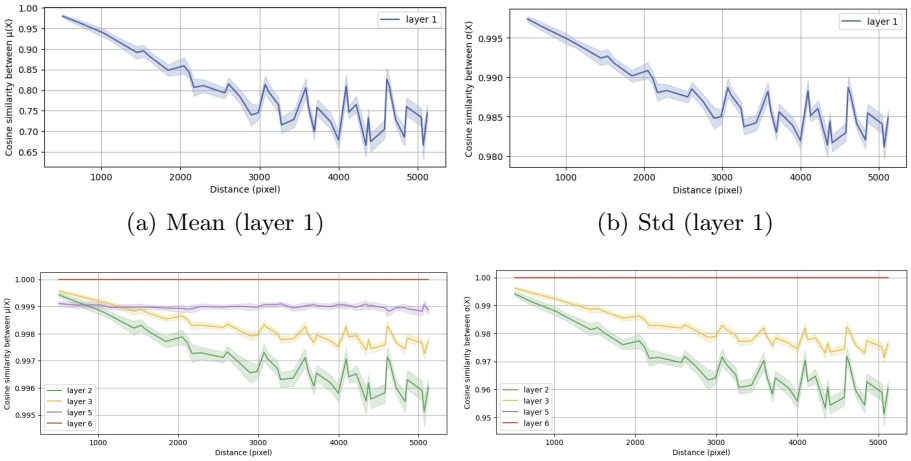
Ultra-high-resolution unpaired stain transformation via Kernelized Instance Normalization (Supplementary Material)

Ming-Yang Ho, Min-Sheng Wu, and Che-Ming Wu

aetherAI, Taipei, Taiwan
{kaminyouho,vincentwu,uno}@aetherai.com
https://www.aetherai.com/

0.1 Analysis



(a) Mean (layer 1) (b) Std (layer 1)

(c) Mean (layer 2-6) (d) Std (layer 2-6)

Fig. S1. Comparison of mean and std calculated in IN between adjacent patches. Mean, and standard deviation (std) of every two adjacent or nearby patches (up to 5,000 pixels far away) were extracted from the IN in the original CycleGAN model and compared. The CycleGAN model comprises 6 layers and each has one or multiple IN: (1) convolutional layer; (2) down-sampling layer; (3) down-sampling layer; (4) residual backbone; (5) up-sampling layer; (6) up-sampling layer. We analyzed mean and std from IN in all layers except the fourth layer, which is a backbone. It can be noticed that there is a great discrepancy in mean and std between faraway patches in the earlier layers.

To verify our hypothesis, we extracted the $\mu(X)$ and $\sigma(X)$ from all the IN layers in $\mathcal{G}$ for patches cropped from one single image, in which $\mu(X), \sigma(X) \in$

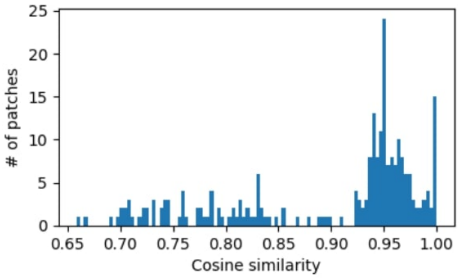


Fig. S2. Distribution of cosine similarity between means of thumbnail and patches calculated in IN. The means of patches and the thumbnail calculated in the IN layer from layer 1 of CycleGAN's generator are extracted and compared. Distribution of the cosine similarity is shown. An obvious discrepancy can be observed, which indicates the inappropriateness of using thumbnail statistics for all cropped patches in the TIN [1].

$R^{1\times C}$, and C is the number of channels. Then, $\mu(X)$ and $\sigma(X)$ were further flattened into vectors with size C to compute the cosine similarity between every pair. Besides, the Euclidean distances between pairs were recorded.

Fig. S1 demonstrates that cosine similarity of $\mu(X)$ and $\sigma(X)$ between two patches would dramatically decrease when two patches are farther apart, especially in the first few layer blocks.

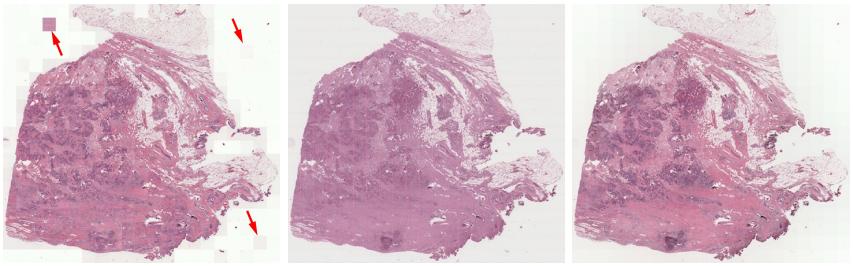
In addition, we adopted the methodology proposed in TIN [1] and measured the $\mu(X)$ and $\sigma(X)$ between the thumbnail and other cropped patches in Figure S2. It shows that extreme inconsistency occurs in the first few layers, implying local contrast and hue information will diminish if μ and σ of thumbnail are used. On the contrary, the convolution mechanism in our KIN can both alleviate this inconsistency issue and further improve the assembly quality when adjacent patches are combined.

0.2 Performance on the classification downstream task

As there is no well-developed metric that can evaluate unpaired ultra-high-resolution (UHR) images, downstream classification task was experimented to address this issue. We conducted a classification task for the ANHIR dataset (breast, lung lesion, and COAD). A ResNet-50 model was trained on the patches cropped from real WSIs in the IHC domain and tested on the patches cropped from translated WSIs generated by patch-wise IN, TIN, and KIN with the CUT framework. We deliberately cropped patches from the attached boundary to evaluate the influence of tilting artifacts. The accuracies of patch-wise IN, TIN, and **KIN** are 98.8%, 88.4%, and **99.2**%, respectively. The results show that KIN achieves the best performance, which might be due to the reduction of tilting artifacts that confused the classifier. TIN obtains the worst performance since using global statistics might lead to the loss of local information.

0.3 Evaluated by SSIM and FSIM metrics

To evaluate KIN with SSIM and FSIM metrics, we experimented with pairwise translating gray images of the ANHIR dataset into H&E. However, both SSIM (patch-wise IN: 0.94, TIN: 0.90, KIN: 0.93) and FSIM (patch-wise IN: 0.79, TIN: 0.74, KIN: 0.78) cannot evaluate the presence of tilting artifacts in patch-wise IN (see Fig. S3).



(a) Patch-wise IN (b) TIN (c) KIN

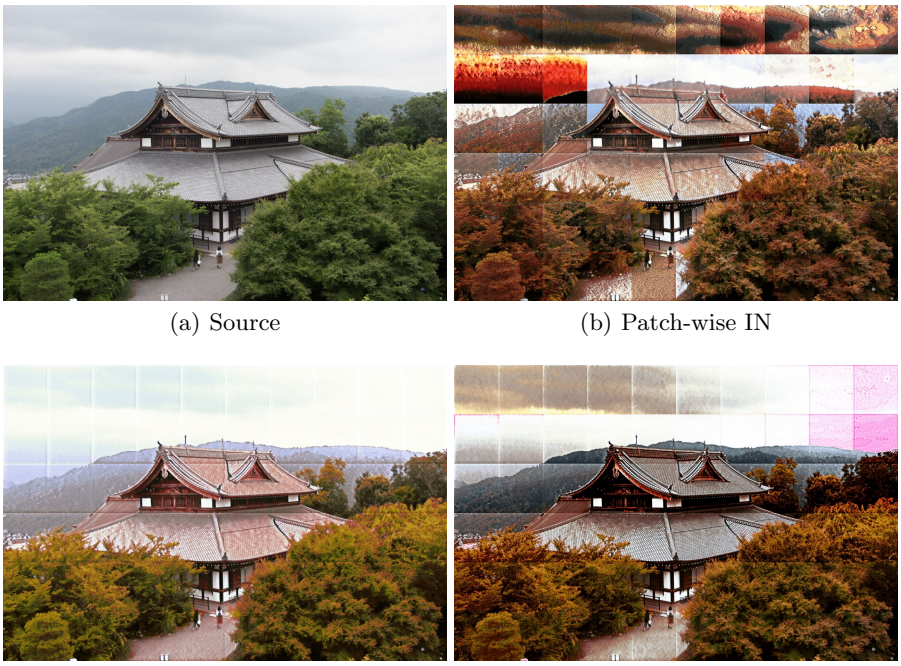
Fig. S3. Generated RGB WSIs by different methods. The presence of tilting artifacts, indicated by red arrows, cannot evaluated by SSIM or FSIM metrics.

0.4 Failure modes of KIN

If the training data lack enough specific scene (e.g., the sky in Kyoto dataset), KIN will be inferior to TIN (see Fig. S4).

References

1. Chen, Z., Wang, W., Xie, E., Lu, T., Luo, P.: Towards ultra-resolution neural style transfer via thumbnail instance normalization. In: Proceedings of the AAAI Conference on Artificial Intelligence (2022)



(a) Source (b) Patch-wise IN

(c) TIN (d) KIN

Fig. S4. Failure modes. KIN will be inferior to TIN if training data lack enough specific scene.

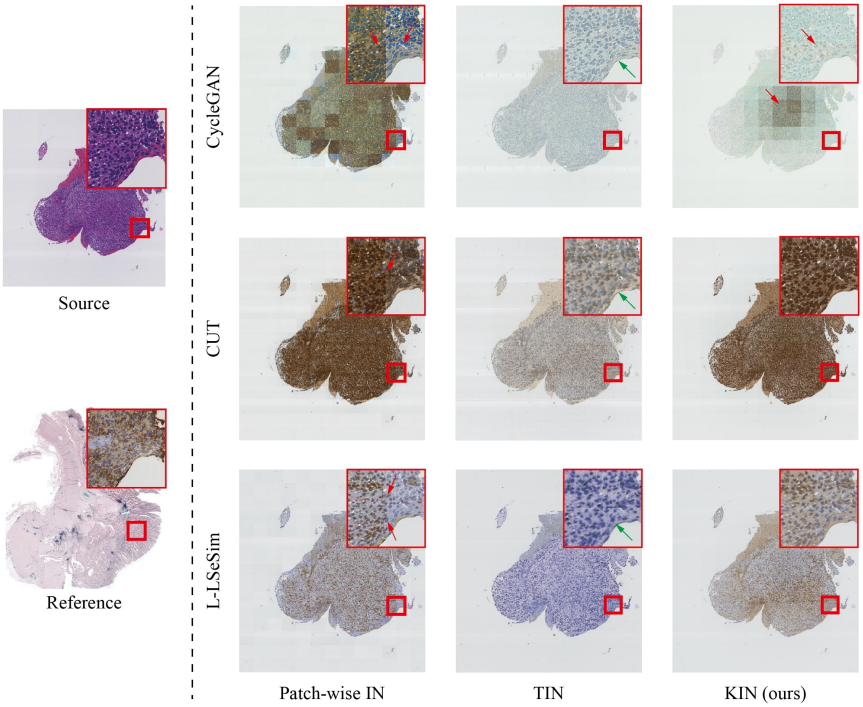


Fig. S5. H&E-to-EGFR stain transformation results on Glioma training set ($7,755 \times 7,109$ pixels) generated by different frameworks with IN, TIN, and KIN layers. Red arrows indicate tilting artifacts; green arrows indicate over/under-colorizing. CUT+KIN achieved the best performance. Zoom in for better view.

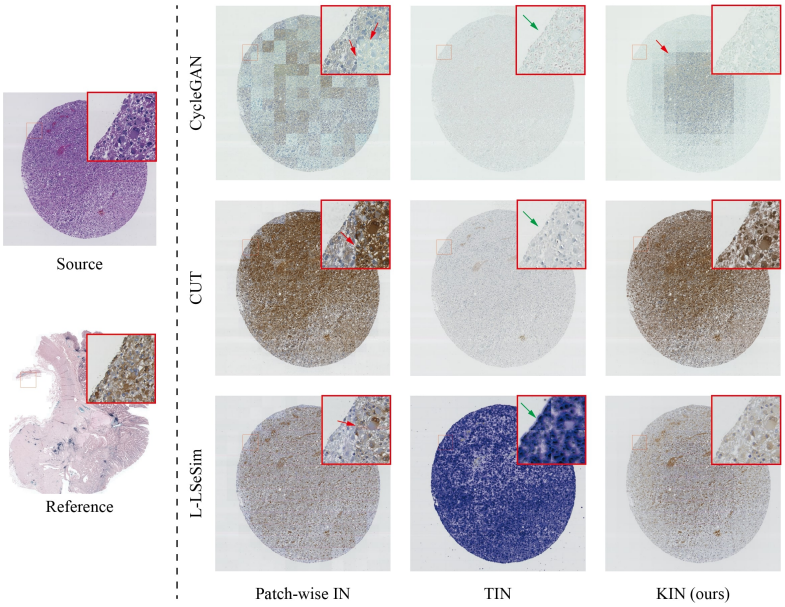


Fig. S6. H&E-to-EGFR stain transformation results on Glioma testing set ($8,078 \times 8,078$ pixels) generated by different frameworks with IN, TIN, and KIN layers. Red arrows indicate tiling artifacts; green arrows indicate over/under-colorizing. CUT+KIN achieved the best performance. Zoom in for better view.

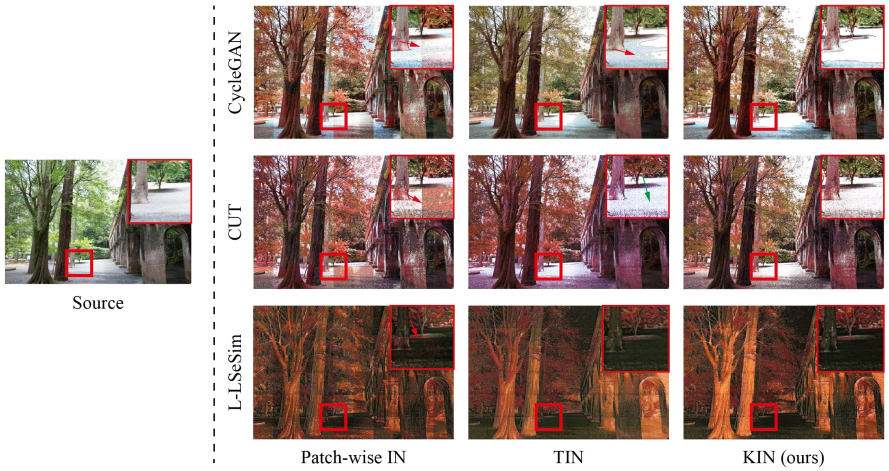


Fig. S7. Image-to-image translation results on Kyoto summer2autumn training set ($3,456 \times 5,184$ pixels) generated by different frameworks with IN, TIN, and KIN layers. Red arrows indicate tilting artifacts; green arrows indicate over/under-colorizing. CUT+KIN achieved the best performance. Zoom in for better view.

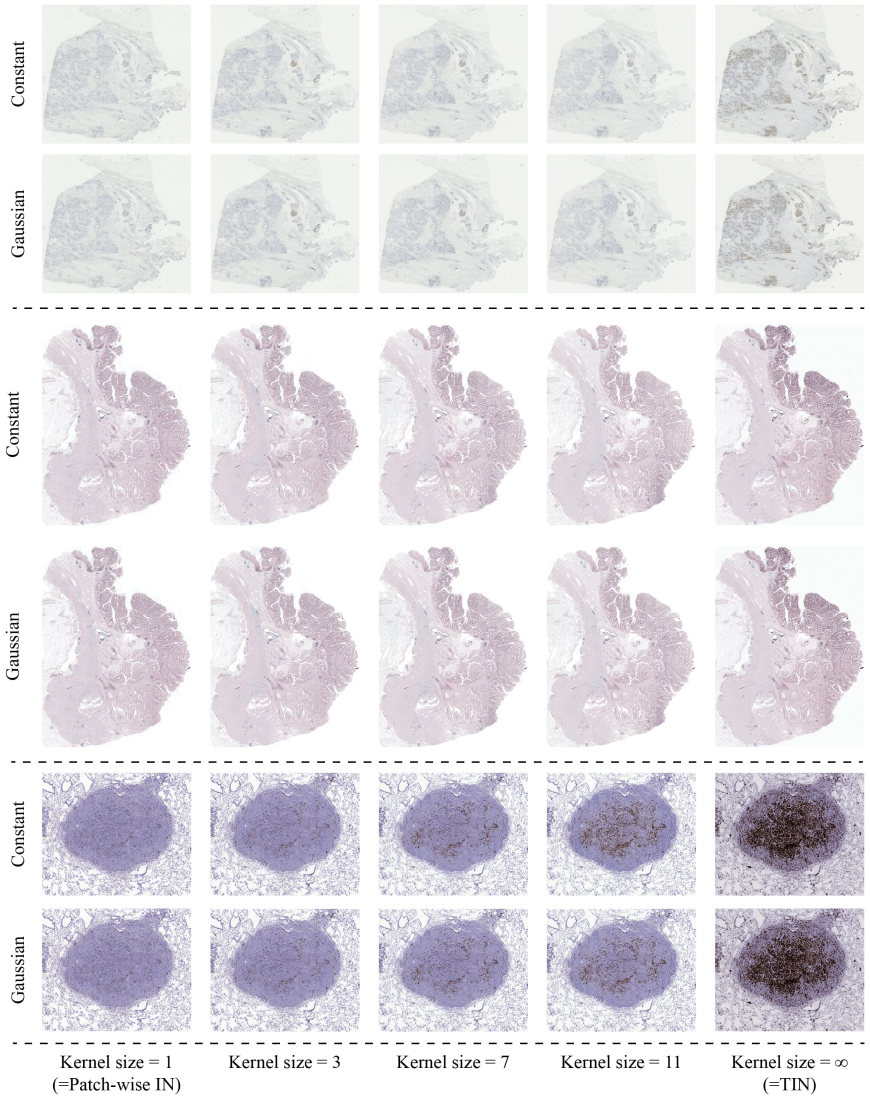


Fig. S8. Ablation study for kernel types on three ANHIR subdatasets. Constant and Gaussian kernels with the size of 1, 3, 7, 11, and ∞ are applied to elucidate the effect of KIN module. When kernel size is set to 1, the KIN module will operate in a manner of patch-wise IN, whereas it would be like TIN when kernel size is set to ∞.

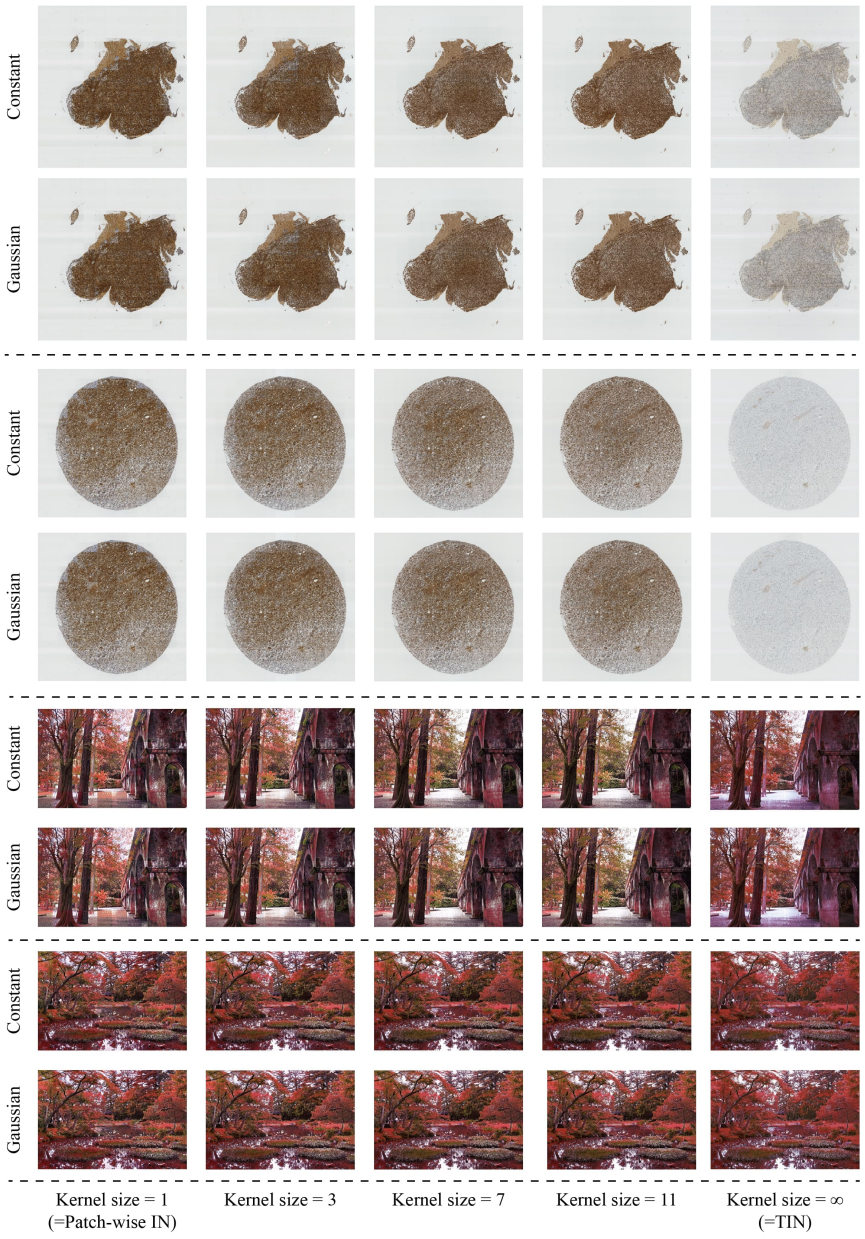


Fig. S9. Ablation study for kernel types on Glioma and Kyoto summer2autumn datasets. Constant and Gaussian kernels with the size of 1, 3, 7, 11, and ∞ are applied to elucidate the effect of KIN module. When kernel size is set to 1, the KIN module will operate in a manner of patch-wise IN, whereas it would be like TIN when kernel size is set to ∞.