

Takeaway: SC-UNSB creates spatially coherent cytology images, while dual-level distillation aligns structural and semantic features for cross-institution cell detection.

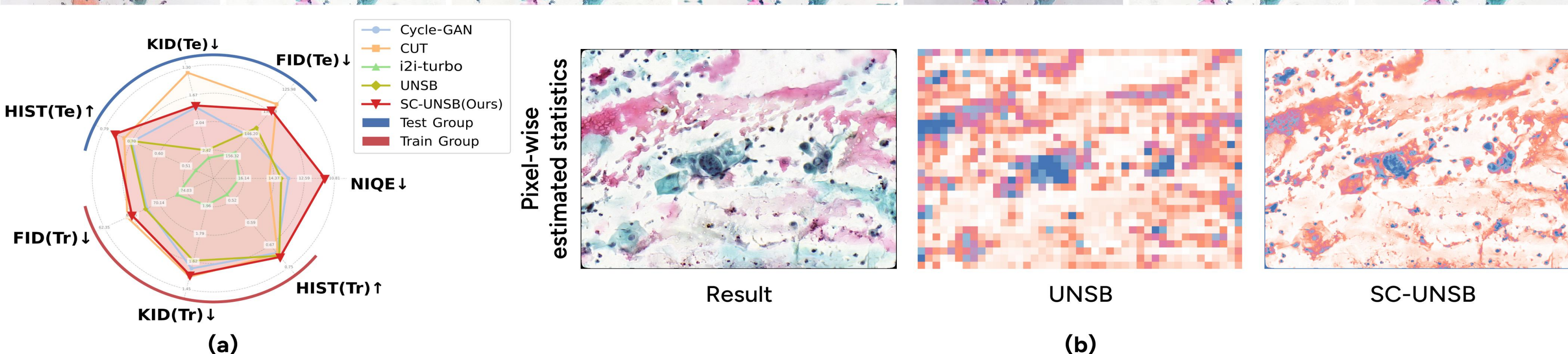
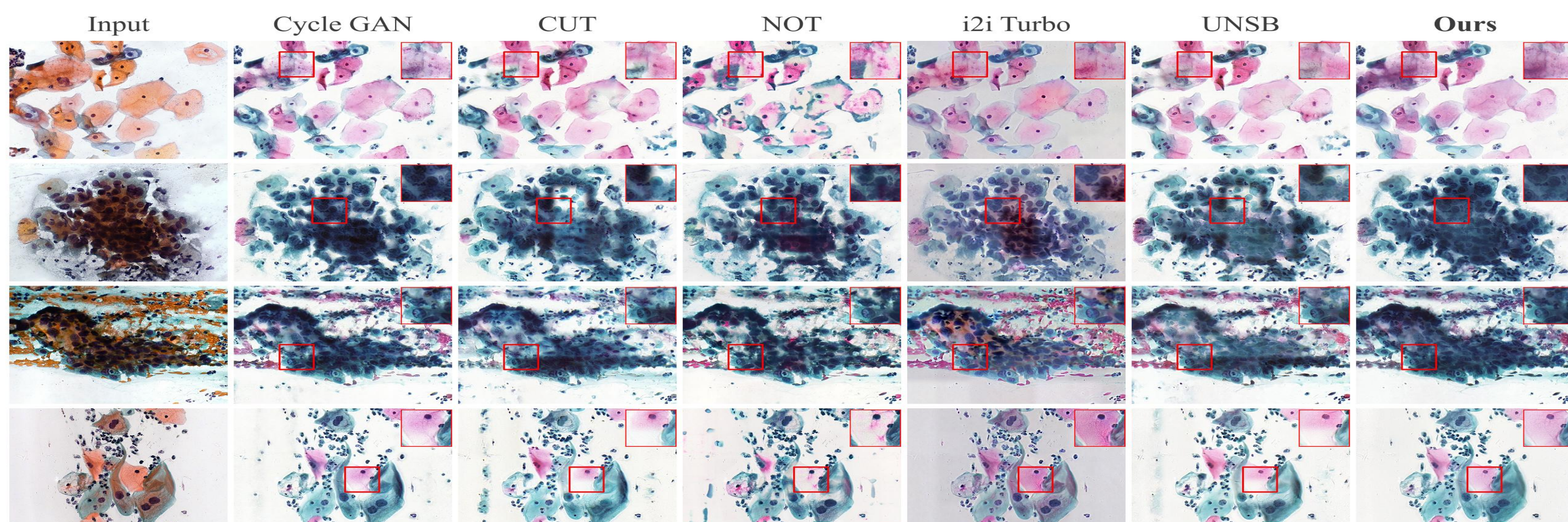
Introduction

Cross-institution cervical cell screening is challenged by variations in sample preparation and staining, which create substantial appearance shifts and obscure the subtle differences between lesion stages. Feature-level adaptation alone may not close this gap, while patch-wise translation of ultra-high-resolution images introduces discontinuous normalization statistics and boundary artifacts. We therefore propose a two-stage framework that first constructs a spatially coherent intermediate domain and then aligns structural and semantic representations through dual-level knowledge distillation.

Experiment Setting

We evaluate the proposed framework on two cervical cytology datasets. CRIC serves as the source domain $\mathcal{D}_s$ and contains 7,839 cervical microscopy bounding boxes, while ComparisonDetector serves as the target domain $\mathcal{D}_t$ and contains 7,410 clustered cell images. The target-domain test set is strictly excluded from training.

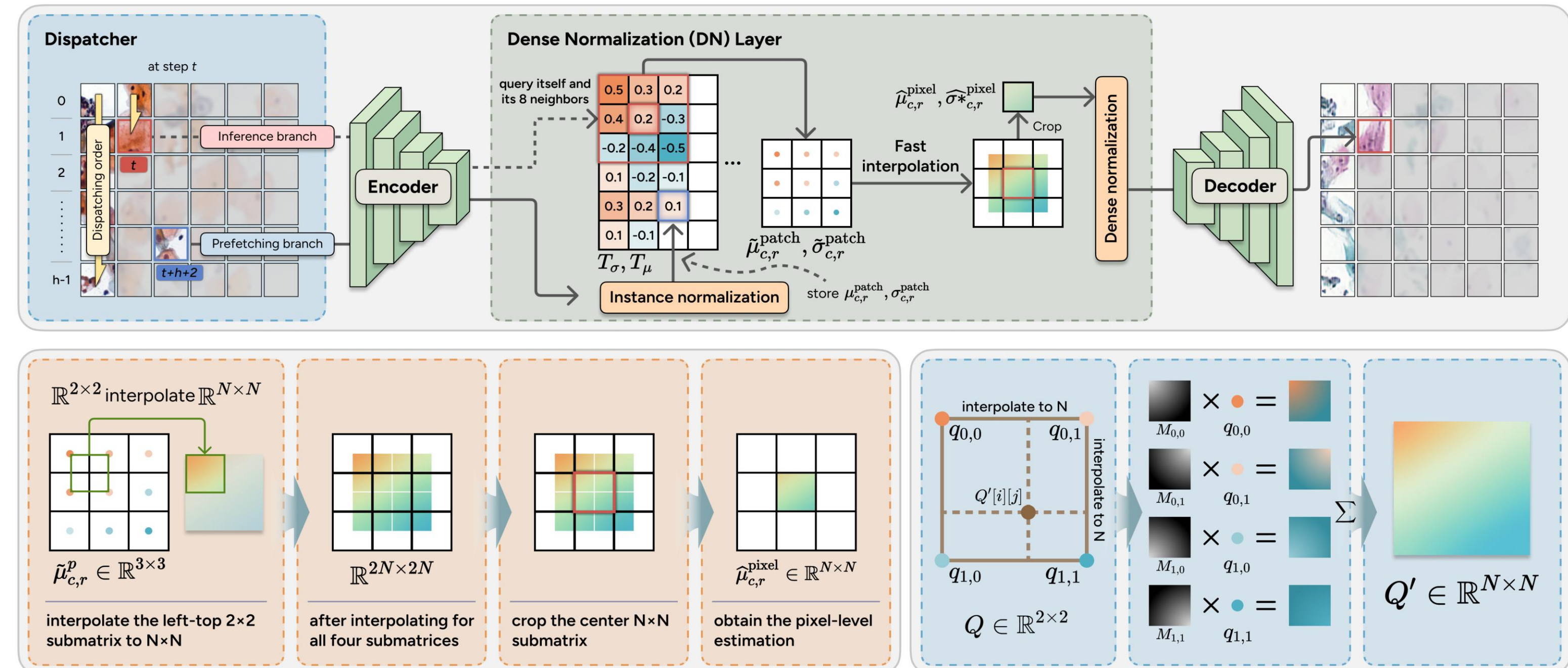
Models are optimized using SGD with a learning rate of 0.001 and momentum of 0.9. Pseudo-labels with confidence scores above 0.9 supervise the target model. Image translation is evaluated using FID, KID, NIQE, and HIST, while downstream detection is evaluated using mAP and mAP50.



Method

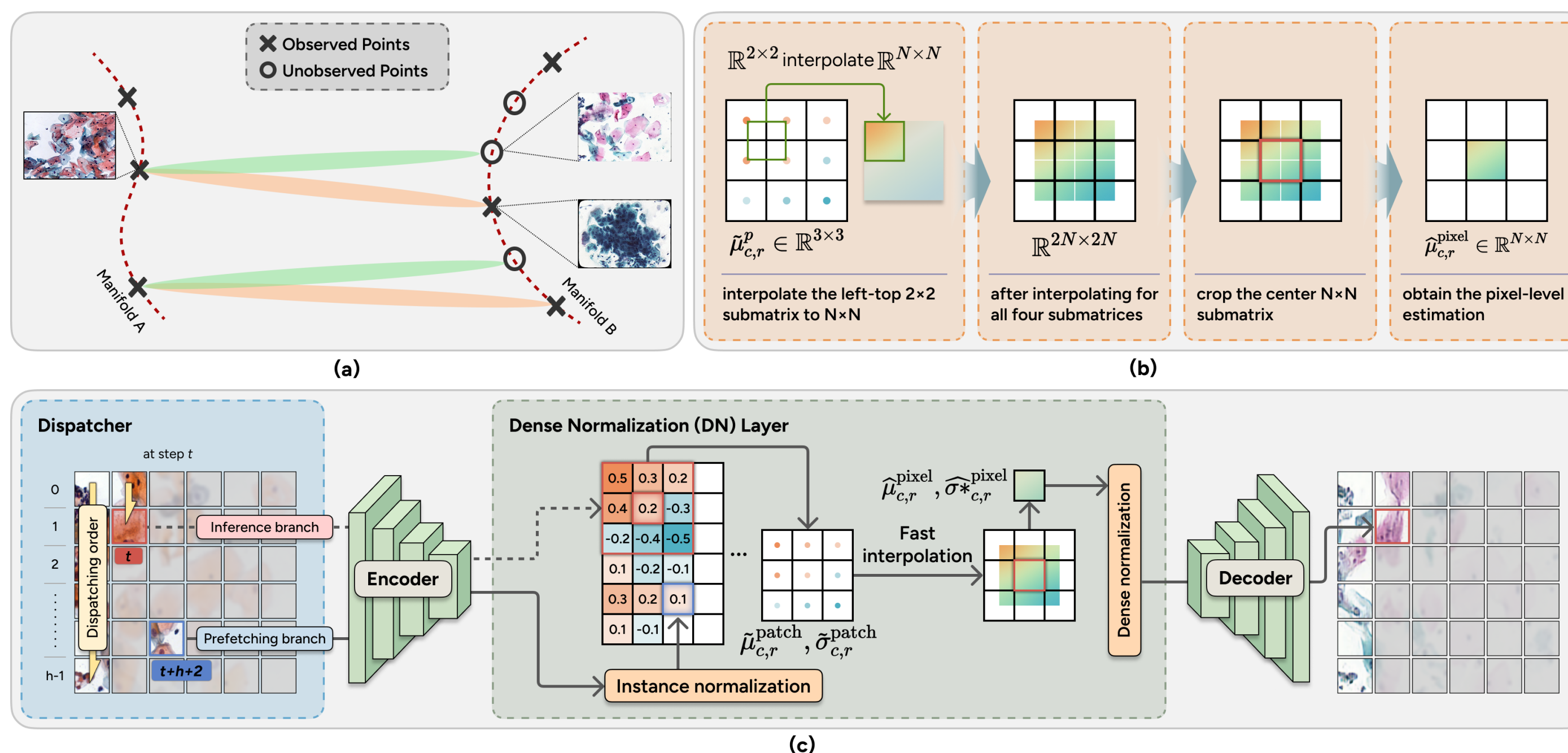
Stage I — Spatially Continuous Image Synthesis

SC-UNSB replaces independently estimated patch statistics with dense pixel-wise moment fields interpolated from neighboring patches. The resulting Dense Normalization layer maintains spatial continuity across patch boundaries, reducing tiling artifacts while preserving cellular morphology and staining consistency.



SC-UNSB constructs continuous pixel-wise statistics from neighboring patches for spatially coherent cytology image translation.

Stage II — Dual-Level Feature Alignment



LFA aligns shallow structural patterns, while CFA transfers compact high-level semantic representations.

Loose Feature Alignment transfers shallow, low-frequency structural patterns, while Compact Feature Alignment aligns high-level semantic representations in a shared embedding space. Together, the two modules progressively transfer domain-invariant knowledge from cellular structure to class-discriminative semantics.

Overall loss function:

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{loc} + \alpha(\mathcal{L}_{LFA} + \mathcal{L}_{CFA})$$

Main Results

SC-UNSB performs best across alignment settings. With LFA and CFA, it achieves 26.9% mAP and 45.8% mAP50, exceeding UNSB by 2.8 and 3.2 points. It achieves the best NIQE and HIST with competitive FID/KID, while results show smoother patch transitions and pixel-wise statistics. Pixel-level interpolation gives the best FID, KID, and HIST; distillation surpasses KD, DKD, and SPD, with $\alpha=0.2$ performing best.

Generator	$\mathcal{D}_s$	CycleGAN[18]	CUT[19]	NOT[20]	i2i-Turbo[21]	UNSB[14]	SC-UNSB
FID ↓	241.05	147.28	132.61	177.43	154.65	143.43	135.31
KIDx100 ↓	10.896	1.831	1.365	5.104	2.514	2.411	1.807
NIQE ↓	14.72	13.51	14.6	12.86	16.39	13.94	11.38
HIST ↑	0.384	0.695	0.722	0.571	0.473	0.701	0.754
Detector	RetinaNet						
mAP ↑	4.6%	10.5%	14.0%	9.3%	10.3%	17.8%	20.2%
mAP50 ↑	9.3%	22.4%	29.3%	18.0%	26.8%	35.4%	41.5%
Detector	RetinaNet + CFA						
mAP ↑	7.8%	12.1%	16.3%	9.6%	13.9%	21.2%	22.7%
mAP50 ↑	19.1%	23.7%	33.8%	20.4%	30.1%	38.7%	43.1%
Detector	RetinaNet + LFA						
mAP ↑	5.1%	12.3%	19.4%	12.1%	12.4%	23.8%	24.1%
mAP50 ↑	11.5%	24.0%	35.4%	18.4%	28.4%	41.1%	43.7%
Detector	RetinaNet + LFA + CFA						
mAP ↑	12.6%	18.8%	22.7%	11.3%	15.1%	24.1%	26.9%
mAP50 ↑	26.6%	31.8%	40.3%	21.9%	30.9%	42.6%	45.8%

Table 1 Overall Generation and Detection Performance

Interpolating Granularity	NIQE↓	SC-UNSB (testing set)			SC-UNSB (training set)		
		FID↓	KID↓ × 100	HIST↑	FID↓	KID↓ × 100	HIST↑
1 pixel	11.38	135.31	1.807	0.754	66.291	1.534	0.711
16 pixels	11.59	147.93	2.22	0.717	70.77	1.8	0.700
32 pixels	11.62	162.59	3.630	0.713	77.1	2.37	0.698
64 pixels	11.34	148.29	2.91	0.708	76.87	2.6	0.688
128 pixels	11.40	143.37	2.47	0.704	70.6	2.07	0.682
256 pixels	13.94	143.43	2.411	0.701	68.232	1.623	0.707

Table 2 Interpolation Granularity

Metric	Knowledge Distillation				Hyperparameter Analysis			
	KD[26]	DKD[27]	SPD[28]	Ours	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$
mAP ↑	21.7%	18.6%	22.3%	26.9%	25.3%	26.9%	26.1%	24.5%
mAP50 ↑	40.6%	36.7%	43.4%	45.8%	44.2%	45.8%	45.4%	45.1%

Table 3 Distillation and Hyperparameter Analysis