KGS-UNet: KAN-Gated Skip Connections for Small-Structure Medical Image Segmentation with Limited Data

1st Chaolin Wu
Jinan University
Guangzhou, China
https://orcid.org/0009-0004-3540-6104

2nd Shun Long
Jinan University
Guangzhou, China
https://orcid.org/0000-0003-2782-0781

***Abstract*—Accurate segmentation of small anatomical structures remains challenging for U-shaped networks because skip connections often fuse encoder and decoder features without explicit selection. We propose KGS-UNet, which repurposes Kolmogorov–Arnold Networks (KANs) as learnable gates at skip interfaces. The KAN-Gated Skip (KGS) module performs pixel-wise routing using spline-parameterized gates with noise-aware lower bounds to suppress inconsistent encoder activations while preserving decoding cues. In addition, the AttnDown module integrates KAN-driven channel attention and spatial gating to better retain thin structures during downsampling. Benchmarked against over 70 mainstream U-Net variants under a unified training framework, KGS-UNet secures the top rank on the challenging, small-scale DRIVE and CHASEDB1 datasets. Specifically, it surpasses competing mainstream models by margins of 1.5% in both IoU and F1 scores. Furthermore, compared with existing KAN-based U-Net variants that primarily utilize KANs as activation replacements, KGS-UNet achieves a substantial improvement of over 6 percentage points in IoU on DRIVE. These results suggest that spline-based gating at skip interfaces is an effective design choice for medical segmentation. Anonymous repository for reproducibility:https://anonymous.4open.science/r/KGS-UNet-C0EE**

***Index Terms*—Medical image segmentation, U-Net, Kolmogorov–Arnold networks, skip connections, retinal vessel segmentation, skin lesion segmentation.**

I. Introduction

Accurate delineation of anatomical structures from medical images is essential for computer-aided diagnosis and treatment planning. In ophthalmology, the topology of retinal vessels is vitally related to diabetic retinopathy, hypertension, and other systemic diseases [1] [2]; in dermatology, the extent and boundary of pigmented lesions has a significant impact on melanoma risk assessment. These tasks share two practical difficulties. First, the targets of interest are often either very thin (e.g., terminal retinal vessels only a few pixels wide) or have low-contrast, irregular boundaries (e.g., skin lesions partially blending into surrounding skin). Secondly, high-quality pixel-wise annotations are expensive and time-consuming to obtain, resulting in public datasets such as DRIVE [3], CHASEDB1 [4], and ISIC 2018 [5] contain of at most a few dozen to a few hundred labeled images. Inevitably,a segmentation model must extract weak structural cues from noisy backgrounds and generalize across different imaging modalities with limited supervision.

U-Net [6] and its variants have become the de facto standard for medical image segmentation due to their encoder–decoder topology with lateral skip connections [7]. The encoder gradually reduces spatial resolution while enriching semantic information, and the decoder progressively restores resolution using encoder features as high-resolution guidance. In most implementations, skip connections are achieved via concatenation or linear addition of encoder and decoder feature maps. While simple and effective in many scenarios, this strategy implicitly assumes that all high-resolution encoder features are equally useful. In small-data medical settings, shallow encoder features may contain substantial noise and modality-specific artefacts. Without an explicit selection mechanism, the decoder may receive a mixture of true structural details and irrelevant high-frequency patterns, making it difficult to distinguish, for example a thin vessel from a noise ridge or a lesion border from background texture. This may lead to over-segmentation or under-segmentation, or broken topology, especially for very fine structures.

Several works have attempted to address these issues by augmenting U-Net [6] with attention mechanisms. Attention U-Net [8] and related models use convolutional gating to suppress irrelevant encoder responses before fusion. Transformer-based hybrids such as TransUNet [9] and Swin-UNETR [10] introduce self-attention to capture long-range dependencies. Despite improving performance on certain benchmarks, these approaches also bring challenges. Convolutional attention typically relies on shallow networks with limited nonlinearity, and its behavior may be hard to tune across very different modalities. Transformer-style self-attention has higher capacity and quadratic complexity, and tends to require larger datasets and careful optimization [11] [12]; on small medical datasets with high-resolution inputs, it may suffer from unstable training or overfitting.

To address these limitations while maintaining strong nonlinear modeling capabilities without the heavy computational burden of self-attention, researchers begun exploring alternative neural representations. Recently, Kolmogorov–Arnold

Networks (KANs) have been proposed as an alternative to conventional multilayer perceptrons [13]. Instead of using fixed activation functions, KANs place learnable spline functions on edges, in order to enable flexible approximation of one-dimensional nonlinear mappings with relatively few parameters. This property is attractive for modeling intensity transitions and other scalar relationships in medical images. Early attempts such as U-KAN [14] and ResU-KAN [15] integrated KAN blocks into encoder and decoder stages, and reported improvements over purely convolutional baselines on several segmentation tasks. However, most of these architectures keep skip connections and downsampling operations unchanged; KAN is mainly used inside feature extraction blocks, while cross-scale feature fusion—arguably the most fragile part of U-shaped networks under limited data—remains governed by simple linear operations.

In this work, we explore a different use of KAN. Instead of only replacing internal MLPs, we investigate whether KANs can serve as explicit, learnable decision functions that regulate how information flows between encoder and decoder. To this end, we propose *KGS-U-Net*, a UNet–style architecture in which KAN-based modules are inserted at key feature transfer interfaces. A KAN-Gated Skip (KGS) module replaces plain concatenation with pixel-wise gating: encoder and decoder features are first projected into a shared space, then combined into a compact interaction descriptor, before being passed through a spline-parameterized gate with a learnable lower bound. This design enables noise-aware fusion that can suppress inconsistent encoder activations while preserving a minimum level of decoder information, which is important for retaining faint structures in low-contrast regions. Complementing this, an *AttnDown* module couples strided convolution with KAN-based channel and spatial attention, so that downsampling not only reduces resolution but also selectively retains features that likely correspond to vessels or lesion boundaries. At the bottleneck, we provide an optional multi-branch module that aggregates semantic, multi-scale, and contextual cues and fuses them through a lightweight attention mechanism; this component can be enabled or disabled,depending on computational budget.

Our contributions can be summarized as follows:

- **KAN-gated skip connections.** We introduce a KAN-Gated Skip (KGS) module, which, to the best of our knowledge, is among the first attempts to embed KAN directly into U-Net [6] skip connections. By treating encoder–decoder interaction as a pixel-wise gating problem parameterized by learnable spline functions, KGS provides a flexible yet compact mechanism for noise-aware cross-scale fusion.
- **Structure-aware downsampling and optional bottleneck fusion.** We propose an AttnDown module that integrates KAN-based channel and spatial attention into the downsampling path, aiming to preserve small anatomical structures during resolution reduction. In addition, we offer an optional bottleneck module for multi-scale and graph-enhanced context aggregation, without compromising the overall simplicity of the encoder–decoder framework.
- **Cross-modality evaluation under limited data.** Using a KGS-UNet configuration and similar training settings, we conducted a unified evaluation on DRIVE [3], CHASEDB1 [4], and ISIC 2018 [5], which cover thin line-like structures and area-based lesions with distinct appearance statistics. The model demonstrated robust results across all three datasets, particularly on the retinal benchmarks. We regard these findings as an initial empirical indication that KAN-based gating is an effective direction for small-structure medical image segmentation.

II. Method

A. Overview

An overview of the architecture and the three main modules is illustrated in **Fig. 1**. KGS-UNet follows a standard encoder–decoder topology with four resolution levels, but replaces the passive information flow of vanilla U-Net [6] with two active mechanisms: (i) *KAN-Gated Skip* (KGS), which reformulates each skip connection as a learnable routing gate; and (ii) *AttnDown*, which couples strided downsampling with KAN-driven channel and spatial attention. At the bottleneck, we provide an optional context module that aggregates multi-scale and graph-enhanced cues via a lightweight scheduler, while preserving the overall U-shaped skeleton. The implementation employs depthwise separable ConvBlocks with GroupNorm and SiLU [16]–[18], bilinear upsampling refinement, and a standard 1×1 prediction head. Deep supervision and class-balance reweighting are employed as optional training stabilizers (Sec. II-E) [19] [20].

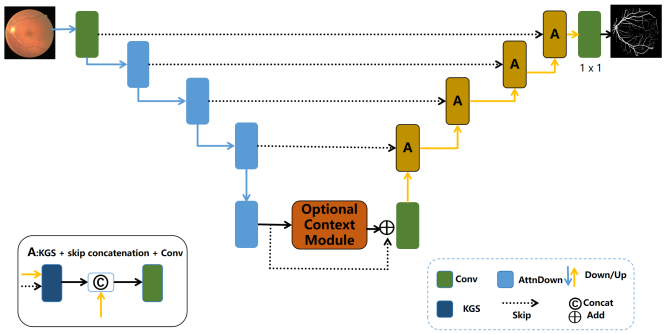

Fig. 1. Overall architecture of **KGS-UNet**. The encoder–decoder backbone is augmented by (i) KAN-Gated Skip (KGS) and (ii) AttnDown; an optional bottleneck aggregates multi-scale context.

B. KAN-Gated Skip (KGS)

instead of fusing encoder- and decoder- features without explicit selection as the traditional skip connections do, the KGS module (**Fig. 2**) introduces a gating mechanism on the encoder stream. Given encoder features $\mathbf{E} \in \mathbb{R}^{C\times H\times W}$ and decoder features $\mathbf{D}$ at the same scale, we first project them into a shared d-dimensional subspace via 1×1 convolutions to

obtain $\theta(\mathbf{E})$ and $\phi(\mathbf{D})$. We then construct an interaction token $\mathbf{Z}$ that explicitly captures feature alignment and discrepancies:

$$\mathbf{Z} = \big[\theta(\mathbf{E}),\ \phi(\mathbf{D}),\ |\theta(\mathbf{E}) - \phi(\mathbf{D})|,\ \theta(\mathbf{E}) \odot \phi(\mathbf{D})\big]. \quad (1)$$

This token is then processed by a mixing convolution and dropout to form a compact descriptor. A spline-parameterized KAN layer then generates a raw pixel-wise gate $\hat{\mathbf{a}} = \sigma(\mathrm{KAN}(\mathbf{Z}))$.

To reduce the risk of completely shutting off skip signals in noisy or low-contrast regions, a parallel branch estimates a noise map and produces a dynamic floor $\mathbf{f}$, which serves as a small positive lower bound. The final rectified gate $\mathbf{a}$ modulates the encoder features:

$$\mathbf{a} = \mathbf{f} + (1 - \mathbf{f}) \cdot \hat{\mathbf{a}}, \qquad \mathbf{S} = \mathbf{E} \odot \mathbf{a}. \quad (2)$$

The gated signal $\mathbf{S}$ is concatenated with $\mathbf{D}$ and passed to the subsequent decoding layers. We add a monotonicity regularizer that encourages the gate to open wider when the semantic discrepancy $|\theta(\mathbf{E}) - \phi(\mathbf{D})|$ is small; in practice, this term is kept modest and mainly acts as a soft prior.

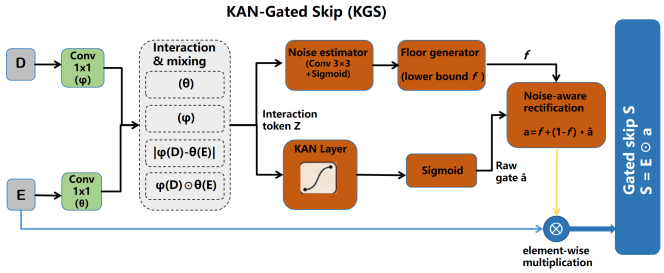

Fig. 2. KAN-Gated Skip (KGS) module. KAN acts as a pixel-wise gate with a learnable floor for noise-aware fusion.

C. Structure-Aware downsampling (AttnDown)

While KGS optimizes feature fusion across scales, preserving fine structural details during the encoding phase is equally critical. downsampling is another stage where fine anatomical structures may be lost. The general AttnDown pipeline is illustrated in **Fig. 3**. To make this step more selective, *AttnDown* replaces standard pooling with a three-step process. First, a stride-2 3×3 convolution reduces the spatial feature dimensions. Second, a *KANSE* (KAN-based Squeeze-and-Excitation [21]) module adaptively recalibrates channel-wise features by mapping global average pooled feature descriptors through two stacked KAN layers, followed by a nonlinear activation function and a sigmoid function. Third, a spatial gating branch predicts a 1-channel saliency mask $\mathbf{g}$ via a 1×1 conv and a spline-based layer with sigmoid activation [22]. As in KGS, this mask is rectified by a learnable lower bound, preventing the model from discarding all responses at any location during downsampling. The final output $\mathbf{Y}$ selectively retains discriminative cues while suppressing background:

$$\mathbf{Y} = \mathrm{KANSE}\big(\mathrm{Conv}_{s=2}(\mathbf{X})\big) \odot \mathbf{g}. \quad (3)$$

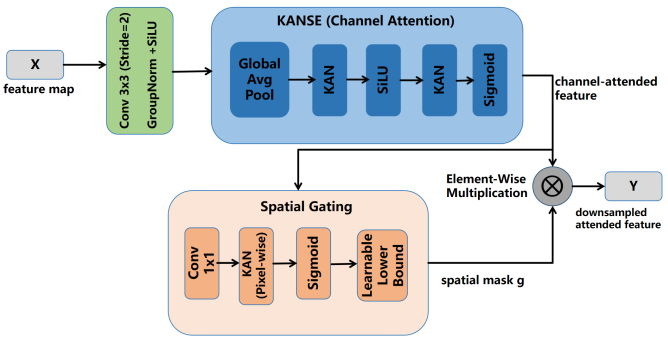

Fig. 3. AttnDown module. Strided convolution is coupled with KAN-based channel attention (KANSE) and spatial gating.

D. Optional Context Module

At the lowest resolution, we introduce an optional context module to aggregate global semantic cues without altering the U-shaped backbone. To balance performance and efficiency, this module fuses three lightweight branches via a learnable attention scheduler: (i) **LiteSPP** for multi-scale pyramid pooling [23]; (ii) a **graph-enhanced** branch that applies GATv2 to model local topology along vessel pathways [24]; and (iii) a **KAN-based context block (KCB)** for non-linear channel recalibration. The module is plug-and-play and can be disabled under tighter computational budgets. Ablation results indicate that it provides a complementary but modest improvement, and architectural details are given in the supplementary material.

E. Implementation Notes

The proposed implementation prioritizes stability on small datasets. We use GroupNorm, which is independent of batch size, and depthwise separable convolutions to reduce parameters [16]–[18]. KAN layers follow a standard implementation and share grids across samples. All gating mechanisms are bounded by sigmoid functions with learnable, noise-aware floors, which helps maintain stable gradients throughout the network without introducing strong inductive biases beyond those described above.

III. Experiments

A. Experimental Setup

This study is built upon the open-source medical image segmentation benchmark U-Bench [25], which provides a unified experimental protocol and ensures fair and reproducible comparisons. In brief, inputs are resized to a fixed resolution and normalized, and lightweight geometric augmentations are applied during training (but not during validation); dataloaders use standard shuffled mini-batches for training and deterministic seeding for reproducibility. The benchmark offers a standardized training framework, 28 cross-modality datasets, as well as the pretrained weights for most U-Net variants, including CNN-, Transformer-, and Mamba-based architectures [26], among others. One of its prime values lies in normalizing experimental pipelines and reducing evaluation bias.

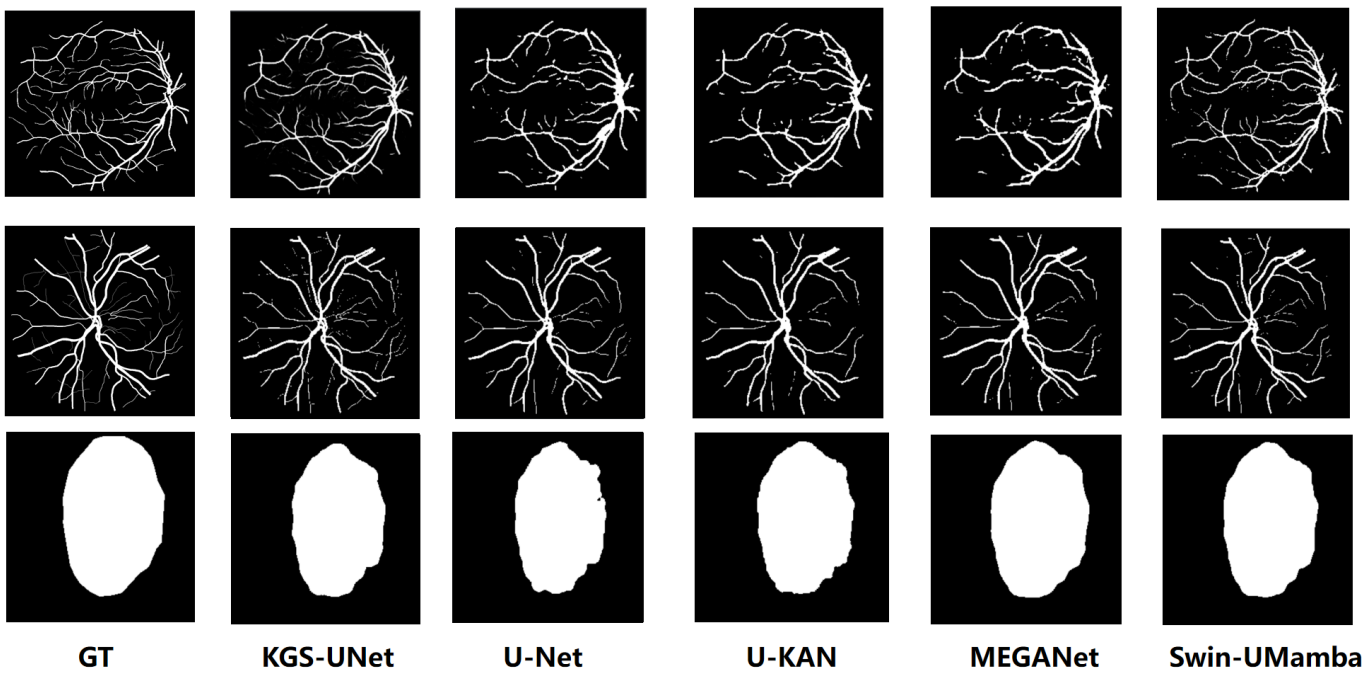

Fig. 4. Qualitative comparison on three datasets: DRIVE/CHASEDB1 (retinal vessels) and ISIC 2018 (skin lesion). KGS-UNet better preserves thin vessel connectivity and produces cleaner lesion boundaries.

We adopted three core datasets from U-Bench, namely DRIVE [3] (20 training, 4 validation, 16 test images), CHASEDB1 [4] (14 training, 3 validation, 11 test images), and ISIC 2018 [5] (2,594 training, 649 validation, 1,000 test images). We followed the default U-Bench training pipeline [25] for preprocessing, optimization, and augmentation, replacing only the segmentation network with our KGS-UNet. Unless otherwise specified, we selected the checkpoint with the best validation performance for reporting test results. All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU with 24GB memory.

B. Visualized segmentation results

Fig. 4 aligns well with the quantitative results in Table I. On DRIVE and CHASEDB1, baselines often miss thin terminal vessels or yield fragmented branches, whereas KGS-UNet maintains more continuous vessel trees with fewer spurious responses. On ISIC 2018, KGS-UNet produces contours that adhere strictly to lesion margins with reduced boundary leakage, indicating improved robustness across heterogeneous segmentation scenarios.

C. Quantitative Comparison with State-of-the-Art Methods

We benchmarked KGS-UNet against 70 models from the U-Bench repository, covering CNNs, Transformers, Mamba-based architectures [26], and KAN-based networks [15]. Table I highlights representative methods, while full results are available in the anonymous GitHub repository.

Topological Versatility vs. Task-Specific Bias. A key observation from our benchmark is the trade-off between segmenting thin vessels (DRIVE / CHASEDB1) and large lesions (ISIC 2018). Strong baselines often exhibit task-specific bias; for instance, MEGANet [27] ranked 1st on ISIC 2018 but dropped to Rank 47 on DRIVE, and Swin-UMamba [29] ranked 2nd on ISIC 2018 but only 11th on CHASEDB1. In contrast, KGS-UNet ranked **1st** on both retinal datasets (DRIVE: 63.75% IoU; CHASEDB1: 84.45% IoU) and maintained a competitive Rank 3 on ISIC 2018, trailing the best model by only 0.04%. The experimental results demonstrate that KGS-UNet is consistently ranked in the top five across these structurally distinct modalities, suggesting that KAN-gated connections adapt to both high-frequency edges and region consistency without architectural tuning.

Comparison with KAN and Mamba Variants. Comparisons with KAN-based models further validated the effectiveness of our design. KGS-UNet outperformed U-KAN [14] and ResU-KAN [15], which use KANs mainly as activation-function replacements. On DRIVE, KGS-UNet increased IoU by 6.03 percentage points over U-KAN (63.75% vs. 57.72%) and by 5.92 percentage points over ResU-KAN (63.75% vs. 57.83%). These results suggest that using KANs as learnable cross-scale gates at skip interfaces is more effective than placing them only inside feature-extraction blocks. Moreover, under the same training protocol, KGS-UNet achieved higher IoU on fine-structure segmentation than heavier models such as TransUNet [9] and Swin-UMamba [29], while retaining a standard U-shaped depth and without additional large-scale pre-training. This indicates that precise pixel-wise gating can be a competitive alternative to global attention on small public

TABLE I
COMPARISON OF KGS-UNET WITH REPRESENTATIVE METHODS ON THREE DATASETS. BEST RESULTS ARE IN **BOLD**, SECOND BEST ARE UNDERLINED.

Method	**DRIVE [3]** (Retinal)			**CHASEDB1 [4]** (Retinal)			**ISIC 2018 [5]** (Skin)		
	Rank	IoU(%)	F1(%)	Rank	IoU(%)	F1(%)	Rank	IoU(%)	F1(%)
KGS-UNet (Ours)	**1**	**63.75**	**77.97**	**1**	**84.45**	**90.87**	3	84.46	90.79
MT_UNet [27]	2	63.21	77.43	8	82.00	89.38	44	82.84	89.62
UTNet [28]	4	63.17	77.40	13	80.83	88.84	21	83.56	90.10
Swin_umamba [29]	6	62.75	77.08	11	81.66	89.07	2	84.49	90.82
MEGANet [30]	47	52.76	69.04	37	73.27	84.36	**1**	**84.50**	**90.85**
AC_MambaSeg [31]	44	54.28	70.34	18	78.72	87.29	30	83.17	89.75
U_Net [6]	12	61.82	76.37	2	84.07	90.48	48	82.78	89.59
AttU_Net [8]	10	62.28	76.73	9	81.89	88.70	53	82.60	89.44
UNet3plus [32]	7	62.54	76.92	3	83.69	90.33	45	82.80	89.61
TransUNet [9]	23	59.76	74.78	15	79.86	88.09	25	83.51	90.06
U_KAN [14]	35	57.72	73.15	27	75.80	85.71	51	82.73	89.42
ResU_KAN [15]	34	57.83	73.26	26	76.11	85.85	26	83.36	89.94

medical segmentation datasets.

D. Ablation Studies

We performed ablation studies on the DRIVE [3] dataset to systematically analyze the contribution of the main components in KGS-UNet and to verify the design decisions within the KGS and AttnDown modules. Unless otherwise stated, all variants were trained under the same protocol as in Sec. III-D1.

1) Impact of Main Components: Table II evaluates the effect of individual and combined modules related to the U-Net [6] backbone. Starting from the baseline (61.82% IoU, 76.37% F1), introducing KGS on the skip connections yielded a clear improvement to 62.98% IoU, indicating that explicit gating of encoder features is beneficial even without other modifications. Similarly, applying AttnDown alone brought a gain over the baseline (62.61% IoU), showing that structure-aware downsampling is also helpful. Combining both KGS and AttnDown further increased IoU to 63.55% and F1 to 77.85%. Finally, the bottleneck context module provided an additional but modest boost, leading to the full KGS-UNet with 63.75% IoU and 77.97% F1. These results suggested that the primary performance gains stem from the proposed skip gating and attention mechanisms, with the bottleneck playing a complementary role.

2) Design of the KGS Module: We next investigated the internal mechanisms of the KAN-Gated Skip (KGS) module while keeping the rest of the network unchanged (Table III). Using plain concatenation corresponded to the U-Net [6] baseline (61.82% IoU). Replacing concatenation with an MLP-based gate improved the score to 62.85%, showing that even a simple learnable gate is helpful. When we used the full interaction token $\mathbf{Z}$ but removed its cross-term components ('w/o interaction", only concatenating $\theta(\mathbf{E})$ and $\phi(\mathbf{D})$), the IoU was 63.02%, slightly below the full KGS design (63.75%), which suggests that explicitly encoding agreement and disagreement between encoder and decoder features brought additional benefits. Furthermore, removing the noise-aware floor and relying solely on a sigmoid gate led to a noticeable degradation (61.50% IoU), indicating that the dynamic lower bound plays an important role in preventing overly suppressed skip signals and stabilizing the gating behaviour in low-contrast regions.

3) Channel vs. Spatial Branches in AttnDown: Finally, we decoupled the channel and spatial branches of AttnDown (Table IV). Replacing AttnDown with a plain strided convolution yielded 62.95% IoU. Adding only the channel branch (KANSE) improved IoU to 63.25%, while using only the spatial gating branch achieved 63.30%. The full AttnDown with both branches obtained the best result (63.75% IoU), indicating that global channel recalibration and pixel-wise spatial gating are complementary for preserving thin structures during resolution reduction.

TABLE II
ABLATION OF MAIN COMPONENTS ON THE DRIVE DATASET. MODULES ARE PROGRESSIVELY ADDED TO THE U-NET BASELINE.

Variant	IoU(%)	F1(%)
U-Net baseline	61.82	76.37
U-Net + KGS (skip gating)	62.98	77.27
U-Net + AttnDown (downsampling)	62.61	76.99
U-Net + KGS + AttnDown	63.55	77.85
Full KGS-UNet (+ bottleneck)	**63.75**	**77.97**

TABLE III
ABLATION ON THE KGS DESIGN ON DRIVE. ALL VARIANTS SHARE THE SAME BACKBONE AND ATTNDOWN; ONLY THE SKIP STRATEGY DIFFERS.

KGS variant	IoU(%)	F1(%)
Plain concat skip (no gate)	61.82	76.37
Gate w/ MLP (no KAN)	62.85	77.14
KGS w/o interaction token $\mathbf{Z}$	63.02	77.27
KGS w/o noise-aware floor	61.50	76.15
Full KGS (ours)	**63.75**	**77.97**

IV. Conclusion

In this work, we have proposed KGS-UNet, a U-shaped architecture that repurposes Kolmogorov–Arnold Networks as active flow controllers rather than mere internal activation replacements. By redesigning the critical interfaces of the U-Net [6] backbone, specifically through the KAN-Gated Skip (KGS) module and the AttnDown module, we explicitly enable the network to filter encoder noise and preserve fine anatomical structures under limited supervision.

Extensive evaluations within the U-Bench framework demonstrate that KGS-UNet achieved substantial performance on retinal vessel segmentation and comparable results on skin lesion segmentation. The model consistently outperformed existing KAN-based baselines such as U-KAN and shows more balanced cross-modality behaviour than several task-specific specialists. These findings suggest that deploying spline-based nonlinearity at cross-scale interfaces can be an effective strategy for small-structure segmentation, beyond using KANs only for feature extraction. In future work, we plan to study parameter–efficiency trade-offs in greater depth, explore lighter variants for deployment on resource-limited devices, and extend this gating mechanism to 3D volumetric segmentation and multi-task clinical workflows.

TABLE IV
DECOUPLING THE ATTNDOWN MODULE ON DRIVE. "KANSE" AND "SPATIAL" DENOTE THE CHANNEL-ATTENTION AND SPATIAL-GATING BRANCHES.

Variant	KANSE	Spatial	IoU(%)	F1(%)
Strided Conv (no attention)	×	×	62.95	77.21
+ Channel only	✓	×	63.25	77.48
+ Spatial only	×	✓	63.30	77.56
Full AttnDown	✓	✓	**63.75**	**77.97**

References

[1] C. Y. Cheung, M. K. Ikram, C. Sabanayagam, and T. Y. Wong, "Retinal microvasculature as a model to study the manifestations of hypertension," Nat. Rev. Cardiol., 2012.

[2] S. Liew et al., "Retinal Microvasculature as a Model to Study the Manifestations of Hypertension," Hypertension, 2012.

[3] J. J. Staal, M. D. Abramoff, M. Niemeijer, et al. "Ridge-based vessel ' segmentation in color images of the retina." IEEE Trans. Med. Imaging (TMI), 2004.

[4] F. G. Venhuizen, B. van Ginneken, M. D. Abramoff, et al. ' "CHASE DB1: A retinal image database for evaluation of segmentation algorithms." Proc. Int. Symp. Biomed. Imaging (ISBI), 2013.

[5] N. C. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2018 ISBI, hosted by ISIC," in *Proc. ISBI*, 2018.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. MICCAI, 2015.

[7] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.

[8] O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," 2018.

[9] J. Chen et al., "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," 2021.

[10] A. Hatamizadeh et al., "Swin UNETR: Transformers for 3D Medical Image Segmentation," 2022.

[11] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," ICLR 2021.

[12] A. Vaswani et al., "Attention Is All You Need," NeurIPS 2017.

[13] Z. Liu, Y. Li, C. He, L. Ni, G. Karypis, and D. Koutra, "Kolmogorov–Arnold Networks," ICLR 2024.

[14] Y. Li et al., "U-KAN: Kolmogorov-Arnold Networks Make Strong Backbones for U-Shaped Medical Segmentation," 2024.

[15] A. Waleed, H. Huang, and Z. Wu, "ResU-KAN: residual Kolmogorov–Arnold networks for medical image segmentation," 2024.

[16] Y. Wu and K. He, "Group Normalization," ECCV 2018.

[17] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions (Swish/SiLU)," 2017.

[18] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[19] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised Nets," in Proc. AISTATS, 2015.

[20] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," CVPR 2019.

[21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," CVPR 2018.

[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," ECCV 2018.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in Proc. European Conf. Comput. Vis. (ECCV), 2014, pp. 346–361.

[24] G. Brody, Y. Alon, and E. Yahav, "How Attentive are Graph Attention Networks? (GATv2)," ICLR 2022.

[25] F. Tang et al., "U-Bench: A Comprehensive Understanding of U-Net through 100-Variant Benchmarking," arXiv preprint arXiv:2510.07041, 2025.

[26] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," arXiv preprint arXiv:2312.00752, 2023.

[27] H. Wang et al., "Mixed Transformer U-Net for Medical Image Segmentation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1091–1095.

[28] Gao, M. Zhou, and D. N. Metaxas, "UTNet: A Hybrid Transformer Architecture for Medical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2021, pp. 61–71.

[29] J. Liu et al., "Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining," arXiv preprint arXiv:2402.03302, 2024.

[30] N.-T. Bui, D.-H. Hoang, Q.-T. Nguyen, M.-T. Tran, and N. Le, "MEGANet: Multi-scale edge-guided attention network for weak boundary polyp segmentation," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2024.

[31] V.-T. Nguyen, V.-T. Pham, and T.-T. Tran, "AC-MambaSeg: An Adaptive Convolution and Mamba-based Architecture for Enhanced Skin Lesion Segmentation," in International Conference on Green Technology and Sustainable Development (GTSD), 2024, pp. 13–26.

[32] H. Huang et al., "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 1055–1059.