HEMERA: H&E-to-HER2 Encoding for Morphology-Enhanced Region Annotation

Daniel Riccio[1,*], Francesco Longobardi[1], Mara Sangiovanni[1], Maria Frucci[2], Nadia Brancati[2], Mariacarla Staffa[3], Lorenzo D'Errico[1] and Antonio Ciccarelli[1]

[1]University of Naples Federico II, via Claudio 21, 80125 Napoli, Italy
[2]National Research Council (CNR) – Institute of High Performance Computing and Networking (ICAR), Via P. Castellino 111, 80131 Naples, Italy
[3]University of Naples Parthenope, Via Amm. F. Acton 38, 80133 Napoli, Italy

Abstract

The evaluation of the Human Epidermal Growth Factor Receptor 2 (HER2) marker is crucial for prognosis and therapy selection, but is hindered by variability in diagnostic results and ambiguity in cases of weak or borderline expression, which require confirmatory tests. Standard physical restaining techniques (IHC, FISH) involve expensive equipment, specialized reagents, and highly trained personnel, leading to diagnostic delays and significant economic burdens, especially in resource-limited settings. Although convolutional neural networks have been applied to automatically infer HER2 from Hematoxylin and Eosin (H&E) slides, they typically rely on an intermediate virtual staining step. In this work, we propose HEMERA (H&E-to-HER2 Encoding for Morphology-Enhanced Region Annotation), a deep learning framework that performs segmentation of HER2-positive regions directly from H&E images, without any intermediate physical or virtual restaining. HEMERA integrates: (i) *GenerationNet*, which leverages attention mechanisms focused on nuclei (the most informative structures for segmentation) and (ii) *SegmentationNet*, a U-Net whose encoder includes a frozen version of GenerationNet to generate synthetic HER2 features on the fly, combining them with H&E features to produce the final segmentation mask. Experiments were conducted on patch pairs extracted from the ACROBAT and BCI Grand Challenge datasets, demonstrating that HEMERA achieves promising segmentation performance across variable expression levels. The experimental results, although preliminary, are encouraging and demonstrate that, with further improvements, HEMERA can become a competitive and more cost-effective solution compared to traditional methods.

Keywords

Breast cancer, HER2, H&E staining, Deep learning, Image segmentation, Generative adversarial networks, U-Net, Digital pathology

1. Introduction

The diagnostic workflow of histological slides relies on routine hematoxylin–and–eosin (H&E) staining to assess cellular and tissue morphology, supplemented by immunohistochemical (IHC) stains and, when equivocal, by fluorescence *in situ* hybridization (FISH) to determine HER2 receptor status [1]. In breast cancers, HER2 is overexpressed in approximately 20–25% of cases and is a key prognostic and predictive marker, yet IHC interpretation can be affected by sample quality, protocol standardization, and pathologist expertise, while confirmatory FISH tests further increase time and cost [2].

These methods require specialized laboratories, costly reagents, and sophisticated instruments, leading to high healthcare expenses and delays in targeted treatments. Nevertheless, their roles remain essential in pathology: histochemical stains (e.g., trichrome, PAS) reveal tissue microstructures, while molecular methods (IHC, FISH) provide key biomolecular data for tumor classification and therapy decisions. Yet, H&E lacks specificity for proteins, and IHC, though precise, is laborious, costly, and time-consuming. Diagnostic efficiency is further limited by sample variability, protocol inconsistencies,

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
daniel.riccio@unina.it (D. Riccio); francesco.longobardi3@unina.it (F. Longobardi); mara.sangiovanni@unina.it (M. Sangiovanni); maria.frucci@cnr.it (M. Frucci); nadia.brancati@cnr.it (N. Brancati); mariacarla.staffa@uniparthenope.it (M. Staffa); lorenzo.derrico@unina.it (L. D'Errico); antonio.ciccarelli8@studenti.unina.it (A. Ciccarelli)

interpretive subjectivity, and tissue exhaustion, especially in resource-constrained settings or borderline cases (e.g., IHC 2+).

In response to these challenges, *virtual staining* has emerged as an innovative alternative: using pixel-based regression algorithms and generative models (GANs, diffusion models), H&E images can be digitally "restained" to simulate IHC or other special stains, reducing turnaround time, costs, and tissue consumption [3, 4, 5]. By leveraging morphological cues already present in H&E slides, these approaches predict the distribution of molecular markers, preserve the original tissue section, and enable multiple virtual stainings on the same specimen.

In this work, we introduce HEMERA, a novel end-to-end deep learning framework designed to segment HER2-positive regions directly from routine H&E images, without any intermediate physical or virtual restaining. HEMERA comprises two complementary CNNs: *GenerationNet*, which employs attention mechanisms to prioritize nuclear features critical for HER2 expression, and *SegmentationNet*, a U-Net–based architecture whose encoder incorporates a frozen GenerationNet backbone to generate synthetic HER2 feature maps on the fly. By fusing these synthetic markers with raw H&E representations, HEMERA produces accurate HER2 segmentation masks in a single pass.

2. State of the Art

In recent years, the need to overcome the constraints of traditional chemical staining has led to exploring "virtual staining" solutions. This research direction began around 2020, when initial studies demonstrated the feasibility of digitally generating immunohistochemical (IHC) or special stains from H&E–stained whole-slide images, aiming to preserve diagnostic material and apply multiple stainings to the same tissue section [3, 4, 5].

Early approaches relied on supervised pixel-level regression models or "vanilla" U-Net architectures to directly map H&E chromatic intensities to the expected IHC intensities [6], often generating overly smoothed results, missing subtle stain variations, and lacking robustness to minor protocol differences, hindering generalization.

Xu *et al.* introduced a conditional CycleGAN (cCycleGAN) for H&E→IHC translation, augmenting the adversarial loss with structural terms to preserve boundaries and morphology and improve color fidelity [7], leading to significantly enhanced membrane pattern reproduction and intensity variations, but susceptible to "hallucinatory" artifacts due to adversarial training dynamics.

More recently, Yang *et al.* proposed a cascaded deep neural network (C-DNN) pipeline that decouples the problem into autofluorescence→H&E and H&E→target stain stages, relaxing the need for perfectly paired data and achieving higher-quality virtual stains with improved structural and color similarity metrics compared to single-stage methods [8].

Generative Adversarial Networks, in particular the Pix2Pix architecture [9], where a generator learns to produce authentic virtual stains while a discriminator distinguishes real from generated images, quickly gained traction. Pix2Pix introduced enhanced color and structural realism via adversarial loss, preserving morphological details, such as cell contours and membrane intensity variations that pure regression methods often lose. Pioneer applications, such as those by de Haan *et al.* on renal biopsies, demonstrated marked diagnostic improvements over U-Net–based techniques [6], though GANs can introduce visual "hallucinations" and require large, well-balanced datasets, making training costly and sometimes unstable [9, 6].

The cutting edge now involves diffusion models [10, 11], which generate the target image by gradually denoising a noise map toward the desired stain. These models excel at reproducing the chromatic diversity of real stains [12] by modeling complex distributions without the oscillations typical of GANs.

High-fidelity virtual staining lays a reproducible chromatic foundation which, when integrated with semantic and instance segmentation models, enables precise delineation of regions of interest (e.g., HER2-positive areas), reducing uncertainty due to staining variability and improving the reliability of biomedical measurements.

Semantic segmentation models delineate exact object boundaries and classify pixels into predefined

classes (e.g., tumor vs. normal tissue) without distinguishing separate instances of the same class [13], while instance segmentation delineates individual objects, enabling counting and characterization of multiple coexisting entities [14].

In the biomedical domain, the U-Net [15] has become the gold standard for 2D and 3D segmentation due to its "U"-shaped architecture that effectively fuses low- and high-level contextual information.

2.1. CNN Architectures for Image-to-Image Transcoding and Segmentation

Before presenting HEMERA's architecture and its modular components, we briefly examine the seminal CNN paradigms, namely U-Net, generative adversarial networks (GANs), and residual networks (ResNet), which form the basis for the adaptations and extensions we introduce in our framework.

Digital pathology expands breast cancer profiling beyond HER2 by integrating markers like hormone receptors, Ki-67, and novel proteins, supporting personalized therapies. CNNs drive this field by extracting features from whole-slide images (WSIs) for diagnostic predictions. For virtual staining, autoencoders and later GANs were explored to generate realistic synthetic images, while in segmentation, U-Nets set a benchmark enabling precise delineation across scales. More specifically, the U-Net is an encoder–decoder architecture with a "U" shape designed for semantic segmentation of biomedical images [15]. In the contracting path, the input is processed by repeated 3×3 convolution blocks, each followed by ReLUs and 2×2 max-pooling layers. Skip connections transfer pre-pooled feature maps to the expanding path, where transposed convolutions restore resolution and further convolutions refine features. A final 1×1 convolution with sigmoid outputs the segmentation mask. Training usually combines cross-entropy (softmax-cross-entropy for multi-class) with a Dice term to balance local accuracy and global overlap.

U-Nets compress input into a latent "bottleneck" before reconstructing the segmentation mask from the latent space. A similar compact and representative embedding can be achieved with CNNs like ResNet, which use residual connections to stabilize deep training [16]. ResNet34 extends this idea with 34 layers: an initial 7×7 convolution and pooling, followed by four stages of residual blocks with increasing channels. Identity shortcuts connect most blocks, while 1×1 convolutions adjust dimensions at stage transitions. After global average pooling, a fully connected layer outputs classifications. These residuals enable efficient optimization and make ResNet34 a powerful feature extractor.

Conditional GANs (cGANs) enable controlled image synthesis using auxiliary inputs, such as class labels or source images. In digital pathology, Pix2Pix framed H&E–to–IHC translation as a cGAN task, with a U-Net generator and a PatchGAN discriminator enforcing realism and stain-specific details.

However, pixel-level losses like L1 are too rigid for histology due to slight, trivial misalignments between slides. Pyramid Pix2Pix addresses this by applying reconstruction loss across multiple resolutions via Gaussian pyramids, reducing penalties for minor shifts while preserving global color patterns and tissue architecture essential for analysis.

3. CNN Architectures Underpinning HEMERA

HEMERA combines two CNNs: *GenerationNet*, a specialized Pyramid Pix2Pix–based U-Net generator with a PatchGAN discriminator, translates H&E images into virtual HER2 stains using a cGAN loss with multiscale L1 terms to relax pixel-level constraints. *SegmentationNet*, an enhanced U-Net, fuses a pre-trained ResNet34 encoder with frozen GenerationNet feature maps; their latent vectors are concatenated before the bottleneck, and the decoder produces binary HER2-positive masks. Training uses a combined cross-entropy and Dice loss to maximize agreement with ground truth.

3.1. The GenerationNet Architecture

GenerationNet adopts the generator–PatchGAN discriminator configuration of Pyramid Pix2Pix, but introduces explicit attention to nuclear structures, which are critical in histopathology, while mitigating artifacts in cytoplasmic regions. Since cell nuclei contain the bulk of pathological information

(pleomorphism, hyperplasia, atypical mitoses), we leverage binary masks M that label nuclear pixels as foreground (set as 1) and cytoplasm plus other tissue components as background. To avoid introducing artifacts by assigning zero in non-nuclear areas, we compute the distance transform of each HER2 binary mask, add one to every entry (ensuring a minimum weight of one), and normalize by the maximum value to obtain a spatial weight map. Concretely, we define $D = \phi(M), \quad D_{i,j} = \min_{(u,v)|M_{u,v}=1} \sqrt{(i-u)^2 + (j-v)^2}$, and then

$$W_{i,j} = \frac{D_{i,j} + 1}{\max_{u,v}\big(D_{u,v} + 1\big)} \in [0, 1]\,, \tag{1}$$

During training, this weight map multiplies both the generated output and the ground truth before applying the loss components, focusing reconstruction error on nuclear regions with less emphasis on cytoplasm. This way, GenerationNet learns to faithfully reproduce nuclear contours and chromatic variations, enhancing the diagnostic quality of virtual HER2 images. The training minimizes a combination of losses, beginning with the adversarial cGAN loss and a weighted reconstruction loss, according to the aforementioned weight matrix W.

3.2. SegmentationNet Architecture

SegmentationNet was designed to leverage the ability of GenerationNet to reconstruct nuclear information missing from H&E images. The encoder of SegmentationNet combines two subnetworks: a pre–trained (on ImageNet) ResNet34 and the frozen GenerationNet, which generates a virtual HER2 image from the H&E input. Each one outputs a feature volume of size $512 \times 16 \times 16$ (after a series of convolutional and pooling layers). Through concatenation along the channel dimension, a latent vector of shape $1024 \times 16 \times 16$ is formed, which serves as the input to the bottleneck.

The decoder follows the classical U-Net design: after the bottleneck, repeated up-sampling stages (via transposed convolutions) and 3×3 convolutions with ReLU are applied, interleaved with skip-connections that link corresponding encoder levels. This preserves the global context while restoring the original spatial resolution.

SegmentationNet is trained using a composite loss that combines pixel-wise cross-entropy with the Dice coefficient using a weighting parameter λ. The Dice term penalizes mask discontinuities and promotes global alignment. During training, GenerationNet's weights are frozen, while ResNet34 and the decoder are optimized, allowing SegmentationNet to leverage synthetic HER2 features for more accurate segmentation of HER2-positive regions.

4. Experiments and Results

For the evaluation of HEMERA, we employed two oncological Whole-Slide Image (WSI) datasets from Grand Challenge: ACROBAT ("AutomatiC Registration Of Breast cAncer Tissue") [17] and BCI ("Breast Cancer Immunohistochemical Image Generation") [18]. The ACROBAT dataset comprises 750 routine clinical cases, each including one H&E–stained WSI and between one and four immunohistochemically (IHC) stained WSIs (HER2, ER, Ki-67, PgR) from primary breast carcinoma patients. In contrast, the BCI dataset consists of 4 873 paired H&E–HER2 images (9 746 total images), split into 3 896 pairs for training and 977 for testing, covering a range of HER2 expression levels. The initial preprocessing step selects only H&E and HER2-IHC WSIs, discarding all other stains. For ACROBAT only, HER2-IHC WSIs are then rigidly and non-rigidly registered to their H&E counterparts using DeeperHistReg [19], the Grand Challenge winner, ensuring precise overlay of stains. Then, from the aligned gigapixel WSIs, we extract 512×512 px patches (BCI images of 1024×1024 px are similarly partitioned). Patches with low tissue content (background), those that are blurred (measured by OpenCV sharpness), and, only for BCI, HER2 patches whose H&E counterpart fails a histogram similarity test are discarded.

Finally, for each retained HER2 patch, a ground-truth segmentation mask is generated using DeepLIIF [20], producing both binary HER2-positive masks and multi-class masks. These annotations serve as the gold standard for training and evaluating SegmentationNet.

4.1. Training and Testing

GenerationNet was trained on 15,000 H&E–HER2 patch pairs (512×512 px) from ACROBAT and BCI, split into 11 000 training and 4 000 validation patches. Training ran for 100 epochs using Adam (LR 2×10^{-4}, $\beta_1 = 0.5$, $\beta_2 = 0.999$) with batch size 2 on a single GPU, applying linear learning rate decay from epoch 50. The U-Net generator has 32 base filters, the discriminator is a PatchGAN, both with batch normalization and no dropout. The total loss combined cGAN adversarial, four-level multiscale L1 (weights 1,25,25,25), and balanced Dice terms. Testing on 1,000 unseen BCI patches used PSNR and SSIM to evaluate luminance, contrast, and structural similarity.

GenerationNet achieved an average PSNR of 21.669 dB (improved over Pyramid Pix2Pix's 21.160 dB) and an average SSIM of 0.355 (lower than the reference 0.477). The higher PSNR reflects reduced pixel-wise noise, particularly in nuclei, which are diagnostically critical. The lower SSIM arises from less emphasis on cytoplasmic regions and overall tissue texture, indicating that GenerationNet prioritizes nuclear fidelity while allowing minor degradation in less clinically relevant structures.

SegmentationNet was trained for 300 epochs with AdamW (LR 2×10^{-4}, batch size 16, BCE loss) on 42 000 HER2 patches from GenerationNet, validated on 8 000, and tested on 2 000 unseen patches. Evaluation yielded precision = 0.771, recall = 0.513, F1-score (Dice) = 0.631, and IoU = 0.487. High precision indicates few false positives, while relatively lower recall reflects substantial false negatives, likely due to the imbalance between nuclear and cytoplasmic regions. Overall, F1 and IoU are acceptable, but suggest that targeted loss weighting or balancing strategies could improve the detection of HER2-positive areas.

4.2. Performance Analysis

The results show that GenerationNet can reasonably approximate HER2 staining from H&E images. While the mean SSIM of 0.355 indicates limitations in reconstructing cytoplasmic structures and local contrast, the high PSNR confirms that nuclei are rendered with high fidelity. This demonstrates the feasibility of the generative–segmentation pipeline using only H&E inputs. The remaining gap compared to real IHC segmentation underscores the importance of generator quality and suggests directions for improvement: employing advanced generative models (e.g., Pix2PixHD, conditional diffusion, attention-guided GANs), next-generation segmentation architectures (U-Net++, DeepLabV3+), and integrating multimodal data to enhance structural consistency and HER2-positive region detection.

5. Conclusions

Histopathology remains the gold standard for breast cancer diagnosis, with H&E revealing morphology and HER2 providing prognostic information, but conventional IHC is costly, time-consuming, and variable. This study presents HEMERA, a pipeline combining GenerationNet (Pyramid Pix2Pix) to create virtual HER2 stains from H&E images and SegmentationNet (U-Net with a hybrid ResNet34–GenerationNet encoder) to segment HER2-positive regions automatically. Results demonstrate the feasibility of virtual staining, achieving particularly high fidelity in nuclear structures.

These results provide a promising proof of concept: while performance does not yet match direct segmentation on real IHC images, they demonstrate that a virtual staining and segmentation workflow is feasible. SSIM-based structural discrepancies highlight the need for stronger generators and advanced segmenters to better preserve contrast and tissue morphology. In conclusion, HEMERA establishes a foundation for extracting IHC-like molecular information directly from H&E images, offering a pathway toward faster, more cost-effective, and reproducible pathological assessment, with further architectural improvements expected to enhance clinical applicability.

Acknowledgments

We acknowledge financial support from the PNRR MUR project PE0000013-FAIR.

Declaration on Generative AI

During the preparation of this work, the authors used **ChatGPT** in order to perform:

- **Grammar and spelling check**: to correct any grammar or spelling errors that may have been missed
- **Paraphrase and reword**: to rephrase some paragraphs in a more concise and clear way

In all cases, authors have revised and, when necessary, edited the affected content and take full responsibility for the publication's content.

References

[1] Associazione Italiana di Oncologia Medica (AIOM), Linee guida AIOM: carcinoma mammario, https://www.aiom.it/linee-guida-aiom-carcinoma-mammario-2020, 2020.

[2] D. J. Slamon, G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich, W. L. McGuire, Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer, Science 235 (1987) 177–182. doi:10.1126/science.3798106.

[3] P. de Haan, A. Flores, I. Koulikov, S. Matsoukas, S. A. Tsaftaris, Deep learning-based transformation of h&e stained tissues into special stains, Journal of Pathology Informatics 12 (2021) 15. doi:10.4103/jpi.jpi_47_20.

[4] L. Chen, N. Mira, G. Wang, Y. Huang, J. Fan, Deep learning-based virtual staining for histopathology images, Scientific Reports 10 (2020) 1453. doi:10.1038/s41598-020-58190-7.

[5] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242–2251. doi:10.1109/ICCV.2017.244.

[6] K. de Haan, Y. Zhang, J. E. Zuckerman, T. Liu, A. E. Sisk, M. F. P. Diaz, K.-Y. Jen, A. Nobori, S. Liou, S. Zhang, R. Riahi, Y. Rivenson, W. D. Wallace, A. Ozcan, Deep learning-based transformation of the h&e stain into special stains, arXiv preprint arXiv:2008.08871 (2020).

[7] Z. Xu, X. Huang, C. Fernández Moro, B. Bozóky, Q. Zhang, Gan-based virtual re-staining: A promising solution for whole slide image analysis, arXiv preprint arXiv:1901.04059 (2019).

[8] X. Yang, B. Bai, Y. Zhang, Y. Li, K. de Haan, T. Liu, A. Ozcan, Virtual stain transfer in histology via cascaded deep neural networks, in: Optica Frontiers in Optics + Laser Science (FIO, LS), 2022.

[9] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1125–1134.

[10] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, arXiv preprint arXiv:1503.03585 (2015).

[11] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Advances in Neural Information Processing Systems, volume 33, 2020, pp. 6842–6853.

[12] P. Dhariwal, T. Salimans, Diffusion models beat gans on image synthesis, in: Advances in Neural Information Processing Systems, volume 34, 2021, pp. 8780–8794.

[13] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.

[14] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.

[15] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 234–241.

[16] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[17] P. Weitz, M. Valkonen, L. Solorzano, C. Carr, K. Kartasalo, C. Boissin, S. Koivukoski, A. Kuusela, D. Rasic, Y. Feng, et al., The acrobat 2022 challenge: automatic registration of breast cancer tissue, Medical image analysis 97 (2024) 103257.

[18] C. Zhu, S. Liu, Z. Yu, F. Xu, A. Aggarwal, G. Corredor, A. Madabhushi, Q. Qu, H. Fan, F. Li, et al., Breast cancer immunohistochemical image generation: a benchmark dataset and challenge review, arXiv preprint arXiv:2305.03546 (2023).

[19] M. Wodzinski, N. Marini, M. Atzori, H. Müller, Deeperhistreg: Robust whole slide images registration framework, arXiv preprint arXiv:2404.14434 (2024).

[20] P. Ghahremani, J. Marino, R. Dodds, S. Nadeem, Deepliif: An online platform for quantification of clinical pathology slides, arXiv preprint arXiv:2204.04494 (2022).