Bidirectional Translation Between ISH and H&E Glioblastoma Tissue Sections via Attention-Enhanced CycleGAN

Ayesha Shahzad[1], Konstantinos Koukoutegos[2,3], Dimitrios Makris[1], and Spyridon Bakas[1,2,3,4,5,†]

[1] Department of Computer Science, Kingston University London, London, UK
[2] Division of Computational Pathology, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine (IUSM), Indianapolis, IN 46202, USA
[3] Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indianapolis, IN 46202, USA
[4] Departments of Neurological Surgery; Biostatistics and Health Data Science; Radiology and Imaging Sciences, IUSM, Indianapolis, IN 46202, USA
[5] Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indiana University, Indianapolis, IN 46202, USA
†Corresponding Author: spbakas@iu.edu

Abstract. Bidirectional translation between in situ hybridization (ISH) and Hematoxylin and eosin (H&E) digitized tissue sections presents an unexplored challenge in computational pathology. ISH reveals the spatial distribution of specific nucleic-acid targets at cellular resolution, while H&E primarily captures tissue morphology. Obtaining both from the same section is rarely feasible, and serial sectioning consumes limited biopsy material, incurs cost, and, importantly, addresses adjacent rather than the same tissue. To this end, we propose an extended CycleGAN for unpaired bidirectional translation between ISH and H&E digitized tissue sections, while leveraging the IvyGAP dataset. We develop a model configuration incorporating Convolutional Block Attention Module (CBAM), corrected loss weightings, and auxiliary structural constraints. Systematic evaluation for CD44 across eleven training checkpoints demonstrates that validation proxy loss is a poor surrogate for distributional quality: the lowest-proxy-loss checkpoints coincide with elevated Fréchet Inception Distance (FID). On the held-out test set, our best checkpoints achieve FID 34.359 for ISH→H&E (epoch 20) and FID 29.577 for H&E→ISH (epoch 42), with cycle-reconstruction SSIM ranging from 0.968 to 0.976. Applied without retraining for IDH1 and EGFR, the model generalizes across glioblastoma molecular markers. A controlled ablation study shows CBAM alone improves performance in both directions, though the full model's combined benefit over a standard CycleGAN baseline is direction-dependent, favoring H&E→ISH. This is, to our knowledge, the first published study of bidirectional translation of ISH ↔ H&E in glioblastoma.

Keywords: Virtual staining · CycleGAN · Glioblastoma · ISH · H&E · CBAM · IvyGAP · Checkpoint selection

1 Introduction

Glioblastoma (GBM) is the most prevalent and lethal primary brain tumor in adults, with a median overall survival of 12–18 months despite maximal surgical resection and temozolomide chemo-radiotherapy [1], and with profound spatial heterogeneity at both morphological and molecular levels. Understanding this heterogeneity requires complementary diagnostic modalities. Hematoxylin and eosin (H&E) staining reveals tissue architecture and nuclear morphology, while in situ hybridization (ISH) maps the subcellular spatial distribution of specific nucleic acid targets, providing direct evidence of gene expression at cellular resolution.

In practice, obtaining both stains from the same tissue region is rarely achievable. Serial sectioning yields adjacent rather than identical sections, typically separated by 10–20 μm, meaning that ISH and H&E images never correspond at the cellular level. Furthermore, each additional stain consumes irreplaceable biopsy material, extends laboratory turnaround, and adds procedural cost and extra time. Virtual staining addresses all these constraints simultaneously: by computationally translating between the two modalities, it enables direct comparison of morphology and gene expression within exactly the same tissue region, which no physical protocol can match, without further material consumption or time delay.

Virtual staining via generative models has been studied across several tissue types. De Haan et al. [6] used paired U-Net supervision for H&E-to-special-stain transfer. Diffusion models later achieved super-resolved virtual staining [20], while Yoshimura et al. [10] applied CycleGAN [3] to highly aligned synthetic MRI, unlike the weakly aligned setting tackled by Chen et al. [11]. Despite this breadth, no prior study has addressed bidirectional ISH $\leftrightarrow$ H&E translation in any tissue type, including GBM.

The Ivy Glioblastoma Atlas Project (IvyGAP) [2] provides paired H&E and ISH digitized tissue sections (also known as whole-slide images - WSIs) used in this work (Sect. 2.1). The inter-section gap precludes pixel-level supervision, necessitating an unpaired image translation approach and ruling out paired methods such as pix2pix [14]. CycleGAN [3] is well-suited for this setting and has demonstrated effectiveness in histopathological stain transfer [4], but the standard architecture presents challenges for translating ISH images, as the sparse, punctate chromogenic ISH signal occupies a small fraction of any given patch and benefits from explicit spatial attention. Also a poorly configured identity loss weight can suppress the color-space change that staining translation requires (Sect. 2.3).

In this article, we make five contributions. **(i)** We present CBAM-CycleGAN, an attention-enhanced CycleGAN incorporating a Convolutional Block Attention Module (CBAM) [5] in the generator bottleneck, with corrected loss weightings and auxiliary structural constraints. **(ii)** We conduct a rigorous multi-checkpoint evaluation across eleven saved states and expose a critical dissociation between validation proxy loss and distributional quality, with direct implications for GAN training practice. **(iii)** We report the first bidirectional translation ISH $\leftrightarrow$ H&E

results on IvyGAP, establishing an initial benchmark for this task. **(iv)** We demonstrate cross-gene generalization: a model trained exclusively on CD44 produces competitive results on IvyGAP IDH1 and EGFR subsets without retraining. **(v)** We conduct a controlled ablation isolating CBAM and the loss reweighting under matched conditions, showing their contribution is direction-dependent rather than uniform.

2 Methods

2.1 Data

IvyGAP [2] comprises 100 matched CD44 ISH and H&E-stained WSIs from 42 GBM donors, each pair obtained from adjacent sections separated by $\sim 20\,\mu$m. Slides measure approximately $15{,}040 \times 18{,}080$ pixels at $0.49\,\mu$m/pixel. WSIs were tiled into non-overlapping 512×512 patches and those with $<70\%$ tissue-covered pixels (per IvyGAP anatomical annotations) were discarded. Data were partitioned at donor level, 31/5/6 for training/validation/testing (57,935/12,378/13,848 patches per partition), normalized to $[-1, 1]$, and random horizontal/vertical flipping was applied during training.

2.2 Proposed Framework

Figure 1 summarizes our proposed bidirectional translation/virtual staining framework. Generator G_{AB} maps ISH images to translated H&E, with discriminator D_B evaluating the realism of the resulting H&E output. Conversely, generator G_{BA} maps H&E images to translated ISH, evaluated by discriminator D_A. Cycle-consistency is enforced by passing each translated image back through the opposing generator to reconstruct the original input (e.g., ISH→H&E→ISH).

The generator follows the 9-block ResNet design of Zhu et al. [3]: a reflection-padded 7×7 input convolution; two strided downsampling convolutions with instance normalization [13] and ReLU activation; nine residual blocks; two transposed convolutional upsampling layers; and a 7×7 tanh output convolution. A CBAM [5] is inserted after every third residual block in the bottleneck (blocks 3, 6, 9), yielding three attention stages at 256-channel, 128×128 feature resolution. Each CBAM applies sequential channel attention (average/max-pool, shared two-layer MLP, reduction ratio 16) and spatial attention (7×7 convolution on channel-averaged and channel-maximum maps), directing the generator to selectively process the spatially sparse ISH chromogenic deposits rather than treating all positions uniformly. Both discriminators use the 70×70 PatchGAN architecture [14], enforcing local texture and color consistency appropriate for staining modality differences.

2.3 Loss Functions

The generator objective combines five complementary loss terms. Let G_{AB} and G_{BA} denote the generators mapping ISH→H&E and H&E→ISH, respectively;

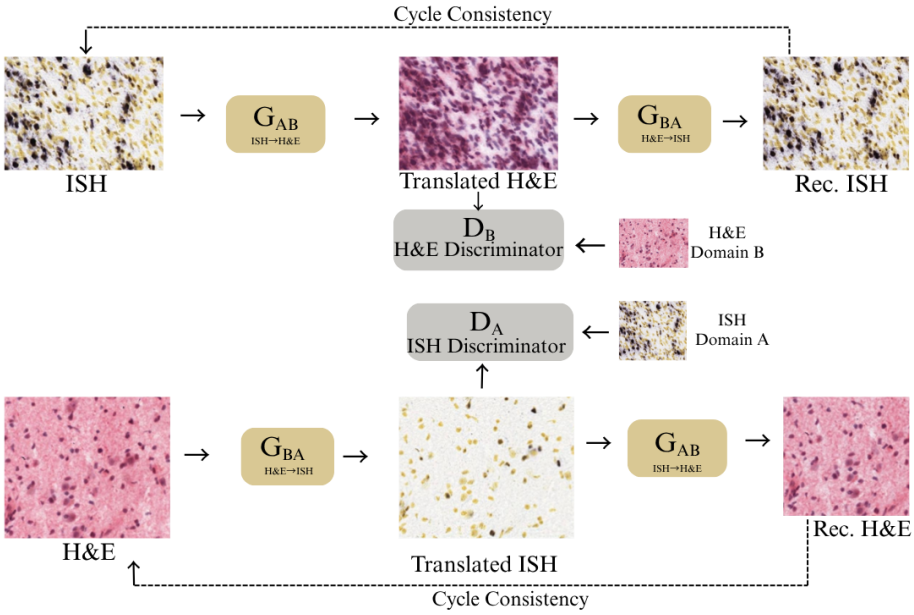

Fig. 1. Bidirectional translation framework. Generator G_{AB} translates ISH→H&E; generator G_{BA} translates H&E→ISH. Cycle-consistency reconstructs each input via the opposing generator.

D_A and D_B denote the corresponding discriminators; and x, y the real ISH and H&E images.

(i) Adversarial LSGAN [15]. Encourages generated images to match the target-domain distribution:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}\Big[\big(D_B(G_{AB}(x)) - 1\big)^2\Big] + \mathbb{E}\Big[\big(D_A(G_{BA}(y)) - 1\big)^2\Big] .$$

(ii) Cycle ℓ_1 ($\lambda_{\text{cyc}} = 10$). Preserves image content through round-trip reconstruction:

$$\mathcal{L}_{\text{cyc}} = \mathbb{E}[\|G_{BA}(G_{AB}(x)) - x\|_1] + \mathbb{E}[\|G_{AB}(G_{BA}(y)) - y\|_1] .$$

(iii) Identity ($\lambda_{\text{id}} = 0.5$). Discourages unnecessary modifications of target-domain inputs:

$$\mathcal{L}_{\text{id}} = \mathbb{E}[\|G_{AB}(y) - y\|_1] + \mathbb{E}[\|G_{BA}(x) - x\|_1] .$$

(iv) Cycle SSIM ($\lambda_{\text{ssim}} = 2.0$). Promotes preservation of structural tissue information:

$$\mathcal{L}_{\text{ssim}} = \mathbb{E}[1 - \text{SSIM}(G_{BA}(G_{AB}(x)), x)] + \mathbb{E}[1 - \text{SSIM}(G_{AB}(G_{BA}(y)), y)] .$$

(v) Cycle Sobel edge ($\lambda_{\text{edge}} = 1.0$). Preserves fine morphological boundaries and edge structure:

$$\mathcal{L}_{\text{edge}} = \mathbb{E}[\|\nabla G_{BA}(G_{AB}(x)) - \nabla x\|_1] + \mathbb{E}[\|\nabla G_{AB}(G_{BA}(y)) - \nabla y\|_1] \,.$$

where ∇ denotes the Sobel magnitude. SSIM and Sobel constraints are applied only to same-domain cycle reconstructions and are therefore unaffected by intersection misalignment.

The identity weight $\lambda_{\text{id}} = 0.5$ and cycle weight $\lambda_{\text{cyc}} = 10$ follow the original specification [3]; λ_{id} values substantially above 0.5 suppress the color-space change staining translation requires. The corrected weighting improves structural self-consistency and H&E→ISH fidelity over a matched $\lambda_{\text{id}} = 5.0$ baseline, with the full comparison reported in Sect. 3.5. The auxiliary $\lambda_{\text{ssim}} = 2.0$ and $\lambda_{\text{edge}} = 1.0$ were tuned via limited manual search to balance structural preservation against the adversarial and cycle terms. Discriminator training uses one-sided label smoothing [16] (real labels 0.9, fake labels 0.0) and a 50-image experience replay buffer per domain [17]; VGG perceptual loss was tested but caused training collapse, consistent with the stain-transfer literature [4], and was excluded. The generators' learning rate is $\text{lr}_G = 2 \times 10^{-4}$, while half that rate ($\text{lr}_D = 1 \times 10^{-4}$) was used for the discriminators, preventing them from converging faster than the generators and rendering their gradient signal uninformative.

2.4 Experimental Progression

Model development began with a standard CycleGAN baseline (256×256, 6 residual blocks, λ_{id}=5.0) trained on an earlier, smaller pilot subset, which yielded preliminary FID scores of 60.25 (ISH→H&E) and 91.50 (H&E→ISH); the resolution- and dataset-matched ablation study in Sect. 3.5 supersedes this as a controlled comparison. Subsequent refinements increased the resolution to 512×512, expanded the generators to 9 residual blocks, reduced λ_{id} to 0.5, and introduced CBAM together with the auxiliary SSIM and Sobel losses described in Sect. 2.3. Each component's individual contribution is isolated in our ablations (Sect. 3.5). The resulting configuration was used for all experiments, which were derived from a single training run and seed, without FID confidence intervals across multiple runs.

2.5 Evaluation Metrics

Because adjacent sections do not correspond at the cellular level (Sect. 2.1), cross-domain pixel-level metrics (e.g., SSIM, PSNR between real ISH and translated H&E) are therefore invalid and excluded. The primary metric is Fréchet Inception Distance (FID) [18], which measures distributional similarity via InceptionV3 feature statistics, requiring no pixel alignment. Kernel Inception Distance (KID) [19] provides an unbiased complement. Both are computed on the full held-out validation and test patch cohorts (12,378 and 13,848 patches respectively, partitioned across both domains; Sect. 2.1) rather than a small subsample, mitigating known

FID instability at small sample sizes. Cycle structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR), computed on same-domain cycle-reconstructions, are self-consistency measures unaffected by inter-section misalignment. They substitute for the paired ground truth unavailable here, following established convention in unpaired translation [21]. Distribution-matching losses can cause generated medical images to include unsupported features [22], motivating both distributional and self-consistency evaluation here. val_G denotes the generator proxy loss (adversarial, cycle ℓ_1, and identity terms) evaluated on 500 validation patches. Eleven checkpoints (epochs 19, 20, 23, 26, 27, 29, 32, 34, 37, 40, and 42), which were retained by the training checkpointing policy rather than an evenly spaced schedule, were evaluated comprehensively on the validation set. Top-ranked validation checkpoints were carried to the test set for confirmation; where validation results were inconclusive, all competing checkpoints were evaluated once to determine the final selection. Excluded checkpoints were not revisited.

3 Results

3.1 Multi-Checkpoint Evaluation (CD44 Validation Set)

Table 1 presents full validation results across all eleven checkpoints, revealing a marked dissociation between proxy validation loss (val_G) and distributional quality. Epoch 26 achieved the lowest val_G of the entire run (1.072) alongside substantially elevated FID ISH→H&E (57.984). Epoch 37, despite an unremarkable, mid-range val_G (1.314), shows comparably elevated FID (56.068). Both epochs satisfied all training objectives through small, reversible transformations rather than genuine staining translation, which FID captures but val_G does not, strongly motivating periodic distributional evaluation (FID or KID) over proxy loss alone for checkpoint selection.

Epoch 20 and epoch 19 emerge as the leading ISH→H&E candidates, with the lowest and second-lowest validation FID respectively. Both were carried forward to verify the stability of this ranking. The H&E→ISH direction tells a different story: epoch 27 minimizes validation FID, but epoch 42 maximizes cycle SSIM-H&E, so neither checkpoint dominates outright. Reconciling distributional realism against structural self-consistency therefore requires both candidates to be carried forward to the test set. Such asymmetric convergence can occur in CycleGAN, whose two generators rarely reach their respective optima at the same epoch.

3.2 Test Set Evaluation (CD44)

On the test set (Table 2), epoch 20 is retained as the ISH→H&E checkpoint: FID and KID are comparable to epoch 19, while cycle SSIM and PSNR are higher, giving epoch 20 the stronger overall profile. The H&E→ISH picture resolves more cleanly: epoch 27's validation lead does not transfer, and epoch 42

Table 1. Validation set results across all eleven checkpoints (CD44). Best per column **in bold**. Shaded rows: proxy-FID dissociation (lowest *val_ G*, elevated FID).

Epoch	FID↓ ISH→HE	FID↓ HE→ISH	KID↓ ISH→HE	KID↓ HE→ISH	SSIM ISH↑	SSIM HE↑	PSNR ISH↑	PSNR HE↑	*val_ G*
19	31.840	36.630	0.020	0.022	0.969	0.970	35.230	35.555	1.218
20	**30.131**	43.308	**0.015**	0.024	0.971	0.976	34.990	34.666	1.363
23	35.969	33.576	0.024	0.017	0.970	0.978	36.315	34.747	1.130
26	57.984	43.448	0.043	0.024	0.971	0.977	36.238	35.102	**1.072**
27	36.492	**32.383**	0.022	**0.013**	0.971	0.977	34.100	**35.601**	1.314
29	38.263	33.947	0.024	0.016	0.972	0.977	**36.722**	33.947	1.294
32	33.919	38.945	0.021	0.023	**0.974**	0.978	35.767	34.990	1.352
34	36.015	39.018	0.023	0.025	0.971	0.979	35.762	31.843	1.191
37	56.068	34.885	0.045	0.017	0.973	0.979	36.037	35.206	1.314
40	33.771	52.517	0.018	0.046	0.971	0.979	35.941	35.239	1.492
42	33.803	33.507	0.020	0.016	0.972	**0.981**	36.147	35.357	1.553

Table 2. Test set results for selected checkpoints (CD44). Best value(s) per column **in bold**.

Epoch	FID↓ ISH→HE	FID↓ HE→ISH	KID↓ ISH→HE	KID↓ HE→ISH	SSIM ISH↑	SSIM HE↑	PSNR ISH↑	PSNR HE↑
19	**33.776**	48.966	**0.019**	0.041	0.969	0.968	34.767	29.842
20	34.359	48.746	0.020	0.037	**0.974**	0.973	35.345	32.187
27	41.945	32.084	0.028	0.020	**0.974**	0.975	34.747	**32.424**
42	56.054	**29.577**	0.043	**0.017**	**0.974**	**0.976**	**35.959**	28.781

instead delivers the best test FID and SSIM-H&E together, settling the direction in its favor. Neither specialist generalizes bidirectionally; for single-checkpoint deployment, epoch 27 offers the lowest mean FID across both directions (37.015) as an alternative. Cycle SSIM stays uniformly high across all four checkpoints, indicating that structural self-consistency holds regardless of which checkpoint best matches the target distribution. Figs. 2 and 3 show representative outputs for the two deployed checkpoints (epochs 20 and 42), illustrating that the translated images preserve nuclear morphology and staining gradients in H&E and reproduce the sparse, punctate chromogenic pattern characteristic of ISH.

3.3 Multi-Gene Generalization

To assess generalization, the CD44-trained model was applied without retraining to IvyGAP ISH and H&E images from the IDH1 and EGFR subsets (Table 3). ISH→H&E generalizes far more readily than H&E→ISH. Relative to the CD44 baseline, FID rises by 37-84% for ISH→H&E (epoch 20, EGFR vs. IDH1) but 107-259% for H&E→ISH (epoch 42), consistent with the more diffuse, lower-density chromogenic pattern observed for IDH1. Cycle SSIM nonetheless remains at or above 0.945 across all genes and directions, confirming structural self-consistency under domain shift even as distributional fidelity (FID) degrades. The CD44 ISH→H&E FID compares favorably with, and the EGFR FID is competitive with, published H&E-to-IHC results (Table 4), though such cross-task comparisons carry the caveats noted in Sect. 3.4.

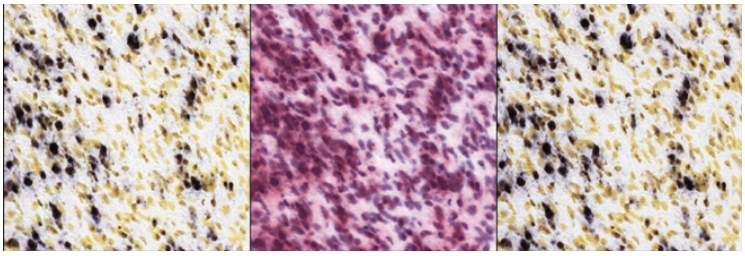

Fig. 2. ISH→H&E translation (epoch 20). Real ISH (left), translated H&E (center), reconstructed ISH (right). Nuclear morphology and staining gradients are reproduced.

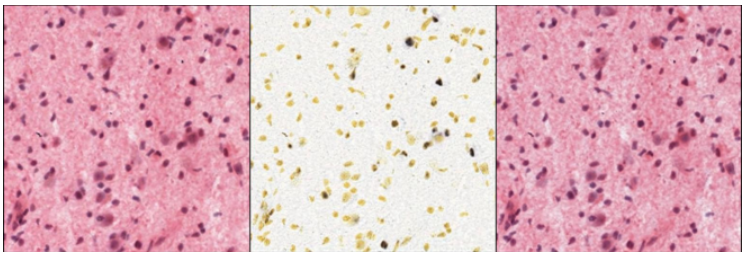

Fig. 3. H&E→ISH translation (epoch 42). Real H&E (left), translated ISH (center), reconstructed H&E (right). The sparse punctate chromogenic pattern is reproduced.

Table 3. Generalization of the CD44-trained model (epochs 20 and 42) to EGFR and IDH1, without retraining. For each gene, the better-performing epoch is shown in **bold**.

Gene	Epoch	FID↓ ISH→H&E	FID↓ H&E→ISH	SSIM ISH↑	SSIM H&E↑
CD44	20	**34.359**	48.746	**0.974**	0.973
CD44	42	56.054	**29.577**	**0.974**	**0.976**
EGFR	20	**47.039**	66.198	0.954	0.963
EGFR	42	50.030	61.130	0.958	0.973
IDH1	20	**63.130**	117.620	0.959	0.945
IDH1	42	66.470	106.230	0.948	0.969

3.4 Contextual Comparison

No prior work has reported ISH ↔ H&E translation. Our comparisons (Table 4) situate the CD44 results within existing literature, but values are not strictly comparable owing to differences in tissue, dataset, and metric implementation.

3.5 Ablation Study

To evaluate the contributions of CBAM and of the corrected loss weighting (Sect. 2.3), we trained two additional configurations matched to the full model's 512 × 512, 9-block architecture, and schedule. Specifically, we quantitatively evaluate i) a standard CycleGAN (λ_{id}=5.0, no auxiliary losses) (indicated as **V**),

Table 4. Contextual comparison with published virtual staining and medical image translation studies. *Foreground FID only. SSIM/PSNR for *Ours* and DeCGAN are cycle-reconstruction metrics, i.e., the only valid same-domain measure given the unpaired dataset, not directly comparable to the paired cross-domain fidelity reported elsewhere.

Study	Task	FID↓	SSIM↑	PSNR↑	Supervision
Ours	ISH→H&E	34.359	0.974	35.345	Unpaired
Ours	H&E→ISH	29.577	0.976	28.781	Unpaired
Pyramid Pix2pix [7]	H&E→IHC	—	0.48	21.16	Paired
PPT [8]	H&E→IHC	49–81	0.35–0.50	17–19	Paired
Confusion-GAN [9]	H&E→IHC	39–77	0.21–0.46	—	Weakly sup.
BBDM [20]	Label-free→H&E	—	0.63–0.85	20–28	Paired
ContourDiff [11]	CT→MRI	126–134*	—	—	Unpaired
CycleGAN-MRI [10]	T1w↔T2w MRI	—	0.993–0.995	40.16–40.51	Unpaired
DeCGAN [21]	CT↔MRI	67–70	0.94–0.98	27–32	Unpaired

Table 5. Ablation study (test set, CD44), 512 × 512, 9-block ResNet throughout. Best checkpoint per direction reported (epoch shown as ISH→H&E/H&E→ISH). V: standard CycleGAN; N: no-CBAM; F: full model (**Ours**). Best per column **in bold**.

Config	CBAM	Aux. losses	Epoch	FID↓ ISH→HE	FID↓ HE→ISH	KID↓ ISH→HE	KID↓ HE→ISH	SSIM↑ ISH	SSIM↑ HE
V: Standard	×	×	23/8	**32.158**	46.569	**0.014**	0.029	0.959	0.944
N: no-CBAM	×	✓	10/20	36.010	36.908	0.018	0.025	0.960	0.974
F: Full (**Ours**)	✓	✓	20/42	34.359	**29.577**	0.020	**0.017**	**0.974**	**0.976**

ii) the full loss configuration CycleGAN but without CBAM (**N**), and iii) our full proposed model (**F**), yielding a three-way ablation study. Table 5 reports test-set results for the best checkpoint per direction.

For H&E→ISH, both components help cumulatively (FID 46.569→36.908 →29.577, -36.5% relative), with cycle SSIM-H&E improving monotonically alongside it (0.944→0.974→0.976). For ISH→H&E, neither component improves FID over the matched standard-CycleGAN baseline (32.158 vs. 36.010/34.359), though cycle SSIM-ISH still improves monotonically (0.959→0.960→0.974). We attribute the FID pattern to the SSIM/Sobel terms trading distributional diversity for structural fidelity on the H&E side, a cost CBAM's ISH-targeted spatial attention does not incur in the reverse direction; the consistent SSIM gains indicate this trade-off does not compromise structural self-consistency. Isolating CBAM alone (N versus F, loss configuration held fixed) shows that adding CBAM improves FID in both directions, 4.6% for ISH→H&E and 19.9% for H&E→ISH, indicating attention helps translate the sparse ISH signal considerably more than it helps read it as input. Representative qualitative outputs across all three configurations are provided in the Supplementary Material (Fig. 4 and Fig. 5).

4 Discussion and Conclusion

In this study, we introduced an extended CycleGAN architecture with CBAM and a loss reweighting under matched conditions for bidirectional virtual staining

between unpaired ISH and H&E WSIs. Quantitative performance evaluation and ablation analysis using the IvyGAP dataset justify the choice of architecture extensions. Trained exclusively on target CD44 and applied without retraining to IDH1 and EGFR, our model demonstrates cross-gene generalization.

The proxy-loss dissociation (epochs 26 and 37) is practically substantial as CycleGAN can satisfy all training objectives, whether at a near-minimal proxy loss or a merely unremarkable one, without achieving genuine staining translation. This degenerate optimum is consistent with theoretical results showing that CycleGAN's loss admits multiple exact solutions, including trivial automorphisms of the target domain [23], and is identifiable here only through distributional evaluation. This indicates that periodic FID/KID computation is a critical requirement alongside proxy-loss monitoring, though these metrics, computed via ImageNet-pretrained InceptionV3 features, quantify distributional plausibility rather than actual biological relevance, a limitation shared across virtual-staining evaluation generally. Our ablation analysis (Sect. 3.5) further justifies CBAM's choice and shows that its contribution is direction-dependent: positive in both directions, but substantially larger for H&E→ISH than for ISH→H&E, indicating its spatial attention helps most when translating the sparse ISH signal rather than merely consuming it as input. The consistently high cycle SSIM values confirm the learned translations are structurally self-consistent, an important property for clinical utility. Additionally, cross-gene generalization suggests that features learned on CD44 ISH WSIs, characterizing the dense, focal chromogenic deposits of CD44 (a transmembrane glycoprotein linked to glioblastoma invasion and poorer prognosis) [12], transfer partially to other genes. EGFR, also relatively focal, generalizes more than IDH1, whose more diffuse, lower-density expression may contribute to its weaker H&E→ISH transfer (Table 3); whether this degree of degradation is acceptable for clinical or research use remains to be established.

Our proposed framework demonstrates the feasibility of bidirectional virtual staining between ISH and H&E, while exhibiting promising cross-gene transfer without retraining. Future work should investigate joint multi-gene training and IDH1-specific fine-tuning to improve low-density marker transfer, characterize how the tissue-coverage filter (Sect. 2.1) may bias representation toward denser tumor regions or underrepresent microenvironments such as the invasive margin and perivascular or necrotic zones, and pursue pathologist or gene-specific biological validation of translated images.

Acknowledgments

This study uses the publicly available, de-identified IvyGAP dataset [2], released under an open license. No additional ethical approval was required.

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Stupp, R., et al.: Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N. Engl. J. Med. **352**(10), 987-996 (2005). https://doi.org/10.1056/NEJMoa043330
2. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., et al.: An anatomic transcriptional atlas of human glioblastoma. Science **360**(6389), 660-663 (2018). https://doi.org/10.1126/science.aaf2666
3. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV, pp. 2223-2232 (2017). https://doi.org/10.1109/ICCV.2017.244
4. Shmatko, A., Ghaffari Laleh, N., Gerstung, M., Kather, J.N.: Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nat. Cancer **3**(9), 1026-1038 (2022). https://doi.org/10.1038/s43018-022-00436-4
5. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: ECCV. LNCS, vol. 11211, pp. 3-19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
6. De Haan, K., et al.: Deep learning-based transformation of H&E stained tissues into special stains. Nat. Commun. **12**(1), 4884 (2021). https://doi.org/10.1038/s41467-021-25221-2
7. Liu, S., et al.: BCI: Breast cancer immunohistochemical image generation through pyramid Pix2pix. In: CVPR Workshops, pp. 1815-1824 (2022). https://doi.org/10.1109/CVPRW56347.2022.00198
8. Zhang, W., et al.: High-resolution medical image translation via patch-alignment based bidirectional contrastive learning. In: MICCAI 2024. LNCS, vol. 15004, pp. 178–188. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72083-3_17
9. Li, J., Dong, J., Huang, S., Li, X., Jiang, J., Fan, X., Zhang, Y.: Virtual immunohistochemistry staining for histological images assisted by weakly supervised learning. In: CVPR, pp. 11259-11268 (2024). https://doi.org/10.1109/CVPR52733.2024.01070
10. Yoshimura, H., Kawahara, D., Saito, A., Ozawa, S., Nagata, Y.: Prediction of prognosis in glioblastoma with radiomics features extracted by synthetic MRI images using cycle-consistent GAN. Phys. Eng. Sci. Med. **47**(3), 1227-1243 (2024). https://doi.org/10.1007/s13246-024-01443-8
11. Chen, Y., Konz, N., Gu, H., Dong, H., Chen, Y., Li, L., Lee, J., Mazurowski, M.A.: ContourDiff: Unpaired medical image translation with structural consistency. Mach. Learn. Biomed. Imaging **3**, 711-727 (2025). https://doi.org/10.59275/j.melba.2025-79a2
12. Inoue, A., Ohnishi, T., Nishikawa, M., Ohtsuka, Y., Kusakabe, K., Yano, H., Tanaka, J., Kunieda, T.: A narrative review on CD44's role in glioblastoma invasion, proliferation, and tumor recurrence. Cancers **15**(19), 4898 (2023). https://doi.org/10.3390/cancers15194898
13. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. CoRR abs/1607.08022 (2016). https://doi.org/10.48550/arXiv.1607.08022
14. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR, pp. 5967-5976 (2017). https://doi.org/10.1109/CVPR.2017.632

15. Mao, X., et al.: Least squares generative adversarial networks. In: ICCV, pp. 2813-2821 (2017). https://doi.org/10.1109/ICCV.2017.304
16. Salimans, T., et al.: Improved techniques for training GANs. In: Advances in Neural Information Processing Systems (NIPS), pp. 2234-2242 (2016). https://doi.org/10.48550/arXiv.1606.03498
17. Shrivastava, A., et al.: Learning from simulated and unsupervised images through adversarial training. In: CVPR, pp. 2107-2116 (2017). https://doi.org/10.1109/CVPR.2017.241
18. Heusel, M., et al.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Advances in Neural Information Processing Systems (NIPS), pp. 6626-6637 (2017). https://doi.org/10.48550/arXiv.1706.08500
19. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying MMD GANs. In: International Conference on Learning Representations (ICLR) (2018). https://doi.org/10.48550/arXiv.1801.01401
20. Zhang, Y., Huang, L., Pillar, N., Li, Y., Chen, H., Ozcan, A.: Pixel super-resolved virtual staining of label-free tissue using diffusion models. Nat. Commun. **16**(1), 5016 (2025). https://doi.org/10.1038/s41467-025-60387-z
21. Cobbinah, M., Nunoo-Mensah, H., Acheampong, F.A., Acquah, I., Tchao, E.T., Agbemenu, A.S., Adinkrah, J., Kponyo, J.J., Abaidoo, E., Asamoah-Ansah, I., Otoo, O.K.: DeCGAN: a diversity-enhanced CycleGAN for unpaired image-to-image translation. Sci. Rep. (2026). https://doi.org/10.1038/s41598-026-55042-6
22. Cohen, J.P., Luck, M., Honari, S.: Distribution matching losses can hallucinate features in medical image translation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. LNCS, vol. 11070, pp. 529-536. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_60
23. Moriakov, N., Adler, J., Teuwen, J.: Kernel of CycleGAN as a principle homogeneous space. In: International Conference on Learning Representations (ICLR) (2020). https://doi.org/10.48550/arXiv.2001.09061