TKS-UNet :TransUNet-based Interpretable Multiscale Feature Fusion Neural Network for Retinal Vessel Segmentation

Zhenyu Chen, Ziwei Zhou

Abstract—**Retinal vessel segmentation is vital for diagnosing eye diseases. Existing models such as U-Net and TransUNet encounter challenges in feature fusion and interpretability. This paper introduces TKS-UNet, a novel neural network architecture that addresses these issues. TKS-UNet incorporates the Kolmogorov-Arnold Network (KAN) to improve interpretability and revamps skip connections for multiscale feature fusion. Through extensive experiments on three public datasets (DRIVE, CHASE_DB1, and STARE), TKS-UNet demonstrates superior performance, achieving F1-scores of 86.09%, 90.89%, and 86.31% respectively. Ablation studies validate the efficacy of both the KAN module and the redesigned skip connections. This research presents a promising method for precise and interpretable retinal vessel segmentation, contributing to the progress of medical image analysis.**

Index Terms—**KAN, Semantic segmentation, Skip Connection, UNet**

I. INTRODUCTION

RETINAL vessel segmentation is crucial in ophthalmic diagnostics for early detection of diabetic retinopathy, glaucoma, and other vascular-related eye diseases. Convolutional neural networks (CNNs), notably U-Net and its variations, are prominent in medical image segmentation due to their hierarchical feature extraction capabilities using encoder-decoder architectures and skip connections [1]. The original U-Net [2] achieved 97.68% accuracy and 85.01% sensitivity on the DRIVE dataset [8], comprising 40 color fundus images. Li et al. (2019) [3] enhanced U-Net with connection-sensitive loss and attention gates, resulting in an 84.35% F1-score and 96.73% accuracy on STARE. Ren et al. (2022) [4] further improved U-Net by incorporating Bi-FPN fusion and data preprocessing techniques such as grayscale conversion, CLAHE, and gamma correction, achieving SP of 0.8604, SE of 0.9767, ACC of 0.9651, and AUC of 0.9787. Nevertheless, two primary challenges persist: inefficiencies in multiscale feature fusion and limited interpretability. The increase in network depth in U-Net-like structures leads to the loss of fine-grained details due to repeated downsampling operations. Additionally, conventional skip connections do not effectively bridge the semantic gap between encoder and decoder layers [7]. Recent advancements like TransUNet have integrated transformer-based encoders to improve global context modeling; however, they still rely on simplistic concatenation-based fusion methods, resulting in suboptimal performance on datasets with intricate vascular structures. Moreover, the opaque nature of deep neural networks impedes clinical acceptance, as healthcare professionals necessitate transparent decision-making processes to validate diagnostic outcomes.

This study presents TKS-UNet, a novel architecture that incorporates KAN into a TransUNet framework and revamps skip connections for integrating multiscale features. Traditional MLPs are replaced with KAN modules to improve interpretability and feature representation, while skip connections are redesigned based on UNet++ for hierarchical fusion of multiscale features. The performance of TKS-UNet is evaluated on three benchmark datasets (DRIVE, CHASE_DB1, STARE), achieving leading F1-scores of 86.09%, 90.89%, and 86.31%, respectively. By amalgamating interpretable neural elements with advanced feature fusion, TKS-UNet offers a transparent and precise solution for retinal vessel segmentation, thereby propelling the practical application of AI in ophthalmology.

II. METHODS

A. MLP replacement

The multilayer perceptron (MLP), a common model in machine learning for approximating nonlinear functions, faces interpretability and model forgetting limitations [5]. These hinder broader adoption, particularly in applications requiring transparency. In contrast, Kolmogorov-Arnold Networks (KAN) mitigate these limitations. Inspired by the Kolmogorov-Arnold representation theorem, KAN structurally resemble MLPs, featuring a fully connected network design [6]. The key difference lies in activation function configuration: while MLPs use fixed activation functions, KAN implement learnable activation functions on edges (weights). This design obviates the need for a linear weight matrix, replacing each weight parameter with a learnable one-dimensional spline function, substantially enhancing the network's expressive power and flexibility.

Kolmogorov-Arnold Networks (KAN) design nodes to perform only summation operations on input signals, avoiding nonlinear activation processing. This design choice preserves interpretability while maintaining model complexity, facilitating understanding of the model's decision-making process. Additionally, KAN exhibit

Manuscript received December 3, 2024; revised May 17, 2025.

Zhenyu Chen is a postgraduate student of the University of Science and Technology Liaoning, Anshan, 114051, China (corresponding author phone: 166-4223-8262; e-mail: chenzhenyu8882023@163.com).

Ziwei Zhou is a Professor at the University of Science and Technology Liaoning, Anshan, 114051, China (e-mail: 381431970@qq.com).

significant advantages in small sample learning scenarios. Their flexible weight representation allows for effective information extraction from limited training data, resulting in satisfactory generalization performance crucial for applications like medical image analysis.

Given these advantages, we propose the TKS-UNet hybrid structure as a novel approach. This hybrid model seeks to optimize feature extraction and representation capabilities of KAN while maintaining interpretability and efficiency. It addresses limitations of traditional deep learning models and encourages broader exploration and advancement in related fields.

B. *Redesign of skip Connections*

In practical image segmentation applications, the use of repeated downsampling operations results in the loss of detail features. The U-Net architecture employs skip connections to recover lost detail information [2]. However, with deep learning techniques, researchers propose enhanced approaches. One such approach is UNet++, which improves the model's capabilities and optimizes image segmentation [7].

The UNet++ network enhances the model's capacity to comprehend local features within a global context through a convolutional feature fusion approach. This improves information flow and feature reuse, improving the accuracy and detail of final segmentation results [7].

In this paper, we present an innovative reconstruction of traditional skip connections, inspired by the UNet++ network design concept. In the new network structure, original skip connections are removed and redesigned skip connections introduced in the first three encoder convolutional layers. This permits decoder nodes to receive feature maps at varying scales, facilitating diverse feature fusion. In particular, by transmitting feature maps of different scales between decoder and encoder, the aggregation layer efficiently fuses information from diverse layers and scales. This enhances feature representation extracted from input data.

Convolutional layer feature maps with similar characteristics are integrated in this design, avoiding data loss and spatial discrepancies. The network extracts and fuses features at different scales, improving understanding of details in complex scenes. This enhances the model's precision.The scale fusion strategy mitigates issues of gradient vanishing and information bottleneck in deep networks, enhancing training efficiency and generalisation. It addresses challenges of insufficient data and overfitting.

C. *Models*

The TKS-UNet network architecture, as depicted in Fig. 1, is a multi-layered design that significantly enhances model accuracy and generalization. The TKS-UNet model, in particular, incorporates three key convolutional layers and a Transformer [23] encoder with the KAN mechanism. These convolutional layers, a fundamental building block in deep learning, effectively extract low-level image features in a progressive, layer-by-layer manner, significantly improving pattern capture and enabling high-dimensional image processing, a critical aspect in computer vision. The Transunet model, another sophisticated model under discussion, employs the aforementioned decoder component, which decodes and reconstructs feature information from the encoder in a manner tailored for image segmentation. The redesigned skip connections in the decoder integrate the three initial valid feature layers from the encoder with increased efficiency through a convolutional feature fusion layer. This strategic integration enhances the model's expressive ability and significantly improves its context-awareness, ensuring accurate image reconstruction. One of the standout features of this model is its multilevel feature fusion, which considerably improves vascular segmentation of fundus images in terms of accuracy and reliability.

III. Technical Details

As part of this study, the raw input images underwent a series of pre-processing stages. The primary objective of these pre-processing steps was to resize the images to a uniform size of 512 pixels by 512 pixels, ensuring uniformity in dimensions for efficient further processing. This step is crucial as it helps maintain aspect ratio while ensuring images are of a standard size efficiently processable by the machine learning model.

Beyond resizing, data enhancement techniques were randomly applied to these pre-processed images to improve various aspects of the model's performance. The application of these techniques improves the model's performance by introducing variability in the training dataset, aiding generalization and enabling better performance when exposed to new data. Moreover, it contributes to more efficient training by exposing the model to diverse representations of the data, facilitating faster learning.

Applying such preprocessing steps before training the model has a larger strategic objective of enhancing model resilience, particularly important for machine learning models deployed to handle real-world scenarios. Real-world scenarios are unpredictable and diverse, and the model must handle variability and perform robustly. Therefore, the aim is to ensure the model is resilient enough to handle variability and perform consistently well in real-world situations.

Moving to the core of the model, the encoder component is critical for downsampling the input image, reducing the spatial dimensionality of the feature map and enabling the extraction of more abstract feature representations from the raw input data. Each individual convolutional block within the encoder consists of a convolutional layer, a batch normalization layer, and a ReLU activation layer. The convolutional layer, combined with the ReLU activation function, synergistically extracts spatial features from the input image, critical for understanding image content.

Additionally, batch normalization and residual connectivity are strategically employed to enhance the training process. Batch normalization aids in normalizing layer output, stabilizing and speeding up training, and reducing internal covariate shift. Meanwhile, residual connectivity helps improve optimization by allowing the model to learn residual functions. These components improve training and optimization efficiency of the model.

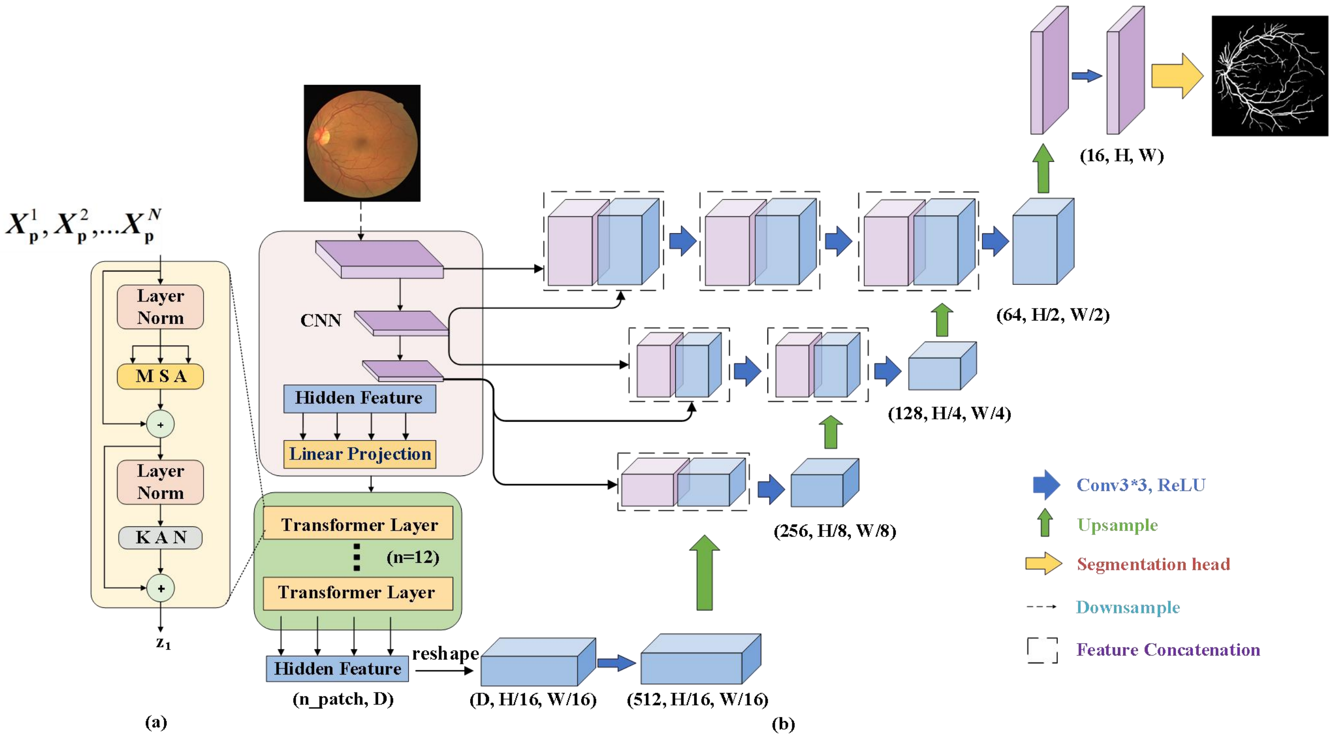

Fig. 1. TKS-UNet Model Structur. (a) Addition of Kan's Transformer layer; (b) Redesign of the TKS-UNet architecture after skip connections.

Applying such preprocessing steps before training the model has a larger strategic objective of enhancing model resilience, particularly important for machine learning models deployed to handle real-world scenarios. Real-world scenarios are unpredictable and diverse, and the model must handle variability and perform robustly. Therefore, the aim is to ensure the model is resilient enough to handle variability and perform consistently well in real-world situations.

Moving to the core of the model, the encoder component is critical for downsampling the input image, reducing the spatial dimensionality of the feature map and enabling the extraction of more abstract feature representations from the raw input data. Each individual convolutional block within the encoder consists of a convolutional layer, a batch normalization layer, and a ReLU activation layer. The convolutional layer, combined with the ReLU activation function, synergistically extracts spatial features from the input image, critical for understanding image content.

Additionally, batch normalization and residual connectivity are strategically employed to enhance the training process. Batch normalization aids in normalizing layer output, stabilizing and speeding up training, and reducing internal covariate shift. Meanwhile, residual connectivity helps improve optimization by allowing the model to learn residual functions. These components improve training and optimization efficiency of the model.

IV. Datasets and Experiments

A. Database Summary

To comprehensively assess the proposed network structure, the method is evaluated on three commonly used datasets: DRIVE [8], CHASE_DB1 [9], and STARE [10]. Key information is summarized in TABLE I.

TABLE I
A SUMMARY OF THE THREE DATASETS

Dataset	Amount	Pixels	Resized	FOV	Observers
DRIVE	40	584× 565	512×512	45°	3
CHASE_DB1	28	999 × 960	512×512	30°	2
STARE	20	605 × 700	512×512	35°	2

a. DRIVE

The DRIVE dataset comprises 40 digital fundus images with a resolution of 584 × 565 pixels, each accompanied by a meticulously hand-labelled vessel segmentation mask providing detailed pixel-level information on vessel status. This data can serve as a reliable ground truth for algorithm training and performance evaluation. The fundus images were captured within a 45-degree field of view, demonstrating the complexity and diversity of the retinal vascular network.

b. CHASE_DB1

In comparison to the DRIVE dataset, the CHASE_DB1 dataset exhibits higher resolution and more intricate detail despite fewer images (28 total). The dataset comprises fundus images with dimensions of 999 × 960 pixels and a field of view (FOV) of 30 mm, accompanied by manually labelled vessel segmentation masks. Furthermore, two independent annotators manually segmented each image, with the first annotator's segmentation serving as the standard reference.

c. STARE

The STARE dataset provides a fundus image dataset with fewer false positives than the DRIVE and CHASE_DB1 datasets due to a region-based detection approach. The STARE dataset contains 20 photographs, each with a resolution of 605 × 700. These images demonstrate that half of the patients have structural blood lesions, while the other half are normal. The complete dataset was manually labelled using a tool developed by Hoover et al. (1994) and visualized with appropriate magnification levels and histogram transformations.

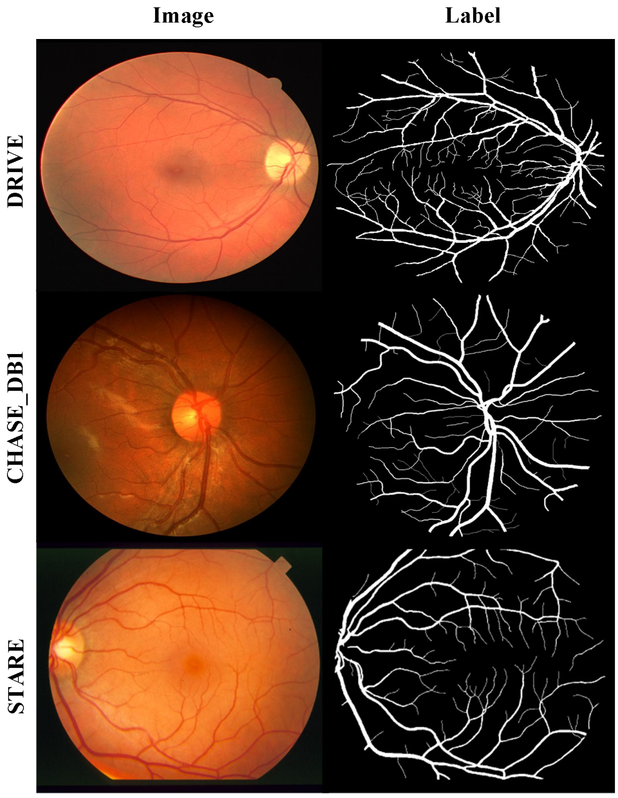

Fig. 2. images and labels for three datasets

B. Evaluation Methods

This paper comprehensively assesses the effectiveness of the TKS-UNet model for retinal blood vessel image segmentation by employing a series of evaluation metrics, including accuracy, F1 score, sensitivity, specificity, and others, to compare the proposed network model with a basic segmentation network model.

Accuracy (Acc): This metric represents the proportion of samples correctly classified by the model across the entire dataset. It is calculated by dividing the number of correctly classified pixels by the total number of pixels, as shown in (1).

$$A_{CC} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Auc（Area Under the Curve）: This metric quantifies the area under the ROC (Receiver Operating Characteristic) curve, serving as a comprehensive measure of classifier performance. It illustrates the relationship between the True Positive Rate and the False Positive Rate across various threshold conditions. By encapsulating these relationships, this metric provides an overall assessment of the model's ability to distinguish between positive and negative samples (see (2)).

$$AUC = 1 - \frac{1}{2}\left(\frac{F_P}{F_P + T_N} + \frac{F_N}{F_N + T_P}\right) \tag{2}$$

F1 Score: The F1 score serves as a comprehensive evaluation metric that integrates two fundamental classification metrics: precision and recall. In the domain of retinal vascular image segmentation, this metric is employed to systematically assess the model's capability to accurately identify vascular regions by balancing two critical performance dimensions (see (3)).

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3}$$

Sensitivity (Se): Also referred to as the true positive rate, sensitivity measures the proportion of positive samples correctly identified by the model out of the total number of actual positive samples. In retinal vascular image segmentation, sensitivity indicates the model's effectiveness in accurately recognizing and detecting vascular regions, capturing and isolating pixels that genuinely belong to the vascular structure (see (4)).

$$S_e = \frac{TP}{TP + FN} \tag{4}$$

Specificity (Sp): Known as the true negative rate, specificity measures the proportion of negative samples (non-vascular regions) correctly identified by the model out of the total number of actual negative samples. In this segmentation task, specificity reflects the model's ability to accurately exclude non-vascular regions, thereby indicating the accuracy with which the model classifies actual non-vascular instances as negative cases (see (5)).

$$S_p = \frac{TN}{TN + FN} \tag{5}$$

Precision (Pre): This metric measures the ratio of true positives (TP) among all positive predictions, including TP and false positives (FP). It evaluates the reliability of the model's positive predictions by quantifying the accuracy of its affirmative classifications (see (6)).

$$\mathrm{Pre} = \frac{TP}{TP + FP} \tag{6}$$

Intersection over Union (IoU): This metric evaluates the overlap between the predicted region and the ground truth region. It is calculated by dividing the area of overlap between the predicted and the actual regions by the area of their union (see (7)).

$$\mathrm{IoU} = \frac{A \cap B}{A \cup B} \tag{7}$$

C. Performance Evaluation

A comparative analysis was conducted using the TKS-UNet model for retinal vessel segmentation across three datasets: DRIVE [8], CHASE_DB1 [9], STARE [10]. The model's robustness and generalizability were comprehensively evaluated for performance. The model was compared with established deep learning architectures and benchmarks in image segmentation.Tables II, III, and IVprovide a comprehensive analysis of the model's performance under different datasets and evaluation metrics.

As evidenced in Table Ⅱ, the performance metrics of the proposed model demonstrate a notable enhancement across all three public datasets. In the DRIVE dataset, the sensitivity (Se) metric reached 85.03%, while the specificity (Sp) metric reached 98.92%. The accuracy (Acc) was 97.65%, while the area under the curve (AUC) reached 99.21%. The F1-score value was 86.09%. These results demonstrate that the model exhibits robust detection performance.

As shown in Table Ⅲ, the model also demonstrated impressive performance in the CHASE_DB1 dataset, with sensitivity, specificity, accuracy, AUC value and F1-score reaching 90.82%, 99.38%, 98.81%, 99.62% and 90.89%, respectively. These results provide further evidence that the model is capable of performing well across a range of datasets.

Furthermore, as illustrated in Table IV, the sensitivity, specificity, accuracy, AUC, and F1 scores of the TKS-UNet model demonstrated notable enhancement, reaching 85.01%, 99.03%, 97.97%, 99.29%, and 86.31%, respectively, in the STARE dataset. In conclusion, the validity and reliability of the proposed model on various public datasets can be seen.

TABLE II
COMPARATIVE ANALYSIS OF OUR MODEL ON THE DRIVE DATASET

Dataset	Model	Se	Sp	Acc	AUC	F1
DRIVE	Att UNet[11]	79.46	97.89	95.64	97.99	82.32
	BCDU-Net[12]	79.84	98.03	95.75	98.11	98.49
	Bio-Net[13]	82.20	98.04	96.09	82.06	98.26
	CTF-Net[14]	78.49	98.13	95.67	97.88	82.41
	CSU-Net[15]	80.71	97.82	95.65	98.01	82.51
	OCE-Net[16]	80.18	98.26	95.81	98.21	83.02
	LDMRes-Net[18]	83.58	98.32	97.02	98.51	83.09
	DA-Res2UNet[21]	81.50	98.56	97.04	98.77	82.77
	Proposed LMBis-net[22]	83.60	98.83	97.08	98.80	83.43
	TKS--UNet	**85.03**	**98.92**	**97.65**	**99.21**	**86.09**

TABLE III
COMPARATIVE ANALYSIS OF OUR MODEL ON THE CHASE_DB1 DATASET

Dataset	Model	Se	Sp	Acc	AUC	F1
CHASE_DB1	Att UNet[11]	80.10	98.04	96.42	98.4	80.12
	BCDU-Net[12]	77.35	98.01	96.18	98.39	79.32
	OCE-Net[16]	81.38	98.24	96.78	98.72	81.96
	LDMRes-Net[18]	85.95	98.88	97.55	98.61	81.94
	DA-Res2UNet[21]	83.18	98.67	97.70	99.12	81.88
	Proposed LMBis-net[22]	86.05	98.96	97.75	98.71	83.54
	TKS-UNet	**90.82**	**99.38**	**98.81**	**99.62**	**90.89**

TABLE IV
COMPARATIVE ANALYSIS OF OUR MODEL ON THE STARE DATASET

Dataset	Model	Se	Sp	Acc	AUC	F1
STARE	Att UNet[11]	77.09	98.48	96.33	97.00	-
	BCDU-Net[12]	78.92	98.16	96.34	98.43	82.30
	CC-Net[19]	80.67	98.16	96.32	98.33	81.36
	OCE-Net[16]	80.12	98.65	96.72	98.76	83.41
	Wave-Net[20]	79.02	98.36	96.41	-	81.40
	G-Net Light[17]	81.70	98.56	97.30	-	81.78
	LDMRes-Net[18]	84.07	98.75	97.64	98.72	84.24
	DA-Res2UNet[21]	82.69	98.85	97.65	98.83	83.96
	Proposed LMBis-net[22]	84.37	98.77	97.69	98.82	84.44
	TKS-UNet	**85.01**	**99.03**	**97.97**	**99.29**	**86.31**

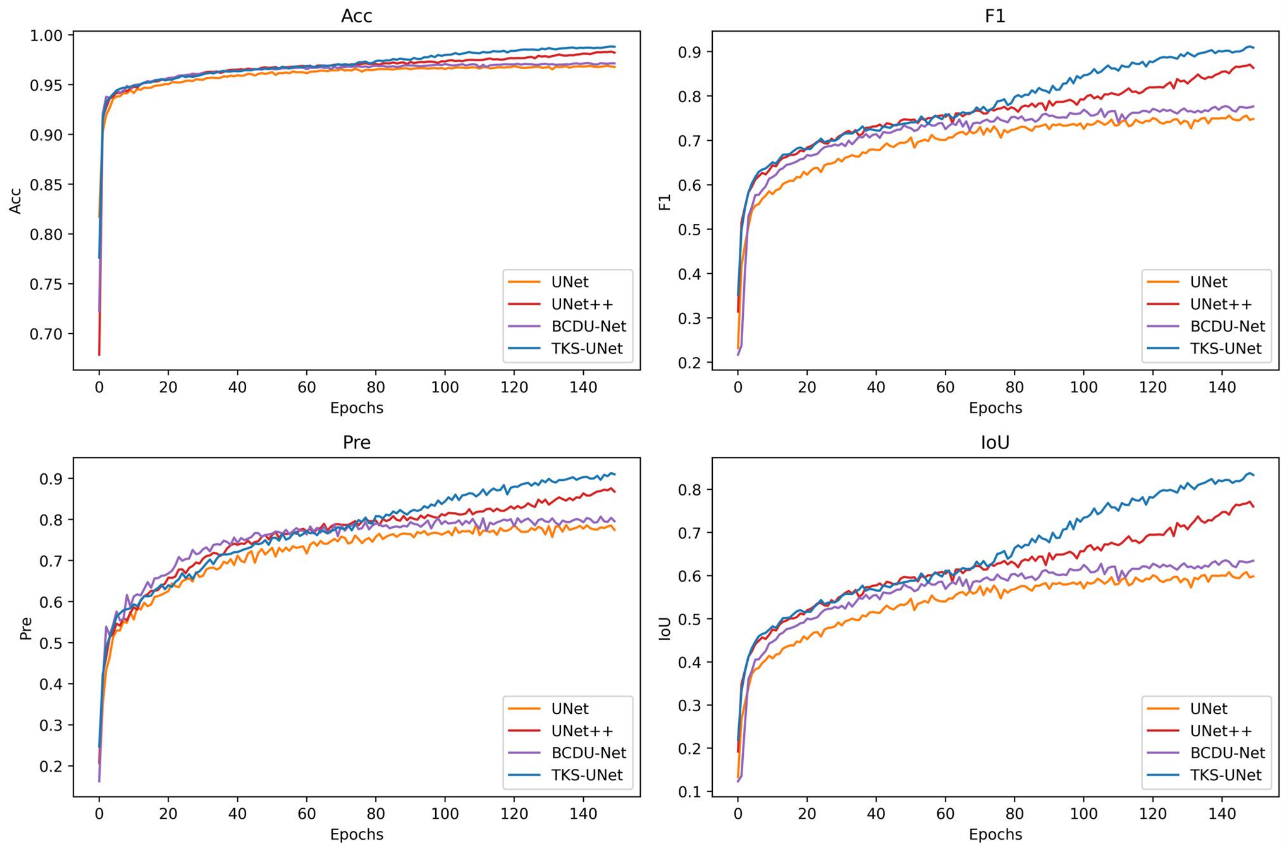

Fig. 3. A comparison of indicator trends between TKS UNet and other models on the CHASE_DB1 dataset

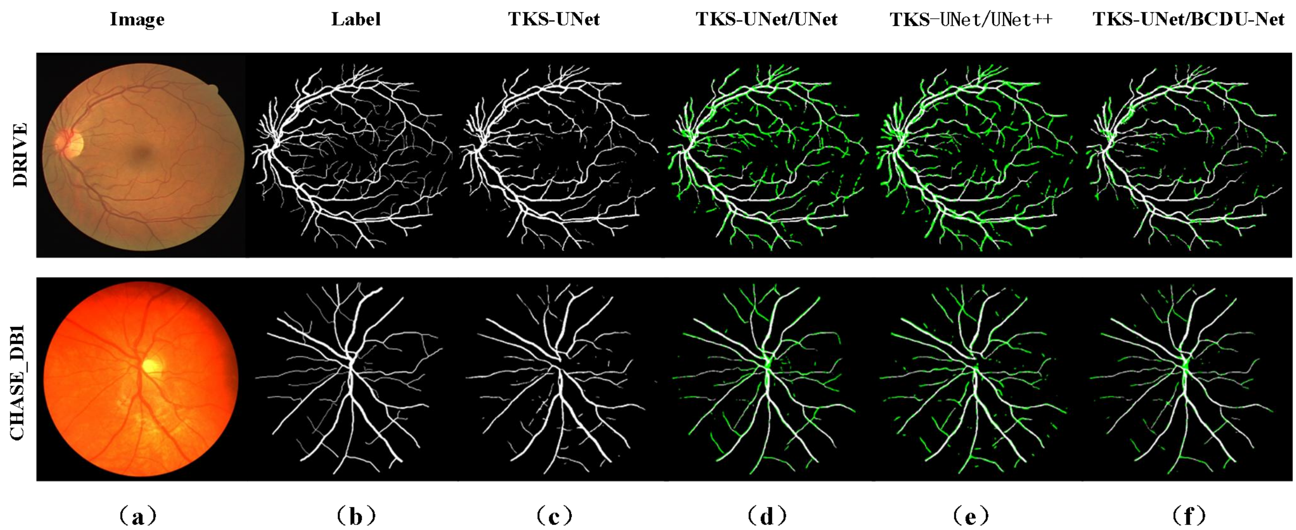

Fig. 4. Retinal vascular segmentation sample image. (a) Original image. (b) Ground truth. (c) is the segmentation prediction graph of TKS-UNnet neural network.(d), (e), and (f) represent the comparison of segmentation results between TKS-UNet and UNet, Unet++, and BCDU-Net, respectively. The green section indicates the segmentation details of our model in comparison to the aforementioned models.

Fig.3 presents a performance comparison between the TKS-Unet model and other similar models on the CHASE_DB1 dataset, specifically focusing on key performance metrics such as accuracy, F1 scores, precision, and intersection over union. Additionally, Fig.4 demonstrates the segmentation results of our proposed retinal segmentation model on both the DRIVE and CHASE_DB1 datasets, along with comparisons to the segmentation results of other models. Through comparative analysis, it is evident that our model exhibits outstanding performance on these two datasets, fully demonstrating its robust capability in accurately segmenting retinal structures.

D. Ablation Experiments

Ablation studies were performed on the CHASE_DB1 dataset to assess the impact of each component on the model's performance. The experiments systematically removed either the KAN module or the redesigned skip connections to isolate their specific effects, with results presented in Table V. The baseline model (TransUNet without modifications) achieved an F1-score of 88.34%, serving as a reference for evaluating improvements.

Incorporating the KAN module alongside traditional skip connections resulted in a significant improvement in the F1-score to 90.67%, with sensitivity improved to 90.14% and AUC to 99.59%. These results indicate that KAN enhances the model's ability to capture complex vascular patterns and generalize from limited data, aligning with its theoretical advantages in interpretability and small-sample learning.

TABLE V
RESULTS OF ABLATION EXPERIMENT

Dataset	Kan	Skip	Se	Sp	Acc	AUC	F1
CHASE_DB1	×	×	87.55	99.26	98.49	99.44	88.34
	√	×	90.14	***99.40***	98.78	99.59	90.67
	×	√	90.66	99.37	98.79	99.62	90.72
	√	√	***90.82***	99.38	***98.81***	***99.62***	***90.89***

Replacing conventional skip connections with the UNet++-inspired hierarchical fusion strategy improved the F1-score to 90.72%, demonstrating the efficacy of multiscale feature integration. Additionally, the model's specificity (99.37%) and accuracy (98.79%) increased, suggesting that the redesigned connections reduce semantic gaps and preserve fine-grained details during upsampling.

The full TKS-UNet architecture, integrating both KAN and redesigned skip connections, achieved the highest F1-score of 90.89%, surpassing all ablated variants. This synergistic effect highlights the reciprocal reinforcement of interpretability from KAN and improved feature fusion.

These findings validate the essential role of both the KAN module and hierarchical skip connections in improving TKS-UNet's performance. The ablation analysis underscores the importance of architectural innovations in optimizing the trade-off between interpretability and segmentation precision, providing empirical evidence for the model's design choices.

V. CONCLUSION

This study presents TKS-UNet, a novel neural network architecture developed for retinal vessel segmentation. TKS-UNet combines the Kolmogorov-Arnold Network (KAN) with a TransUNet framework and redesigns skip connections for effective multiscale feature fusion. This approach aims to overcome the limitations observed in conventional U-Net and TransUNet models related to feature fusion efficiency and interpretability.

The proposed model achieves substantial improvements in segmentation accuracy on three benchmark datasets: DRIVE, CHASE_DB1, and STARE. On the DRIVE dataset, TKS-UNet attained an F1-score of 86.09%, surpassing leading approaches like Att U-Net and BCDU-Net. Additionally, on the CHASE_DB1 and STARE datasets, the model achieved F1-scores of 90.89% and 86.31%, respectively, demonstrating its robustness and generalizability.

The key innovations of TKS-UNet involve employing KAN modules to improve interpretability and feature representation, as well as the redesigned skip connections inspired by UNet++ for hierarchical multiscale feature fusion. These enhancements elevate the model's capacity to capture complex vascular structures and improve transparency, rendering it more suitable for clinical scenarios where interpretability is critical.

Ablation studies have verified the efficacy of both the KAN module and the redesigned skip connections, demonstrating their substantial contributions to the model's performance. The integration of KAN modules enhances the model's ability to learn intricate patterns from limited data, while the hierarchical feature fusion strategy improves retention of fine-grained details and reduces the semantic gap between encoder and decoder layers.

Despite its success, TKS-UNet has limitations, particularly in terms of computational complexity that may hinder its deployment on resource-constrained devices. Future research will focus on optimizing the architecture to minimize computational overhead while preserving performance. Additionally, exploring advanced feature fusion techniques and enhancing interpretability through visualization tools are potential avenues for improvement.

In conclusion, TKS-UNet represents a significant advancement in retinal vessel segmentation by merging high performance with interpretability. Its effectiveness across diverse datasets underscores its clinical potential, contributing to the development of AI-driven diagnostic tools in ophthalmology.

REFERENCES

[1] J. Chen, Y. Lu, Q. Yu, et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv, 2021, doi:10.48550/arxiv.2102.04306.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention, vol. 234, 2015, pp. 234–241.

[3] R. Li, M. Li, J. Li, and Y. Zhou, "Connection Sensitive Attention U-NET for Accurate Retinal Vessel Segmentation," arXiv, 2019, doi:10.48550/arxiv.1903.05558v2.

[4] K. Ren, L. Chang, M. Wan, et al., "An improved U-net based retinal vessel image segmentation method," Heliyon, 2022, doi:10.1016/j.heliyon.2022.e11187.

[5] I. Tolstikhin, N. Houlsby, A. Kolesnikov, et al., "MLP-Mixer: An all-MLP architecture for vision," arXiv, 2021, doi:null.

[6] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov–Arnold networks," arXiv preprint, arXiv:2404.19756v4, 2024.

[7] Z. Zhou, M. R. Siddiquee, N. Tajbakhsh, et al., "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE Trans. Med. Imag., vol. 39, no. 6, pp. 1856–1867, 2020, doi:10.1109/tmi.2019.2959609.

[8] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Trans. Med. Imag., vol. 23, pp. 501–509, 2004.

[9] M. M. Fraz, et al., "An ensemble classification-based approach applied to retinal blood vessel segmentation," IEEE Trans. Biomed. Eng., vol. 59, pp. 2538–2548, 2012.

[10] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Trans. Med. Imag., vol. 19, pp. 203–210, 2000.

[11] O. Oktay, et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint, arXiv:1804.03999, 2018.

[12] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-net with densely connected convolutions," in IEEE International Conference on Computer Vision Workshops, 2019.

[13] T. Xiang, et al., "BiO-Net: Learning recurrent bi-directional connections for encoder-decoder architecture," in Medical Image Computing and Computer-Assisted Intervention, pp. 74–84, 2020.
[14] K. Wang, X. Zhang, S. Huang, Q. Wang, and F. Chen, "CTF-Net: Retinal vessel segmentation via deep coarse-to-fine supervision," in IEEE International Symposium on Biomedical Imaging, 2020, pp. 1237–1241.
[15] B. Wang, et al., "CSU-Net: A context spatial U-Net for accurate blood vessel segmentation in fundus images," IEEE J. Biomed. Health Inform., vol. 25, pp. 1128 – 1138, 2020.
[16] X. Wei, K. Yang, D. Bzdok, and Y. Li, "Orientation and context entangled network for retinal vessel segmentation," Expert Systems with Applications, vol. 217, pp. 119443, 2023.
[17] S. Iqbal, S. S. Naqvi, H. A. Khan, A. Saadat, and T. M. Khan, "G-Net light: A lightweight modified Google-Net for retinal vessel segmentation," Photonics, vol. 9, p. 923, 2022.
[18] S. Iqbal, et al., "LDMRes-Net: A lightweight neural network for efficient medical image segmentation on IoT and edge devices," IEEE J. Biomed. Health Inf., 2023.
[19] S. Feng, Z. Zhuo, D. Pan, and Q. Tian, "CcNet: A cross-connected convolutional network for segmenting retinal vessels using multi-scale features," Neurocomputing, vol. 392, pp. 268–276, 2020.
[20] Y. Liu, J. Shen, L. Yang, H. Yu, and G. Bian, "Wave-Net: A lightweight deep network for retinal vessel segmentation from fundus images," Comput. Biol. Med., p. 106341, 2022.
[21] R. Liu, T. Wang, X. Zhang, et al., "DA-Res2UNet: Explainable blood vessel segmentation from fundus images," Alexandria Engineering Journal, vol. 68, pp. 539–549, 2023, doi:10.1016/j.aej.2023.01.049.
[22] M. Abbasi, S. Iqbal, A. Naveed, et al., "LMBiS-Net: A lightweight multipath bidirectional skip connection based CNN for retinal blood vessel segmentation," arXiv, 2023, doi:10.48550/arxiv.2309.04968.
[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need (Journal Paper style)," Neural Information Processing Systems (NIPS), vol. 30, 2017, pp. 6000–6010.

ZHENYU CHEN was born in Shanxi Province, P. R. China, received the B.S. degree in Communication Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2026.

He is currently pursuing the M.S. degree in Electronic Information with University of Science and Technology Liaoning, Anshan, P. R. China. He research interest is artificial intelligence.

ZIWEI ZHOU (1974-), male, from Anshan, Liaoning, associate professor, master's supervisor, received bachelor's and master's degrees from Liaoning University of Science and Technology in 1997 and 2007, respectively; Ph.D. from Harbin Institute of Technology in 2013, with main research directions in artificial intelligence, 3D vision, deep learning and robotic system research. Email: 381431970@qq.com.