U-STAR: AN ASYMMETRIC U-SHAPED NETWORK BASED ON ELEMENT-WISE MULTIPLICATION TO SEGMENT NUCLEI IN H&E STAINED HISTOLOGICAL IMAGES

Guangzhengao Yang[1] *Li Zhang*[1⋆] *Jie Zhao*[2,3] *Zifan Chen*[1] *Haoshen Li*[1] *Bin Dong*[4,5,6]

[1] Center for Data Science, Peking University, China
[2] National Engineering Laboratory for Big Data Analysis and Applications, Peking University, China
[3] Peking University Changsha Institute for Computing and Digital Economy, China
[4] Beijing International Center for Mathematical Research, Peking University, China
[5] Center for Machine Learning Research, Peking University, China
[6] National Biomedical Imaging Center, Peking University, China

ABSTRACT

Nuclei segmentation in Hematoxylin and Eosin (H&E) stained images plays a crucial role in cancer diagnosis and pathological evaluation, enabling pathologists to identify abnormal cells and assess their morphology and distribution. While current automated nuclei segmentation methods predominantly employ convolutional neural networks and attention mechanisms, the potential of element-wise multiplication has been largely unexplored. This paper introduces U-Star, a novel asymmetric segmentation network based on the star block that leverages element-wise multiplication. U-Star adopts the classic encoder-decoder architecture and innovatively implements star-connection as an alternative to traditional skip-connections. In experiments on an H&E stained image dataset, U-Star achieved superior performance with a Dice coefficient of 0.8783, accuracy of 0.9089, and IoU of 0.7929, significantly outperforming baseline models. Extensive ablation studies validate the effectiveness of the star-connection and demonstrate the advantages of our proposed framework. Beyond advancing the application of element-wise multiplication techniques, U-Star shows promising potential for broader applications in medical image segmentation.

***Index Terms*—** element-wise multiplication, nuclei segmentation, tokenized MLP

1. INTRODUCTION

Hematoxylin and Eosin (H&E) staining remains the gold standard technique in pathological diagnostics, as it effectively enhances the contrast between nuclei and cytoplasm, enabling pathologists to identify and analyze diverse cell types and tissue structures. However, automated analysis of H&E stained images presents several technical challenges [1]. The heterogeneous morphology of nuclei, varying staining intensities, and frequent nuclear overlaps due to inconsistent tissue section thickness all contribute to the complexity of accurate image segmentation.

In recent years, neural network technology has made significant progress in the field of computer vision, particularly demonstrating great potential in medical image processing applications. These advanced technologies have opened up new possibilities for solving the problem of nucleus segmentation. The creation of AlexNet [2] represents a major breakthrough in deep learning within the field of computer vision and has directly promoted the widespread application of deep convolutional networks. VGG [3] and GoogLeNet [4] significantly improved network performance and efficiency by promoting the optimization and innovation of convolutional neural network structures. The residual learning introduced by ResNet [5] effectively tackled the challenges of training difficulty as network depth increased. In the field of medical image analysis, U-Net [6] marks a milestone, specifically designed for medical image segmentation; its unique symmetrical structure and skip connections significantly improved the capability of image processing. Transformer [7] employs Self-Attention that breaks through the limitations of traditional sequence processing models, performing excellently not only in text processing but also in visual tasks. Swin Transformer [8], through its hierarchical structure and sliding window mechanism, effectively reduces computation while maintaining the ability to capture global information, further extending the application of Transformers in the visual field.

Recently, researchers have increasingly focused on a computational method distinct from convolution and attention mechanisms: element-wise multiplication. FocalNet [9], HorNet [10], and VAN [11] analyzed why element-wise multiplication is effective based on intuition and assumptions, while Xu Ma et al. [12] referred to it as the "star operation", demonstrating its ability to map inputs to high-dimensional nonlinear feature spaces and introduced the concept validation network, StarNet.

The core of StarNet, star block, is a lightweight mod-

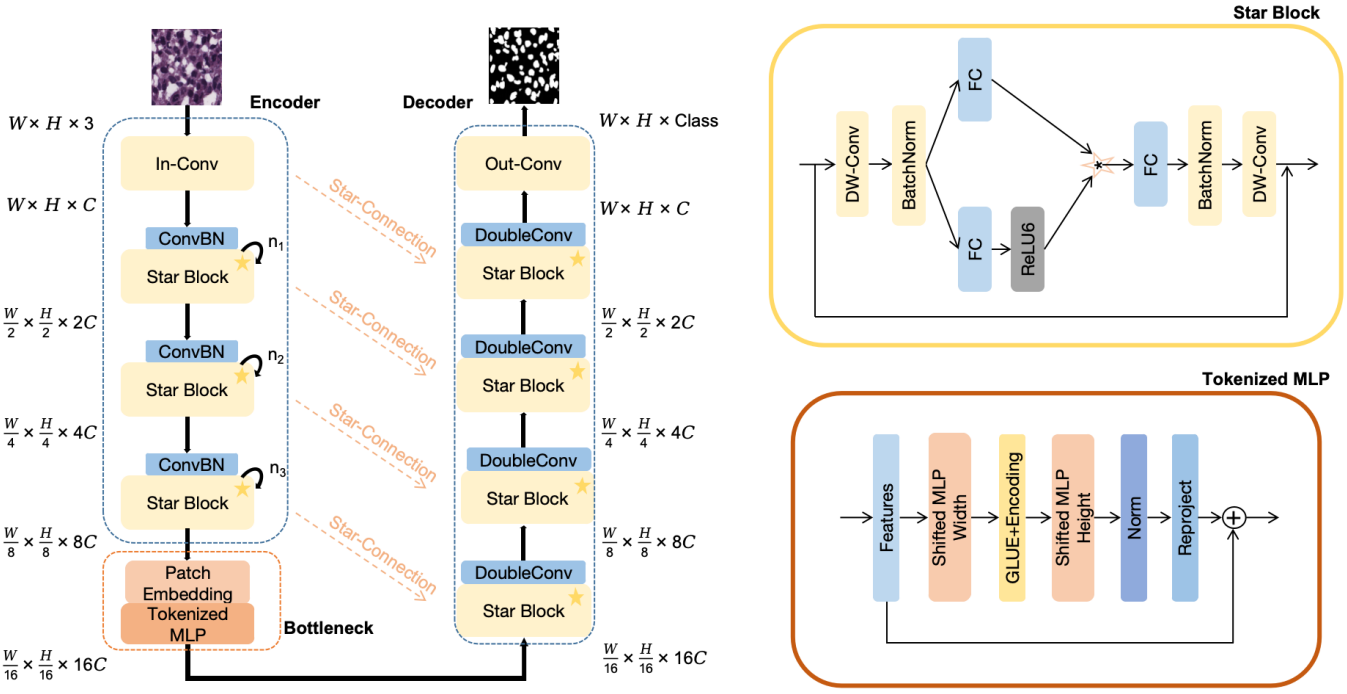

Fig. 1: Illustration of U-Star, an asymmetric U-shaped network based on element-wise multiplication. U-Star extracts features of the input image layer by layer through its decoder and achieves local interaction effects similar to a Swin Transformer block in a lightweight manner. Ultimately, through the layer-by-layer output of the decoder block, U-Star produces precise image segmentation results.

ule based on element-wise multiplication that exhibits exceptional performance. However, StarNet is mainly suitable for image classification since it primarily uses the star block for feature extraction and lacks an integrated decoding structure; also, a fully symmetric encoder-decoder structure is inappropriate. Because element-wise multiplication extracts hidden features by increasing the number of blocks, excessive stacking raises computational complexity and can lead to issues like overfitting or the curse of dimensionality. Therefore, different numbers of star blocks are required at various encoding and decoding stages to ensure optimal performance, explaining why U-Star adopts a non-strictly symmetric architecture. "Symmetry" is reflected in the same size of feature maps in the encoding and decoding stages, while "non-strict symmetry" is shown in the asymmetric connections of the star-connection and the different numbers of star blocks used by the encoder and decoder.

The bottleneck connects the encoder and decoder, enabling the model to learn deep feature representations. Swin-Unet [13] uses two Swin Transformer Blocks to optimize window locality, but this approach increases the number of parameters, which may hinder model convergence. Thus, we employ the lightweight token MLP [14] introduced by U-Next [15], which not only significantly reduces model parameters but also efficiently utilizes locality.

Our key contributions include:

1. Development of U-Star, a novel network architecture incorporating star blocks and star-connections for H&E stained nuclei segmentation, achieving robust performance without extensive hyperparameter tuning.
2. Introduction of an asymmetric network design that optimizes computational efficiency and feature extraction by strategically varying the distribution of star blocks between encoder and decoder stages, offering a more effective alternative to conventional symmetric U-shaped architectures.
3. Implementation of a lightweight token MLP bottleneck that efficiently connects the encoder and decoder while maintaining local feature contexts.
4. Comprehensive experimental validation of the proposed architecture through ablation studies and comparative analyses, providing empirical evidence for the effectiveness of each architectural component.

2. METHODS

2.1. Overall Framework

The overall architecture of the U-Star proposed in this study is shown in the figure 1. U-Star utilizes a non-rigidly symmetric encoder-decoder architecture, where the bottleneck section links the encoder and decoder. The encoder is primarily

responsible for feature extraction, while the decoder is tasked with mapping these features back to their original size to produce segmentation results. In this architecture, ConvBN consists of a convolution layer and a batch normalization layer, while DoubleConv is made up of two ConvBNs, with each ConvBN output connected to a ReLU activation function.

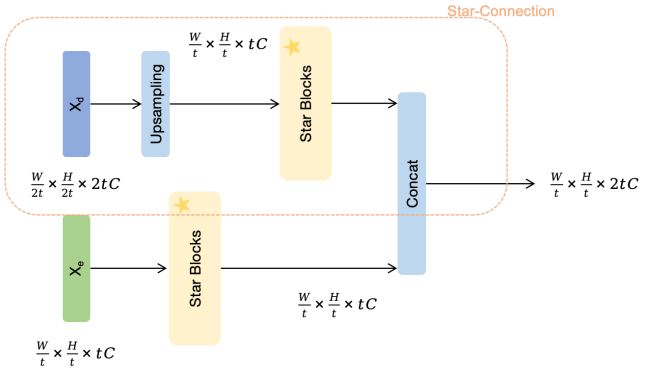

Fig. 2: Illustration of Star-Connection. Star-Connection first uses upsampling to make the features from the decoder match the dimensions of the features from the encoder, then independently passes them through a star block, and finally concatenates them together.

2.2. Encoder and Decoder

In the encoder, the input image first passes through the In-Conv module, which projects the feature dimensions to the integer C. The In-Conv module consists of a DoubleConv block. The features then pass through three encoding blocks, each containing a ConvBN and multiple star blocks, which reduce the feature resolution by a factor of two. ConvBN, responsible for downsampling, is configured with a 3×3 convolution kernel, padding of 2, and dilation of 1 in the batch normalization layer. Notably, in the encoder, the star blocks are stacked multiple times, utilizing their element-wise multiplication capability to map inputs to a high-dimensional nonlinear feature space, for more effective feature extraction.

In the decoder, features first pass through four decoding blocks, each consisting of a star block and a DoubleConv. Each decoding block aims to restore features, doubling their resolution. Before passing through DoubleConv, features undergo a star-connection with other features in the encoder. Finally, the output is processed through a 1×1 single convolution layer (Out-Conv) to complete segmentation or classification tasks. During the decoding phase, only a single star block is used to reduce disturbances caused by multiple stackings. Additionally, the decoding phase requires the fusion of feature maps from different layers (star-connection), where excessive stacking might reduce the robustness of the model.

2.3. Star Block

The structure of the star block is also shown in the figure 1. Specifically, a DWconv sized 7×7, with padding of 3 and stride of 1, is used to extract features, thereby preserving the size of the feature map. Utilizing DWconv substantially lowers parameter count and computational cost, thereby boosting computational efficiency. After processing with batch normalization, mapping is performed using two fully connected layers (FC). Once mapping is complete, one feature is activated using the ReLU6 function and then engaged in element-wise multiplication with another feature. Ultimately, features are extracted using batch normalization followed by another DWconv.

2.4. Star-Connection

We have adopted a novel feature fusion method called star-connection. The primary difference between star-connection and skip-connection is that the star block is used to map features during the connection process. As shown in the figure 2, x_e are features from the encoder with dimensions $\frac{W}{t} \times \frac{H}{t} \times tC$; x_d are features from the decoder with dimensions $\frac{W}{t} \times \frac{H}{2t} \times 2tC$. First, x_d is upsampled to the dimensions of $\frac{W}{t} \times \frac{H}{t} \times tC$ and passed through a star block for processing. Simultaneously, x_e is also processed through another star block. Then, these two processed features are concatenated to form features of dimensions $\frac{W}{t} \times \frac{H}{t} \times 2tC$, completing the star-connection. Experimental results show that star-connection performs better than skip-connection.

2.5. Bottleneck

The bottleneck part consists of patch embedding and tokenized MLP. Patch embedding is used for tokenizing features, and the subsequent tokens are processed by the tokenized MLP to achieve window locality in the network.

Patch embedding is implemented with a convolution kernel of size 3, stride 2, and padding 1. Subsequently, these tokens are input into the tokenized MLP block. The structure of the tokenized MLP block is shown in the diagram, with the shifted mlp as its core. Specifically, the shifted mlp splits features into m chunks along the channel, and each chunk is rolled by different units in height or width, mapped through an MLP to enhance feature locality.

3. EXPERIMENTS

3.1. Datasets

The Histology Image Dataset [15] used in this study is a composite of the publicly available MoNuSeg dataset [16] and another private dataset [17]. This dataset contains 462 H&E stained images, each with a resolution of 512×512 pixels.

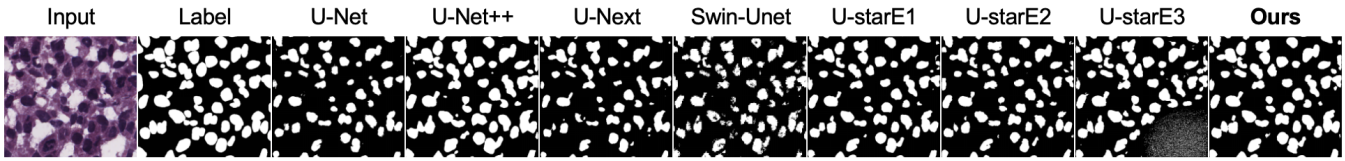

Fig. 3: Segmentation results using different models. The segmentation results include four baseline models and three sets of ablation studies. Our model is the closest to the label.

Models	DICE(%)	Acc(%)	IoU(%)	ErCnt(%)
U-Net	80.14	86.91	69.15	13.09
U-Net++	85.57	89.31	76.06	10.69
U-Next	82.04	87.52	71.33	12.48
Swin-Unet	70.57	81.52	57.99	18.48
U-StarE1	85.81	89.63	76.15	10.37
U-StarE2	79.07	83.43	68.04	16.56
U-StarE3	82.97	86.98	73.01	13.02
U-Star	**87.83**	**90.89**	**79.29**	**9.11**

Table 1: Performance metrics for different models

The labels are binarized and converted to black and white images to facilitate segmentation experiments. The dataset is divided in the ratio of training set: validation set: test set = $0.8 : 0.1 : 0.1$.

3.2. Implementation

In the U-Star implementation, the first convolutional layer in the DoubleConv block alters the number of channels, whereas the second one maintains it; the stack counts n_1, n_2, and n_3 for the star blocks are respectively set at 3, 3, and 5. We compared U-Net [6], U-Net++ [18], U-Next [14], and Swin-Unet [13] as baselines, and designed three additional architectures to enrich the experiment.

U-StarE1, Compared to U-Star, simplifies by replacing the star-connection with a regular skip-connection. U-StarE2 retains the star-connection but adds additional star blocks in each decoding block. U-StarE3 combines skip-connections with multiple star blocks stacked in the decoding blocks. All models are trained using the Adam optimizer [19] with a training duration set for 50 epochs. The loss function is a weighted combination of cross-entropy loss and DICE loss, with weights of 0.4 and 0.6. We selected the model with the best DICE coefficient on the validation set for testing.

4. RESULTS AND DISCUSSION

As demonstrated in Table 1 and Figure 3, U-Star consistently achieves superior performance across all evaluation metrics compared to baseline models.

To validate our architectural choices, we conducted comprehensive ablation studies. U-StarE1, which replaces star-connections with traditional skip-connections, shows performance degradation across all metrics: a 2.02% decrease in DICE score, 1.26% decrease in accuracy, and 3.14% decrease in IoU score. These results substantiate the effectiveness of our proposed star-connection mechanism in facilitating feature integration between encoder and decoder paths.

In U-StarE2, we investigated the impact of increasing star blocks in the decoder while maintaining star-connections. The observed significant performance deterioration indicates that star blocks are more effectively utilized in the encoder path, where feature extraction is paramount, rather than in the decoder path.

U-StarE3 combines traditional skip-connections with multiple star blocks in the decoder. While it slightly outperforms U-StarE2, possibly due to skip-connections mitigating the adverse effects of stacked star blocks in the decoder, its performance remains notably inferior to U-Star. The segmentation results visually confirm the presence of artifacts, suggesting that the combination of these modifications disrupts the network's feature reconstruction capability.

These ablation studies conclusively demonstrate that the optimal architecture comprises star-connections and a single star block in each decoder stage, as implemented in our final U-Star design. This configuration achieves the best balance between feature extraction, information flow, and computational efficiency.

5. CONCLUSION

In this paper, we present U-Star, a novel segmentation network architecture that leverages the power of star blocks for H&E stained image segmentation. Our work makes two significant architectural innovations: the implementation of an asymmetric design that challenges traditional U-shaped paradigms, and the introduction of star-connections for enhanced feature integration. Comprehensive ablation studies demonstrate that this design achieves superior segmentation performance while maintaining computational efficiency, without requiring complex parameter tuning. The success of U-Star not only advances the application of element-wise multiplication in medical image analysis but also opens new possibilities for its utilization in broader computer vision tasks. Future work could explore the adaptation of U-Star to other medical imaging modalities and investigate its potential in various segmentation applications beyond pathology.

6. REFERENCES

[1] Mitko Veta, Paul J Van Diest, Robert Kornegoor, André Huisman, Max A Viergever, and Josien PW Pluim. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PloS one*, 8(7):e70221, 2013.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[4] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[8] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[9] Jianwei Yang, Chunyuan Li, Xiyang Dai, and Jianfeng Gao. Focal modulation networks. *Advances in Neural Information Processing Systems*, 35:4203–4217, 2022.

[10] Yongming Rao, Wenliang Zhao, Yansong Tang, Jie Zhou, Ser Nam Lim, and Jiwen Lu. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems*, 35:10353–10366, 2022.

[11] Menghao Guo, Chengze Lu, Zhengning Liu, Mingming Cheng, and Shimin Hu. Visual attention network. *Computational Visual Media*, 9(4):733–752, 2023.

[12] Xu Ma, Xiyang Dai, Yue Bai, Yizhou Wang, and Yun Fu. Rewrite the stars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5694–5703, 2024.

[13] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[14] Jeya Maria Jose Valanarasu and Vishal M Patel. Unext: Mlp-based rapid medical image segmentation network. In *International conference on medical image computing and computer-assisted intervention*, pages 23–33. Springer, 2022.

[15] Zhenqi He, Mathias Unberath, Jing Ke, and Yiqing Shen. Transnuseg: A lightweight multi-task transformer for nuclei segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 206–215. Springer, 2023.

[16] Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng-Ann Heng, Jiahui Li, Zhiqiang Hu, et al. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 39(5):1380–1391, 2019.

[17] Jing Ke, Yizhou Lu, Yiqing Shen, Junchao Zhu, Yijin Zhou, Jinghan Huang, Jieteng Yao, Xiaoyao Liang, Yi Guo, Zhonghua Wei, et al. Clusterseg: A crowd cluster pinpointed nucleus segmentation framework with cross-modality datasets. *Medical Image Analysis*, 85:102758, 2023.

[18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, Cham, 2018. Springer International Publishing.

[19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.