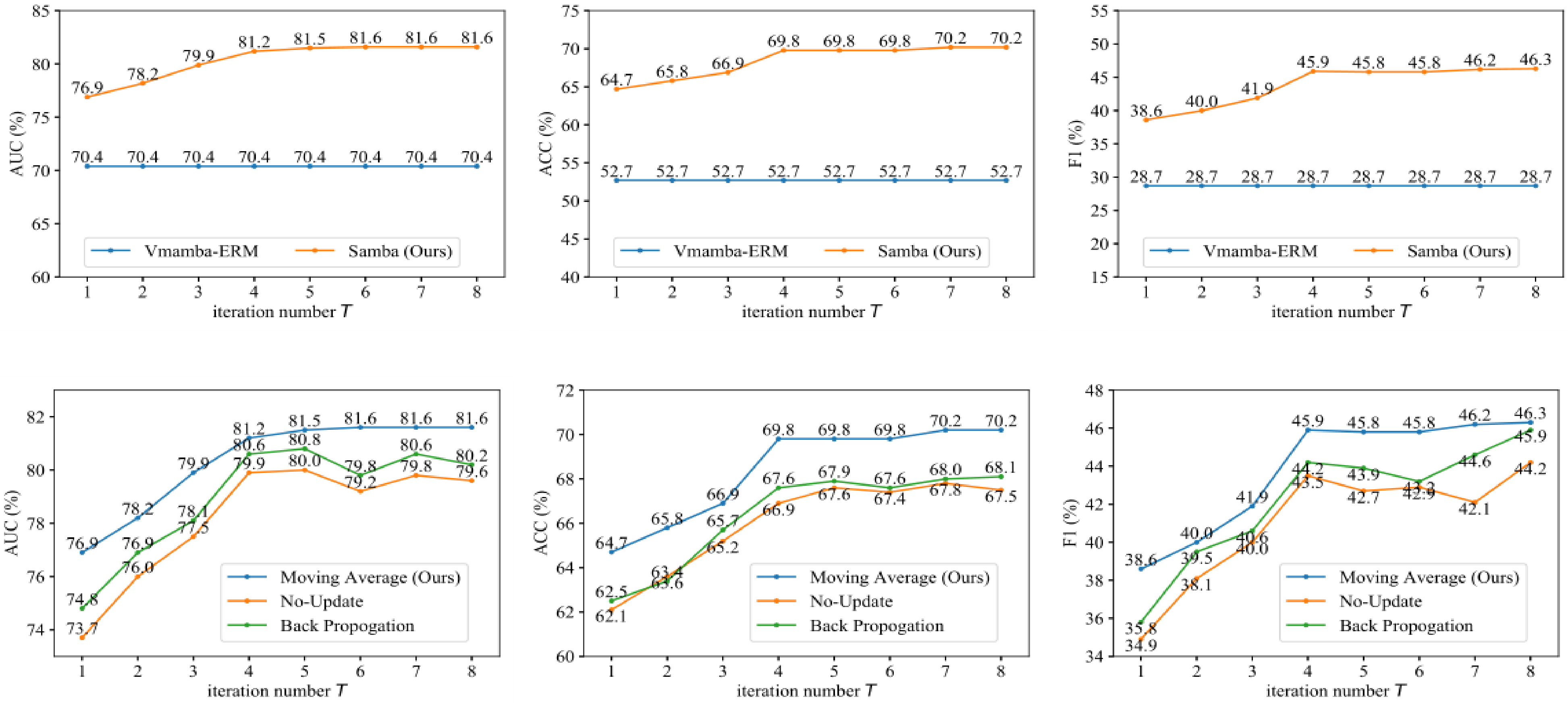
Cross-domain Breast Cancer Grading Benchmark consists of a total of 3644 H&E stained breast invasive ductal carcinoma pathological images from two domains.[2] The first domain contains 2,486 images under the 20× magnification (denoted as Domain-1). The second domain contains 1,158 images under the 40× magnification (denoted as Domain-2). Different magnifications make the image appearance dramatically different. For each experiment setting, one is used as the source domain and the other is used as the unseen target domain. According to the severity of breast invasive ductal carcinoma, three grades, namely, rare, frequent and abundant, are annotated. For simplicity, we denote them from level-1 to level-3, respectively.

Cross-domain Diabetic Retinopathy Grading Benchmark consists of six DR retinal image datasets, namely, DeepDR [33], Messidor [1], IDRID [40], APTOS [3], FGADR [67], and RLDR [49]. Following recent work [11], the single-domain generalization protocol is adapted. Specifically, one of the above six datasets is used as the source domain, and all the rest datasets are used as unseen target domains. Following [11], two extra large-scale datasets, DDR [29] and EyePACS [15] are used to enrich the source domain for each experiment setting. The development of DR is graded into five levels according to the severity, namely, normal, mild nonproliferative diabetic retinopathy (npdr), moderate npdr, severe npdr and pdr. For simplicity, we denote them from level-1 to level-5.

Experiment

- Experiment 1:

Impact of iteration number T & severity base updating approaches



Experiment

- Experiment 2:

  Recurrent Patch Modeling & Each Component

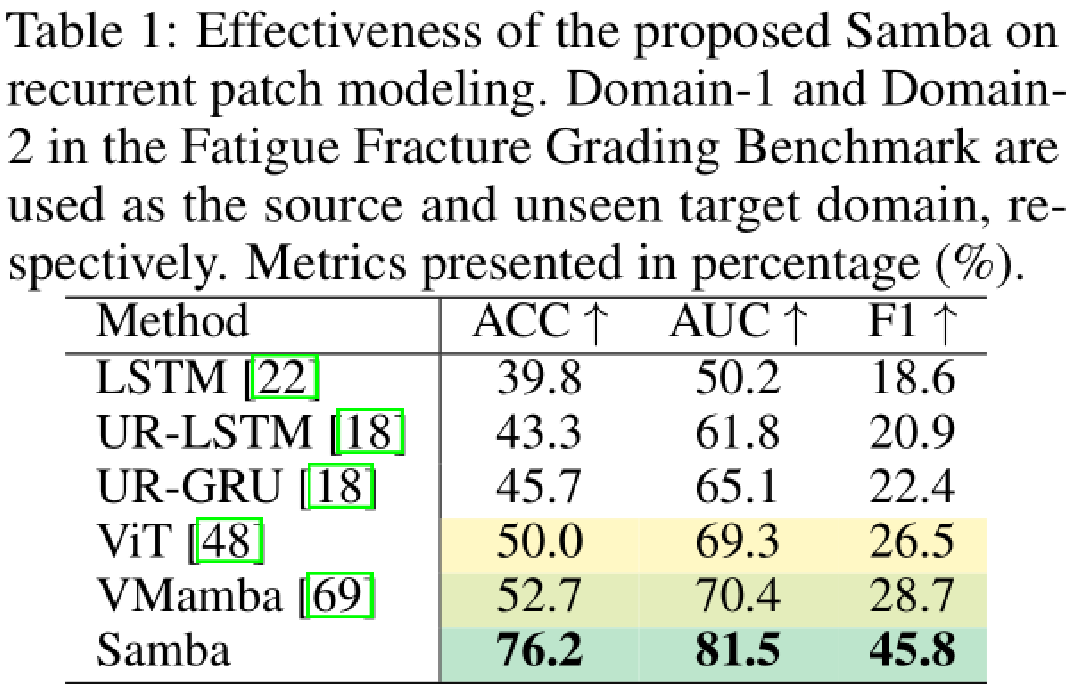


Table 1: Effectiveness of the proposed Samba on recurrent patch modeling. Domain-1 and Domain-2 in the Fatigue Fracture Grading Benchmark are used as the source and unseen target domain, respectively. Metrics presented in percentage (%).

Method	ACC ↑	AUC ↑	F1 ↑
LSTM [22]	39.8	50.2	18.6
UR-LSTM [18]	43.3	61.8	20.9
UR-GRU [18]	45.7	65.1	22.4
ViT [48]	50.0	69.3	26.5
VMamba [69]	52.7	70.4	28.7
Samba	**76.2**	**81.5**	**45.8**

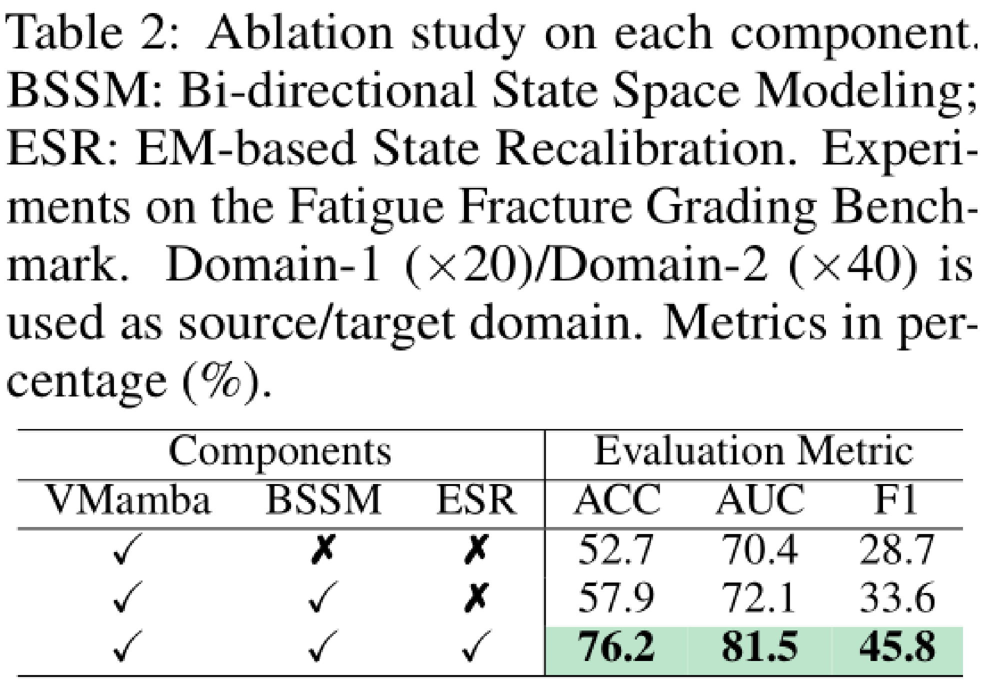


Table 2: Ablation study on each component. BSSM: Bi-directional State Space Modeling; ESR: EM-based State Recalibration. Experiments on the Fatigue Fracture Grading Benchmark. Domain-1 (×20)/Domain-2 (×40) is used as source/target domain. Metrics in percentage (%).

Components			Evaluation Metric		
VMamba	BSSM	ESR	ACC	AUC	F1
✓	✗	✗	52.7	70.4	28.7
✓	✓	✗	57.9	72.1	33.6
✓	✓	✓	**76.2**	**81.5**	**45.8**

Experiment

- Experiment 3:

  Category-wise performance & Impact of GMM number

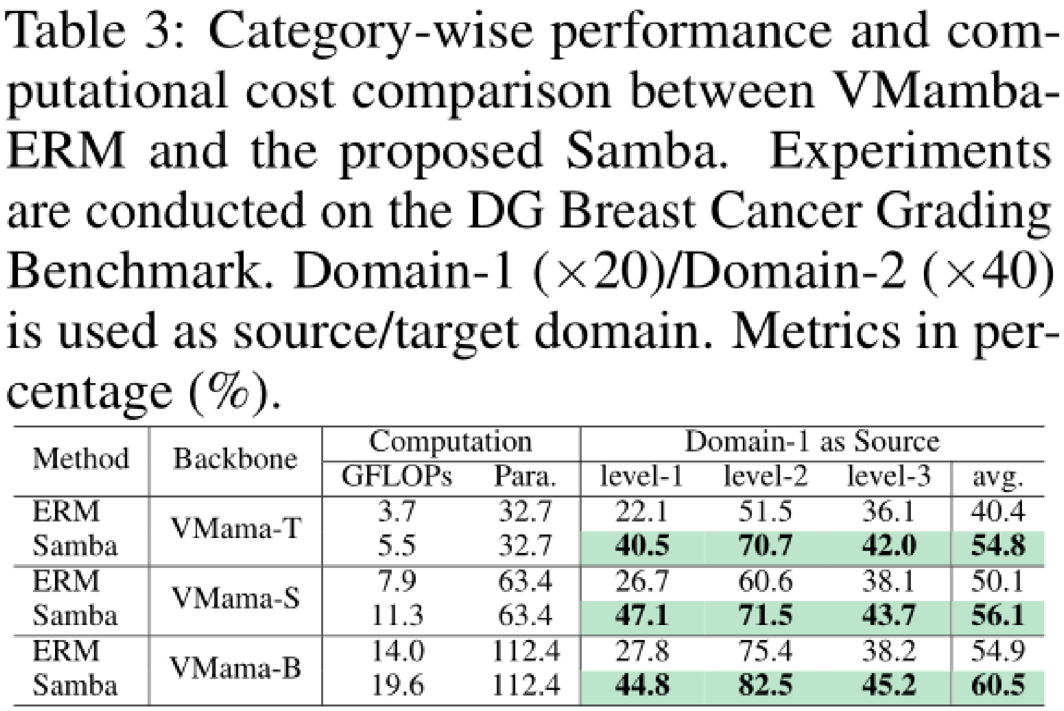


Table 3: Category-wise performance and computational cost comparison between VMamba-ERM and the proposed Samba. Experiments are conducted on the DG Breast Cancer Grading Benchmark. Domain-1 (×20)/Domain-2 (×40) is used as source/target domain. Metrics in percentage (%).

Method	Backbone	Computation		Domain-1 as Source			
		GFLOPs	Para.	level-1	level-2	level-3	avg.
ERM	VMama-T	3.7	32.7	22.1	51.5	36.1	40.4
Samba		5.5	32.7	**40.5**	**70.7**	**42.0**	**54.8**
ERM	VMama-S	7.9	63.4	26.7	60.6	38.1	50.1
Samba		11.3	63.4	**47.1**	**71.5**	**43.7**	**56.1**
ERM	VMama-B	14.0	112.4	27.8	75.4	38.2	54.9
Samba		19.6	112.4	**44.8**	**82.5**	**45.2**	**60.5**

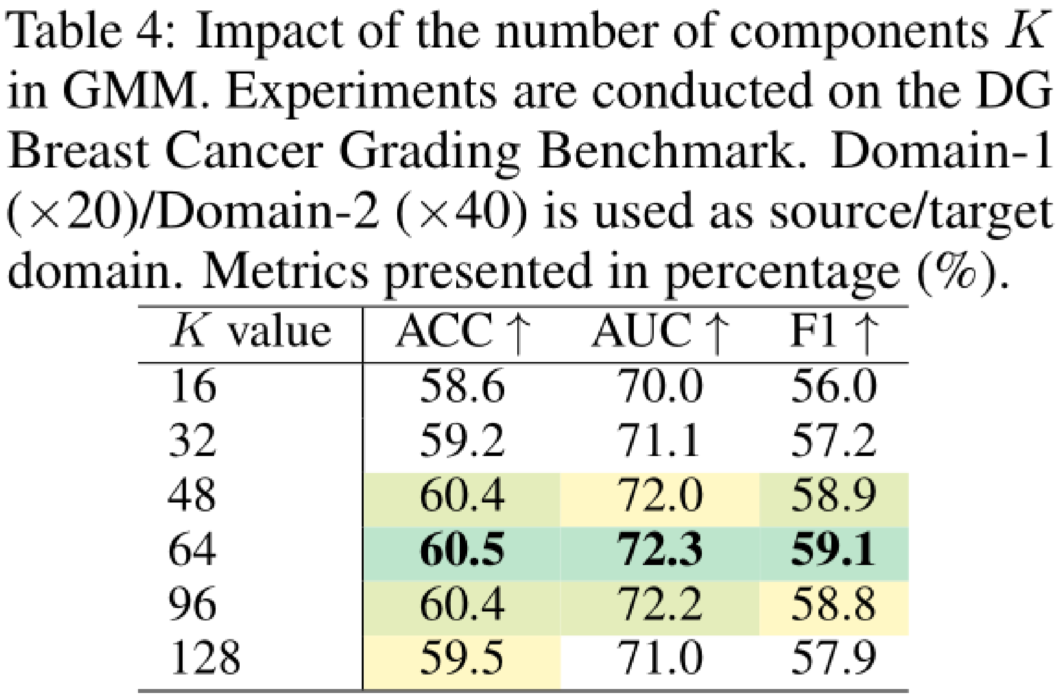


Table 4: Impact of the number of components K in GMM. Experiments are conducted on the DG Breast Cancer Grading Benchmark. Domain-1 (×20)/Domain-2 (×40) is used as source/target domain. Metrics presented in percentage (%).

K value	ACC ↑	AUC ↑	F1 ↑
16	58.6	70.0	56.0
32	59.2	71.1	57.2
48	60.4	72.0	58.9
64	**60.5**	**72.3**	**59.1**
96	60.4	72.2	58.8
128	59.5	71.0	57.9

Experiment

- Experiment 5: Comparison with State-of-the-art

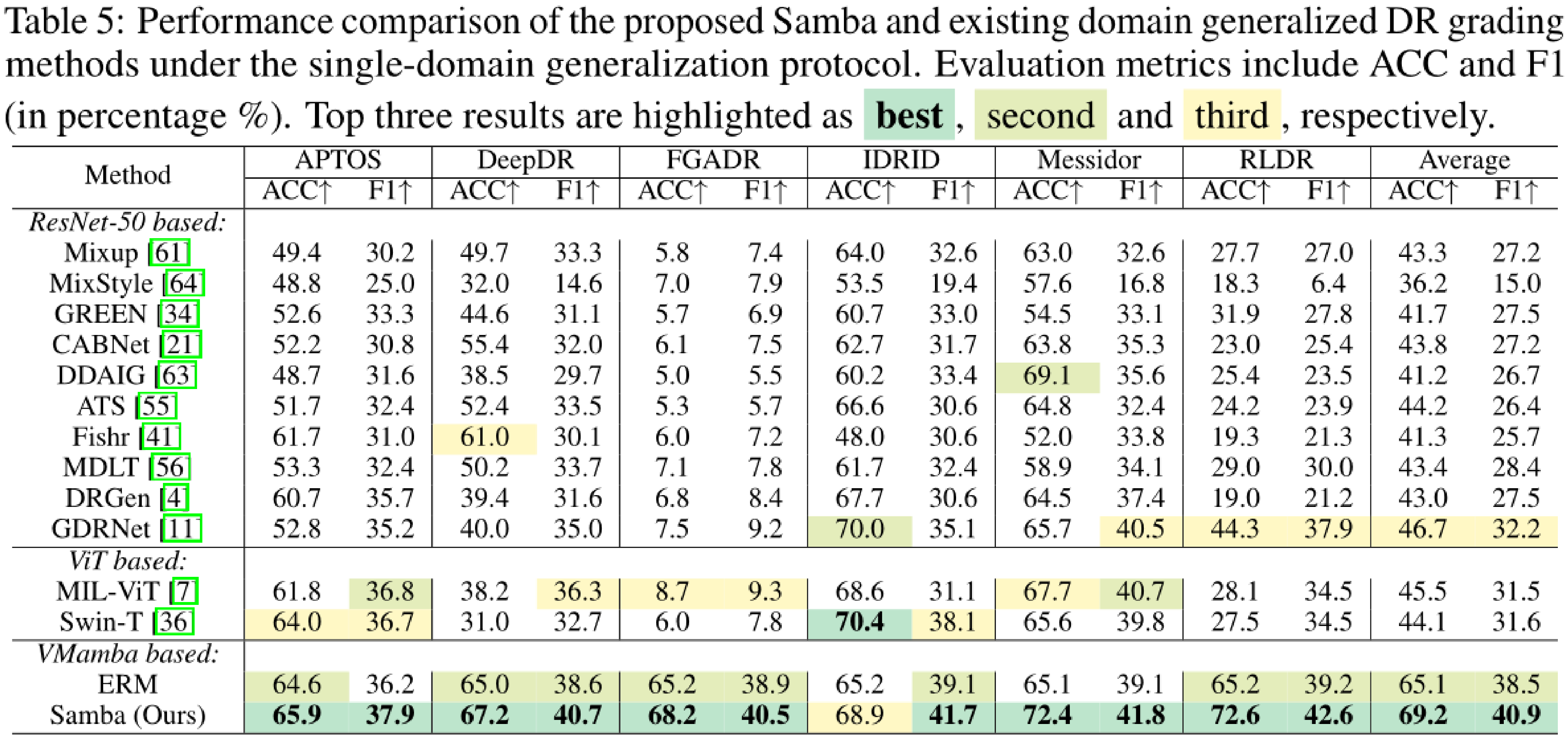


Table 5: Performance comparison of the proposed Samba and existing domain generalized DR grading methods under the single-domain generalization protocol. Evaluation metrics include ACC and F1 (in percentage %). Top three results are highlighted as **best**, second and third, respectively.

Method	APTOS		DeepDR		FGADR		IDRID		Messidor		RLDR		Average	
	ACC↑	F1↑	ACC↑	F1↑	ACC↑	F1↑	ACC↑	F1↑	ACC↑	F1↑	ACC↑	F1↑	ACC↑	F1↑
ResNet-50 based:														
Mixup [61]	49.4	30.2	49.7	33.3	5.8	7.4	64.0	32.6	63.0	32.6	27.7	27.0	43.3	27.2
MixStyle [64]	48.8	25.0	32.0	14.6	7.0	7.9	53.5	19.4	57.6	16.8	18.3	6.4	36.2	15.0
GREEN [34]	52.6	33.3	44.6	31.1	5.7	6.9	60.7	33.0	54.5	33.1	31.9	27.8	41.7	27.5
CABNet [21]	52.2	30.8	55.4	32.0	6.1	7.5	62.7	31.7	63.8	35.3	23.0	25.4	43.8	27.2
DDAIG [63]	48.7	31.6	38.5	29.7	5.0	5.5	60.2	33.4	69.1	35.6	25.4	23.5	41.2	26.7
ATS [55]	51.7	32.4	52.4	33.5	5.3	5.7	66.6	30.6	64.8	32.4	24.2	23.9	44.2	26.4
Fishr [41]	61.7	31.0	61.0	30.1	6.0	7.2	48.0	30.6	52.0	33.8	19.3	21.3	41.3	25.7
MDLT [56]	53.3	32.4	50.2	33.7	7.1	7.8	61.7	32.4	58.9	34.1	29.0	30.0	43.4	28.4
DRGen [4]	60.7	35.7	39.4	31.6	6.8	8.4	67.7	30.6	64.5	37.4	19.0	21.2	43.0	27.5
GDRNet [11]	52.8	35.2	40.0	35.0	7.5	9.2	70.0	35.1	65.7	40.5	44.3	37.9	46.7	32.2
ViT based:														
MIL-ViT [7]	61.8	36.8	38.2	36.3	8.7	9.3	68.6	31.1	67.7	40.7	28.1	34.5	45.5	31.5
Swin-T [36]	64.0	36.7	31.0	32.7	6.0	7.8	**70.4**	38.1	65.6	39.8	27.5	34.5	44.1	31.6
VMamba based:														
ERM	64.6	36.2	65.0	38.6	65.2	38.9	65.2	39.1	65.1	39.1	65.2	39.2	65.1	38.5
Samba (Ours)	**65.9**	**37.9**	**67.2**	**40.7**	**68.2**	**40.5**	68.9	**41.7**	**72.4**	**41.8**	**72.6**	**42.6**	**69.2**	**40.9**

Experiment

- Visualization

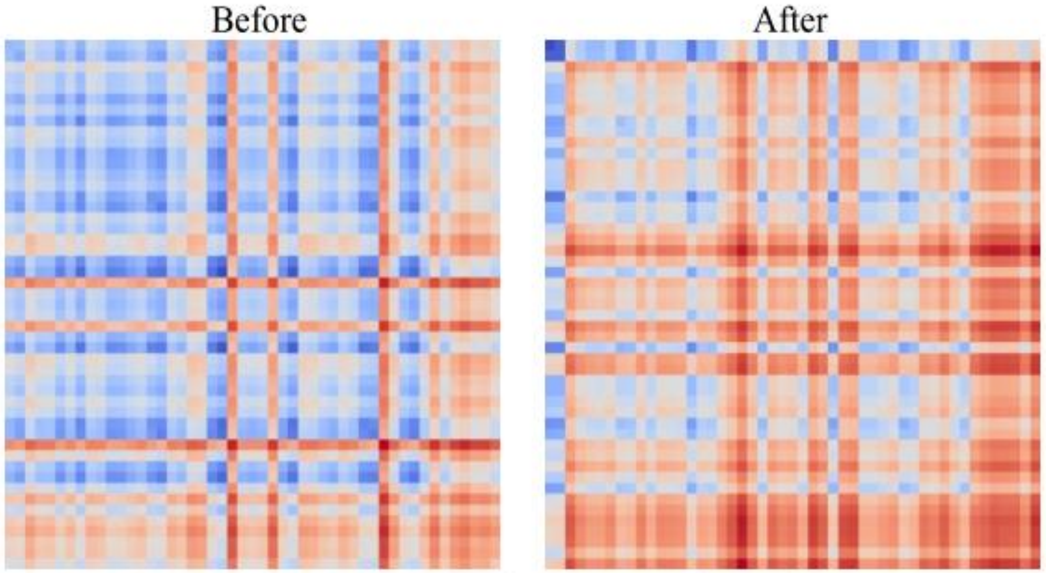


Figure 4: The correlation matrix of each patch embedding before and after processed by the recurrent patch modeling in the forward direction, denoted as ‘Before’ and ‘After’ respectively. The higher correlation, the more red a cell is.

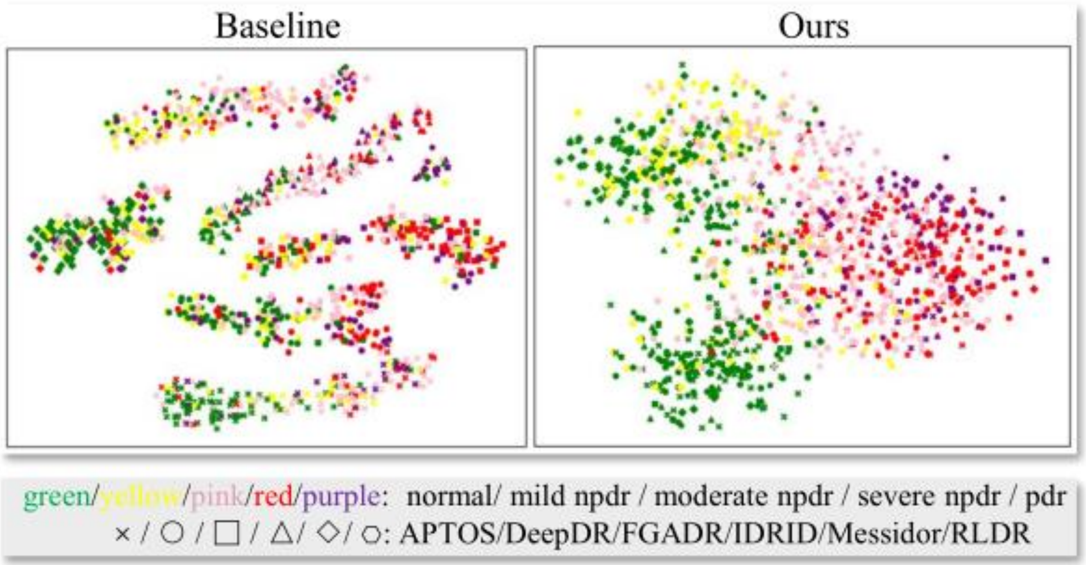


Figure 5: T-SNE visualization of the feature space from the ERM baseline (left), and the proposed Samba (right). APTOS is chosen as the source domain and the rest datasets are used for as target domains.

Experiment

- Visualization

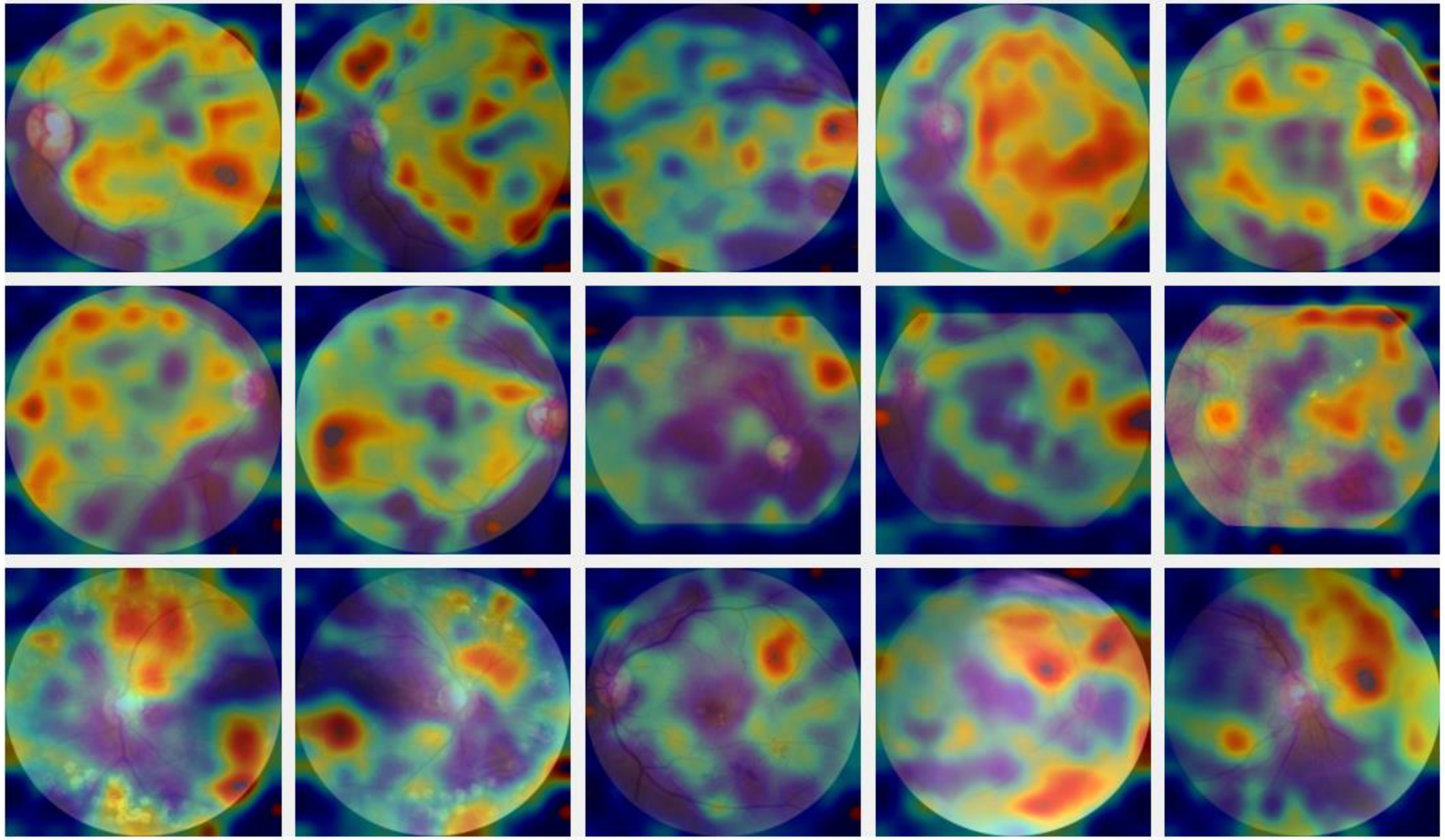


Figure 9: Attention maps of the proposed Samba on retinal images from unseen domains.

Thanks for your attention!