

ContiStain: Cross-Domain Relation-Preserving Distillation for Continual Multi-Domain Virtual IHC Staining

Fuqiang Chen¹, Yifeng Wang², Hongpeng Wang¹, and Yongbing Zhang¹

¹ School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China

ybzhang08@hit.edu.cn

² School of Computer Science and Technology, Tsinghua University, Beijing, China

Abstract. A unified multiplex virtual staining model enables scalable and non-destructive multiplex analysis from H&E slides while promoting parameter efficiency, shared pathological knowledge, and consistent cross-biomarker representations. However, in clinical practice, data for new biomarkers are typically acquired sequentially over time. Fine-tuning on such temporally arriving data leads to severe performance degradation on previously learned biomarkers, as sequential optimization disrupts the structured relationships among biomarker representations in the latent space. To address this issue, we propose ContiStain, an IHC multi-domain relational distillation framework for continual virtual staining. We first (i) construct a domain-aware structured feature space using a mixture-of-experts (MoE) feature extractor to reduce representation interference across biomarker domains. Based on this stabilized feature space, we then (ii) propose a relation-preserving distillation strategy that explicitly enforces the consistency of cross-domain token-level cosine similarity matrices between learned biomarker domains during continual adaptation. By maintaining cross-domain structural coherence, ContiStain mitigates forgetting while retaining adaptability to new domains. Experiments on the MIST dataset under a four-domain sequential virtual IHC staining setting show improved stability, reducing FID and ConchFID by 11.1 and 60.9 compared to sequential fine-tuning, enabling scalable and robust multi-domain virtual staining. Code is released at <https://github.com/ccitachi/ContiStain>.

Keywords: Virtual immunohistochemistry · Continual learning · Relational distillation.

1 Introduction

With the growing clinical demand for efficient pathological analysis, virtual immunohistochemistry (IHC) staining has emerged as a practical approach to generate biomarker-specific images directly from routine H&E slides [9]. Beyond single-biomarker translation [3, 15, 6], developing a unified multiplex virtual staining model [12, 21, 13, 2, 20, 5] capable of generating multiple biomarker domains

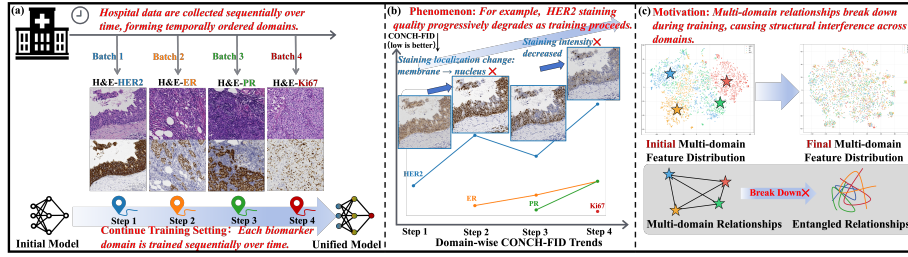


Fig. 1. Challenges of multi-domain continual virtual staining. (a) Sequential arrival of multiple biomarkers. (b) Naïve fine-tuning causes catastrophic forgetting (CONCH-FID increases). (c) Cross-domain relationships gradually collapse, motivating relationship-aware learning.

within a shared framework is highly desirable for parameter efficiency, shared pathological knowledge, and consistent cross-biomarker representation. In real-world pathology practice, however, new biomarkers are often adopted gradually based on evolving diagnostic needs and treatment decisions, rather than being introduced simultaneously. This naturally leads to a continual multi-domain virtual staining setting in Fig. 1 (a), where a unified model must incrementally adapt to new biomarker domains while preserving performance on previously learned ones.

In continual learning, sequential training often leads to the degradation of previously learned domains—in the case of virtual IHC staining, the generation quality of earlier biomarkers collapses in Fig. 1 (b). Therefore, a number of methods have been proposed, primarily in the context of classification tasks. Replay-based approaches such as iCaRL [17] mitigate forgetting by storing and revisiting a subset of samples from previous tasks. Distillation-based methods, including Learning without Forgetting (LwF) [11], encourage consistency between the outputs of the old and updated models to preserve prior knowledge. Regularization-based strategies such as Elastic Weight Consolidation (EWC) [8] constrain important parameters to remain close to their previously learned values. These approaches have shown effectiveness in alleviating catastrophic forgetting in sequential supervised settings.

However, in continual multi-domain virtual IHC staining, preserving cross-biomarker structural relations is of central importance. Different IHC biomarkers exhibit distinct staining patterns—for example, HER2 primarily localizes to the cell membrane, whereas ER, PR, and Ki-67 are expressed in the cell nucleus. Despite these differences in molecular expression and spatial localization, all biomarkers are observed within the same underlying tissue architecture, sharing consistent glandular structures, cellular arrangements, and tumor regions. Consequently, their latent representations are inherently structured and interrelated. If these shared structural relations are disrupted during sequential adaptation, the model gradually loses its alignment to the underlying tissue morphology, resulting in representation drift and unstable generation as shown in Fig. 1 (c).

To address the above challenges, we propose an **IHC multi-domain relational distillation continual learning framework (ContiStain)**. To the best of our knowledge, this work is the first to formally establish continual multi-domain virtual IHC staining as a dedicated learning setting and to explicitly emphasize cross-biomarker structural preservation as a core objective. To maintain stable cross-biomarker relations throughout sequential domain adaptation, ContiStain is built upon two progressively structured components. First, we construct a domain-aware feature extractor using a Mixture-of-Experts module [18], where biomarker-specific experts model domain characteristics while shared components preserve morphology-aligned structures, providing a stable foundation for continual adaptation. Second, we introduce a relational distillation strategy that preserves cross-domain structural consistency within the MoE-structured latent space established in the first step, ensuring coherent structural relations among learned biomarkers during sequential updates. Extensive experiments demonstrate that ContiStain effectively mitigates forgetting while maintaining cross-domain stability in continual multi-domain virtual IHC staining.

The main contributions are as follows:

- 1) We introduce a relational distillation mechanism that operates within the structured latent space to explicitly maintain cross-domain structural consistency across biomarkers.
- 2) We propose a domain-aware structured feature modeling strategy based on MoE, which disentangles biomarker-specific characteristics from morphology-aligned structures to support stable continual adaptation.
- 3) Extensive experiments on the MIST dataset demonstrate the effect of the proposed framework under a sequential multi-domain setting.

2 Method

We consider a sequential multi-domain scenario in which biomarker domains $\{D_1, \dots, D_K\}$ are introduced progressively. At training step t , only data from the current domain D_t are accessible, while previous data are unavailable.

After training on D_{t-1} , the generator is frozen as a teacher model G^{t-1} , serving as a stable reference. During adaptation to D_t , the student model G^t is optimized using current-domain supervision together with a structural preservation objective computed against G^{t-1} on previous domains. Unlike replay-based methods, we preserve knowledge via cross-biomarker structural relations.

2.1 Base Multi-domain Virtual IHC Framework

We build our framework upon Adaptive Supervised PatchNCE (ASP), an encoder-decoder generative model for H&E-to-IHC translation trained with adversarial and patch-level contrastive supervision.

Given an H&E image x and a biomarker domain index $d \in \{1, \dots, K\}$, the generator produces a stain-specific output:

$$\hat{y}_d = G(x, d) \quad (1)$$

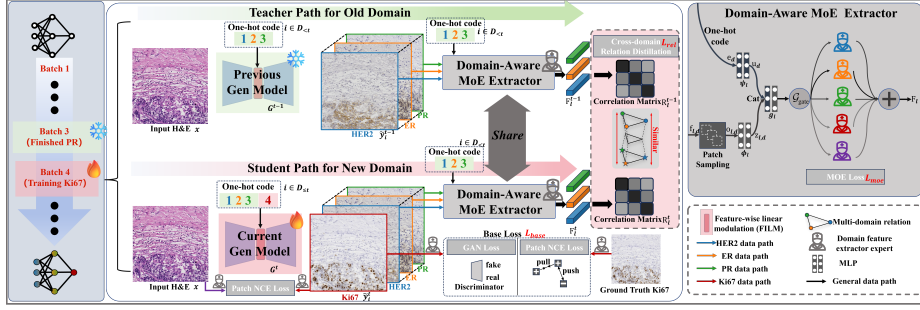


Fig. 2. ContiStain: continual virtual staining via MoE-structured features and cross-domain relational distillation. The fourth domain process is used for illustration.

In our setting, domain indices correspond to different biomarkers (e.g., 1: HER2, 2: ER, 3: PR, 4: Ki67).

Bottleneck-only domain injection preserves domain-invariant morphology and confines biomarker-specific modulation to a compact latent space by applying feature-wise linear modulation (FiLM) [16] at the bottleneck, where the domain index is encoded as a one-hot vector $\mathbf{e}_d$ and mapped to an embedding $\mathbf{u}_d = \psi(\mathbf{e}_d)$.

Let $\mathbf{h}_b$ denote the bottleneck feature following [12]. FiLM modulates it as:

$$\mathbf{h}'_b = (1 + \gamma_b(d)) \odot \mathbf{h}_b + \beta_b(d) \quad (2)$$

where $[\gamma_b(d), \beta_b(d)] = \phi_b(\mathbf{u}_d)$. This builds a lightweight unified conditional virtual IHC backbone for continual cross-domain relation preservation.

2.2 Domain-Aware MoE Extractor for Relational Modeling

A key challenge in continual multi-domain virtual staining is *representation interference*: as the generator adapts to a new biomarker domain, the latent representations of previously learned domains are implicitly reshaped. Such interference destabilizes cross-domain structural relations, making relation preservation difficult to enforce. To address this issue, we construct a *domain-aware* feature extractor F_{MoE} that produces structured and disentangled representations for different biomarker domains.

Patch Sampling and Shared Projection: For each biomarker domain $d \in \mathcal{D}$, we extract intermediate encoder features $\{\mathbf{f}_{\ell,d}\}_{\ell \in \mathcal{L}}$, where $\mathbf{f}_{\ell,d} \in \mathbb{R}^{B \times N_\ell \times C_\ell}$ denotes patch-level tokens at layer ℓ (N_ℓ is the number of spatial tokens).

We randomly sample P spatial tokens to obtain a compact set of structural descriptors, which are then projected into a shared embedding space:

$$\mathbf{o}_{\ell,d} = \text{Sample}(\mathbf{f}_{\ell,d}) \in \mathbb{R}^{B \times P \times C_\ell}, \quad \mathbf{z}_{\ell,d} = \phi_\ell(\mathbf{o}_{\ell,d}) \in \mathbb{R}^{B \times P \times C_\ell} \quad (3)$$

where ϕ_ℓ is a two-layer MLP applied to each token independently. The projection function is shared across domains to align patch-level representations in a unified

latent space. This shared space is designed to emphasize morphology-related structures rather than biomarker-specific appearance, enabling meaningful cross-domain relational comparison.

Domain-Aware Mixture-of-Experts Modeling: To explicitly incorporate domain structure while preserving content awareness, we introduce a domain-aware mixture-of-experts module.

The biomarker domain d is encoded as a one-hot vector $\mathbf{e}_d$ and mapped to a compact embedding $\mathbf{u}_d = \psi(\mathbf{e}_d)$. Importantly, the expert routing is conditioned on both the domain embedding and the image content representation. We denote the gating network by $\mathcal{G}_{\text{gate}}$, which corresponds to the $\mathcal{G}_{\text{gate}}$ block shown in Fig. 2. For each token embedding $\mathbf{z}_{\ell,d}$, the gating network $\mathcal{G}_{\text{gate}}$ predicts domain- and content-adaptive expert weights:

$$\mathbf{w}_\ell = \text{softmax}\left(g_\ell([\mathbf{z}_{\ell,d}, \mathbf{u}_d])\right) \quad (4)$$

where M experts $\{h_{\ell,m}\}_{m=1}^M$ produce the domain-aware structured feature:

$$\mathbf{F}_{\ell,d} = \text{Normalize}\left(\sum_{m=1}^M \mathbf{w}_{\ell,m} \odot h_{\ell,m}(\mathbf{z}_{\ell,d})\right) \quad (5)$$

This domain-aware structured feature space reduces cross-domain interference and provides a stable basis for relation preservation.

MoE Regularization. To enforce domain-consistent routing and avoid expert collapse, we regularize the gating probabilities, with the first term promoting domain-aligned expert preference and the second discouraging imbalanced expert usage:

$$\mathcal{L}_{\text{MoE}} = \lambda_1 \text{BCE}(\mathbf{w}_{\ell,m^*}, y_d) + \lambda_2 \left(\max_m \bar{p}_m - \frac{1}{M} \sum_{m=1}^M \bar{p}_m \right) \quad (6)$$

where $\mathbf{w}_{\ell,m^*}$ denotes the routing probability of reference expert m , y_d indicates the domain label, and $\bar{p}_m$ is the mini-batch average routing probability.

2.3 Multi-Domain Relational Distillation

While the domain-aware MoE builds a structured feature space, continual updates can still *distort cross-domain geometry*. We thus distill *relations* among previously learned biomarker domains using a frozen teacher from the last step.

Teacher–Student Generation for Old Domains. At continual step t , we keep the previous generator G^{t-1} frozen as the teacher (**preGAN**) and optimize the student G^t . Given an input H&E image x , for each old domain $i \in \mathcal{D}_{<t}$, we generate old-domain outputs from both teacher and student:

$$\tilde{y}_i^{t-1} = G^{t-1}(x, i), \quad \tilde{y}_i^t = G^t(x, i) \quad (7)$$

We then extract multi-layer encoder features and apply the same patch sampling and MoE projection module F_{MoE} to obtain structured feature tokens:

$$\mathbf{F}_{\ell,i}^{(\cdot)} = F_{\text{MoE}}\left(E_\ell(\tilde{y}_i^{(\cdot)}), i\right) \in \mathbb{R}^{P \times C} \quad (\cdot) \in \{t-1, t\}, \quad \ell \in \mathcal{S}. \quad (8)$$

where $\mathcal{S}$ denotes the set of selected feature scales.

Importantly, the domain index i and patch feature $E_\ell(\tilde{y}_i^{(\cdot)})$ are explicitly provided to the MoE router to ensure domain-aware structured features.

Cross-Domain Relation Distillation. For each pair of old domains (i, j) , we compute a patch-level cosine correlation matrix in the structured space:

$$\mathbf{R}_\ell^{(\cdot)}(i, j) = \text{Normalize}(\mathbf{F}_{\ell, i}^{(\cdot)}) \cdot \text{Normalize}(\mathbf{F}_{\ell, j}^{(\cdot)})^\top \in \mathbb{R}^{P \times P} \quad (9)$$

We then preserve cross-domain geometry by minimizing the discrepancy between the teacher and student relation matrices across selected layers and pairs:

$$\mathcal{L}_{\text{rel}} = \frac{1}{|\mathcal{P}||\mathcal{S}|} \sum_{(i, j) \in \mathcal{P}} \sum_{\ell \in \mathcal{S}} \|\mathbf{R}_\ell^t(i, j) - \mathbf{R}_\ell^{t-1}(i, j)\|_1 \quad (10)$$

In practice, we use output-level ℓ_1 distillation when learning the second domain (as cross-domain supervision is underdetermined with only one domain pair), and switch to the above relational distillation from the third domain onward.

2.4 Overall Objective

At training step t , the student model G^t is optimized using supervision from the current domain D_t together with the relational preservation objective.

The overall loss is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{base}} + \lambda_{\text{rel}} \mathcal{L}_{\text{rel}} + \lambda_{\text{moe}} \mathcal{L}_{\text{moe}} \quad (11)$$

where $\mathcal{L}_{\text{base}}$ denotes the original ASP training objective (including adversarial and contrastive terms) on the current domain samples $\tilde{y}_{D_t}^t$ and corresponding label, $\mathcal{L}_{\text{rel}}$ is the relational distillation loss on the previous domains, and $\mathcal{L}_{\text{moe}}$ represents optional routing regularization on the MoE feature extractor.

3 Experiments

Datasets: Experiments are conducted on the Multi-IHC Stain Translation (MIST) dataset [10], which contains paired 1024×1024 H&E–IHC patches at 0.4661 μm per pixel ($20\times$) for four biomarkers: HER2, Ki67, ER, and PR. The training set includes 4,642, 4,361, 4,153, and 4,139 pairs for HER2, Ki67, ER, and PR, respectively, with 1,000 testing pairs for each biomarker. The availability of multiple biomarker domains derived from whole-slide images enables systematic evaluation of continual multi-domain virtual IHC staining.

Implementation Details: We adopt ASP [10] as the backbone for virtual staining. All continual learning baselines are implemented on the same ASP framework, differing only in the adopted continual learning strategy. All input H&E–IHC patches are resized to 512×512 . In the continual multi-domain setting, biomarker domains are learned sequentially without accessing samples from previously learned domains. Unless otherwise specified, each domain is trained

Table 1. Domain-wise retention analysis (HER2→ER→PR→Ki67). The *Initial* row shows performance after the first-stage training, and the remaining rows report final results after continual learning. Best results are highlighted in **bold** (excluding *Initial*).

Method	HER2					ER				
	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓
Initial	67.0973	88.8182	14.9335	0.1544	0.3045	41.2629	45.6739	13.3665	0.1368	0.2739
Seq-FT	80.9059	265.9206	13.7123	0.1323	0.2898	62.7876	98.2752	13.7309	0.1526	0.2797
LWF	84.7352	148.6154	14.1219	0.1286	0.2830	53.1903	62.7852	14.6164	0.1766	0.2765
EWC	78.0549	197.3761	13.3334	0.1130	0.2828	69.1685	78.3096	13.1532	0.1336	0.2776
iCaRL	76.6097	198.2103	14.8120	0.1644	0.2923	48.4756	59.1977	13.2304	0.1365	0.2729
Ours	66.1568	111.9423	13.6269	0.1153	0.2798	47.4210	47.1893	13.8484	0.1419	0.2677
Method	PR					Ki67				
	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓
Initial	41.8015	36.3097	14.3173	0.1572	0.2668	—	—	—	—	—
Seq-FT	63.9035	98.4476	13.7745	0.1579	0.2853	39.3537	33.0343	14.4889	0.1615	0.2478
LWF	46.1695	47.8434	13.9698	0.1473	0.2648	43.2973	38.9575	14.5766	0.1590	0.2503
EWC	78.0882	104.4298	13.7501	0.1377	0.2737	64.4175	49.9091	13.1746	0.1316	0.2655
iCaRL	51.3151	69.0636	13.5846	0.1381	0.2729	50.5395	60.1791	14.7017	0.1602	0.2619
Ours	44.8295	38.1893	14.7375	0.1599	0.2561	44.3424	54.6830	14.9143	0.1739	0.2603

Table 2. Performance under reversed domain order (Ki67→PR→ER→HER2). The *Initial* row shows performance after the first-stage training, and the remaining rows report final results after continual learning. Results are averaged across all four domains to evaluate robustness to domain sequence variation. Best results are highlighted in **bold** (excluding *Initial*).

Method	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓
Initial	48.3852	50.5980	14.5974	0.1564	0.2666
Seq-FT	64.1188	161.7601	13.5295	0.1444	0.2887
LWF	56.4544	70.2778	14.8404	0.1694	0.2766
EWC	70.5058	193.4794	14.1416	0.1621	0.2923
iCaRL	55.9826	81.6900	13.4792	0.1389	0.2776
Ours	51.9917	63.9831	14.7410	0.1637	0.2711

for 50 epochs with a batch size of 1 on a single NVIDIA RTX 3090 GPU. The learning rate is set to 1×10^{-4} using the Adam optimizer.

Evaluations: We assess image quality using PSNR [1] and SSIM [19], and distribution-level similarity via FID [7]. To better capture histopathology-specific features, we report CONCH-FID [2] using a pathology-pretrained CONCH encoder [14], which captures pathology-aware semantic similarities through pathology image-text contrastive pretraining and provides a more clinically relevant assessment of staining quality, and DISTS [4] for structural and textural perceptual consistency.

Comparison Study: Table 1 reports domain-wise retention after continual training (HER2→ER→PR→Ki67). Our method achieves competitive or best performance, with lower FID/ConchFID on HER2/ER, superior PR results, and highest PSNR/SSIM on Ki67, indicating better pathological consistency and structural reconstruction. Table 2 reports averaged results under the reversed domain order (Ki67→PR→ER→HER2), indicating robustness to domain sequence variation. Fig. 3 illustrates that our method preserves HER2 membrane staining and achieves more accurate nuclear staining for ER, PR, and Ki67. The t-SNE

distributions in Fig. 4 show Seq-FT features scattered, while Ours aligns closely with Init, demonstrating improved cross-domain retention.

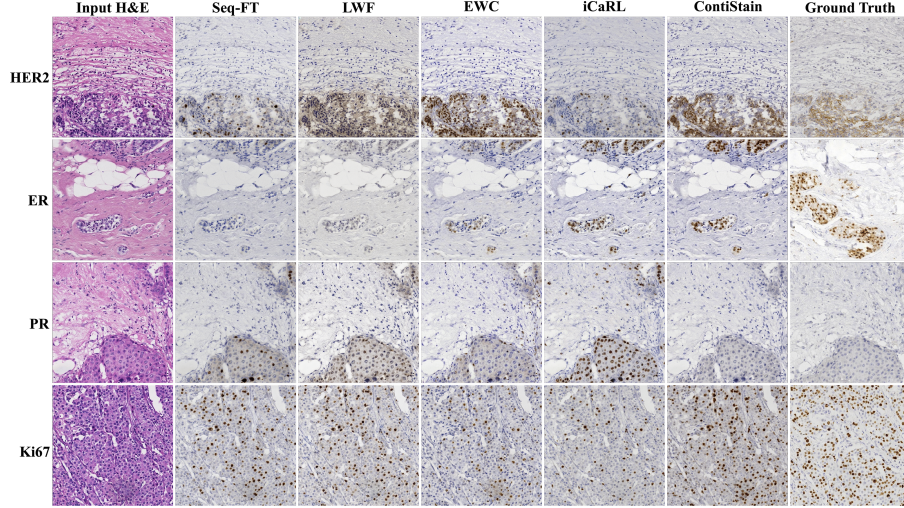


Fig. 3. Qualitative visualization of the final results on HER2, ER, PR, and Ki67.

Ablation Study: Table 3 presents ablation results on average performance. Adding the relation distillation significantly improves FID and ConchFID. Further incorporating MoE yields consistent gains across perceptual and fidelity metrics. For the loss weight, $\lambda_{rel} = 1$ achieves the best trade-off, while overly large weights degrade stability. Regarding expert numbers, 5 experts (IHC and H&E) provide the most balanced performance.

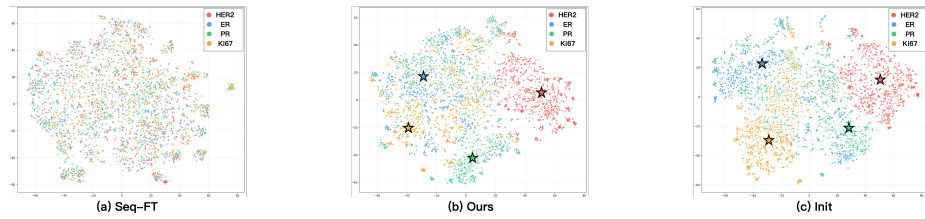


Fig. 4. Qualitative comparison of distributions for Seq-FT(final), Ours(final), and Init.

Table 3. Ablation study on the average performance across all domains.

Type	Setting	FID↓	ConchFID↓	PSNR↑	SSIM↑	DISTS↓
Module	Baseline	61.7377	123.9194	13.9267	0.1527	0.2757
	+ Rel Distill	57.9892	76.3321	13.8988	0.1454	0.2725
	+ Rel Distill+ MoE	50.6874	63.0010	14.2818	0.1478	0.2660
Weight (λ_{rel})	0.1	57.1981	74.1813	14.2812	0.1497	0.2722
	1	50.6874	63.0010	14.2818	0.1478	0.2660
	10	89.8493	163.6526	14.6272	0.1628	0.2923
	100	121.3326	235.9494	14.7138	0.1614	0.3081
Experts	2	65.8156	93.5733	14.2793	0.1451	0.2793
	5	50.6874	63.0010	14.2818	0.1478	0.2660
	10	58.9102	86.5717	14.1437	0.1462	0.2762

4 Conclusion

We formalize continual multi-domain virtual IHC staining and identify cross-biomarker structural preservation as its key challenge. To mitigate representation interference and structural drift, we propose ContiStain, a domain-aware relational distillation framework with structured MoE features and cross-domain structural consistency. Experiments on the four-domain MIST show it effectively reduces forgetting while maintaining structural coherence, suggesting that preserving latent cross-domain relations is a practical mechanism for scalable and robust continual virtual staining under real clinical constraints.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under 62331011.

Disclosure of Interests. The authors declare no competing interests in the paper.

References

1. Avciabas, I.s., Sankur, B.I., Sayood, K.: Statistical evaluation of image quality measures. *Journal of Electronic Imaging* **11**(2), 206–223 (2002)
2. Chen, F., Zhang, R., Hu, W., Abera, D.E., Peng, Y., Zheng, B., Sun, Y., Cai, J., Qin, W.: Pgvms: A prompt-guided unified framework for virtual multiplex ihc staining with pathological semantic learning. *IEEE Transactions on Medical Imaging* (2026)
3. Chen, F., Zhang, R., Zheng, B., Sun, Y., He, J., Qin, W.: Pathological semantics-preserving learning for h&e-to-ihc virtual staining. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 384–394. Springer (2024)
4. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. *IEEE transactions on pattern analysis and machine intelligence* **44**(5), 2567–2581 (2020)
5. Dubey, S., Chong, Y., Knudsen, B., Elhabian, S.Y.: Vims: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains. In: *International Workshop on Machine Learning in Medical Imaging*. pp. 143–155. Springer (2024)
6. Guan, X., Zhang, Z., Wang, Y., Li, Y., Zhang, Y.: Supervised information mining from weakly paired images for breast ihc virtual staining. *IEEE Transactions on Medical Imaging* **44**(5), 2120–2130 (2025)

7. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30** (2017)
8. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences* **114**(13), 3521–3526 (2017)
9. Klöckner, P., Teixeira, J., Montezuma, D., Fraga, J., Horlings, H.M., Cardoso, J.S., Oliveira, S.P.: H&e to ihc virtual staining methods in breast cancer: an overview and benchmarking. *Npj digital medicine* **8**(1), 384 (2025)
10. Li, F., Hu, Z., Chen, W., Kak, A.: Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, October 8–12, 2023, Proceedings, Part VI*. pp. 632–641. Springer (2023)
11. Li, Z., Hoiem, D.: Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence* **40**(12), 2935–2947 (2017)
12. Lin, Y., Zeng, B., Wang, Y., Chen, Y., Fang, Z., Zhang, J., Ji, X., Wang, H., Zhang, Y.: Unpaired multi-domain stain transfer for kidney histopathological images. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 36, pp. 1630–1637 (2022)
13. Liu, Z., He, Y., Zhang, T., Ma, C., Song, F., Wu, H., Cai, R., Guo, H., Zhang, H., Wen, B., et al.: Stainexpert: A unified multi-expert diffusion framework for multi-target pathological stain translation. *IEEE Transactions on Medical Imaging* (2025)
14. Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. *Nature Medicine* **30**(3), 863–874 (2024)
15. Peng, Y., Xiong, B., Chen, F., Abera, D.E., Zhang, R., Hu, W., Cai, J., Qin, W.: Usigan: Unbalanced self-information feature transport for weakly paired image ihc virtual staining. *IEEE Transactions on Image Processing* (2026)
16. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 32 (2018)
17. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. pp. 2001–2010 (2017)
18. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538* (2017)
19. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004)
20. Xiong, B., Peng, Y., Zhang, R., Chen, F., He, J., Qin, W.: Unpaired multi-domain histopathology virtual staining using dual path prompted inversion. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 8780–8787 (2025)
21. Zhang, C., Xie, Y., Li, Y., Liang, X., Lin, L., Chen, Y.W.: Pd-unist: Prompt-driven universal model for unpaired h&e-to-ihc stain translation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 463–472. Springer (2025)