KAN-AINet: Kolmogorov-Arnold Network with Adaptive Illumination Modulation for Generalizable Polyp Segmentation

Watcharapong Timklaypachara[1,2], Monrada Chiewhawan[1,2], Abhiraj Pudhota[3], Stephanie Marie Aguilera[5], Sandeep Bodduluri[3,4], and Titipat Achakulvisut[1]

[1] Department of Biomedical Engineering, Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand
[2] Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[3] Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, The University of Alabama at Birmingham, Birmingham, AL, USA
[4] Department of Electrical and Computer Engineering, The University of Alabama at Birmingham, Birmingham, AL, USA
[5] Department of Biomedical Engineering, The University of Alabama at Birmingham, Birmingham, AL, USA

Abstract. Automated polyp segmentation is critical for early colorectal cancer detection, yet performance often degrades across clinical centers due to illumination variability from heterogeneous endoscopic systems and imaging protocols. We propose KAN-AINet, a ConvNeXt–U-Net architecture that integrates Kolmogorov–Arnold Networks (KAN) for adaptive feature modulation through learnable nonlinear activation functions, improving robustness to diverse lighting conditions. KAN-AINet includes two modules: (1) a KAN-based Illumination Modulation Module (KAN-IMM) that adaptively scales encoder features according to brightness patterns, and (2) a KAN-based Boundary Attention Module (KAN-BAM) that refines decoder features for improved structural consistency. Trained on Kvasir-SEG, SUN-SEG, and PolypGen (C1–C5), and evaluated on Kvasir-Sessile, CVC-ColonDB, ETIS-LaribPolypDB, and PolypGen-C6. KAN-AINet outperforms ESPNet and nnU-Net on the combined unseen datasets, improving mDice by 2.20% and 42.27%, and mIoU by 3.05% and 55.51%, respectively. On ETIS-LaribPolypDB and PolypGen-C6, mDice increases by 3.24% and 7.26% over ESPNet, with substantially lower HD95 and ASD, demonstrating strong cross-dataset generalization. Ablation studies show that KAN modules improve mDice by up to 5.81% and mIoU by 6.45% while reducing HD95 and ASD. Stratified analysis on 822 external images shows consistent mDice gains under dark (+4.98%), medium (+3.47%), and bright (+5.06%) conditions. Brown–Forsythe testing further confirms lower prediction variance ($F = 11.28$, $p < 0.001$; $\sigma_1^2/\sigma_2^2 = 0.68$), supporting robust and stable performance across imaging variations. The source code and implementation are publicly available at https://github.com/biodatlab/kanainet.

1 Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death and third most common cancer across the globe [17]. Many studies show that early detection and polyp removal with colonoscopy significantly reduce CRC mortality rates [24]. However, manual polyp detection and delineation is time-consuming and depends on operator's subjectivity. Automated polyp segmentation therefore plays a major role in assisting clinicians for the task.

Deep learning has been the primary approach for this task, evolving into several architectural paradigms. CNN-based approaches such as U-Net, U-Net++, and nnU-Net [8,19,25], established dense pixel-wise prediction through skip connections and feature aggregation, yet their limited receptive fields hinder the capture of long-range dependencies. PraNet [7] proposed parallel reverse attention for high-level features with background noise suppression. SANet [23] employed shallow attention for feature filtering. Subsequently, vision transformer has been used to increase global context awareness of the models. Polyp-PVT [3] leverages pyramid vision transformers for multi-scale feature fusion. With the importance of edge detection in the task, MEGANet [4] introduced edge-guided attention to preserve boundary information, at multiple scales and ESPNet [22] proposed an edge-aware feature shrinkage pyramid upon a DeiT encoder that demonstrated strong out-of-distribution generalization across clinical centers. However, these state-of-the-art models address endoscopic challenges through fixed, input-agnostic operations that cannot fully adapt to local imaging context. Specifically: (1) illumination variation from the endoscope's point-light source is typically handled through data augmentation or normalization rather than learned adaptive photometric correction, leaving feature quality vulnerable to specular highlights and shadows; and (2) boundary detection relies on fixed convolutional edge filters whose responses are modulated by downstream attention but cannot themselves adapt to the local polyp morphology and imaging condition, limiting discrimination between true polyp margins and visual artifacts.

Kolmogorov-Arnold Network (KAN) replaces fixed activations with learnable univariate functions, allowing each network unit to adapt its nonlinear response to specific tasks and local context. U-KAN [12] first demonstrated this potential for medical image segmentation by integrating KAN layers into a U-Net architecture and achieving improved representational flexibility over fixed-activation baselines. However, U-KAN applies KAN layers as general-purpose replacements without addressing the domain-specific challenges of endoscopic imaging.

Here, we propose KAN-AINet which integrates ConvNeXt–U-Net with Kolmogorov–Arnold Network (KAN) modules. KAN-IMM (Illumination Modulation Module) learns adaptive per-channel scaling conditioned on local brightness and contrast to stabilize features in varying conditions. KAN-BAM (Boundary Attention Module) fuses multi-scale edge cues with KAN-based nonlinear discrimination to refine boundary sensitivity.

2 Method

We design KAN-AINet as a modular framework (see **Fig. 1**) in which KAN-based components act as plug-and-play modules for integration into existing segmentation architectures. The framework consists of three components: (i) the KAN Utility Block (KANLinear and SpatialKAN2D), which bridges KAN-based learnable activations with conventional convolutional operations and serves as the fundamental building unit; (ii) the KAN Illumination Modulation Module (KAN-IMM), built on KANLinear, which learns adaptive per-channel scaling functions to normalize illumination variations in encoder features; and (iii) the KAN Boundary Attention Module (KAN-BAM), built on SpatialKAN2D, which generates attention maps for multi-scale edge detection to refine polyp boundaries under varying illumination. We adopt a ConvNeXt-Tiny encoder with a U-Net decoder. KAN-IMMs are inserted after each encoder stage, where features preserve low-level intensity and contrast cues that directly reflect lighting conditions, enabling early normalization and preventing illumination-induced noise from propagating into deeper semantic representations. KAN-BAMs are applied within decoder stages, where boundary refinement is most effective, as multi-scale semantic features aggregated through skip connections provide the spatial context needed to distinguish true polyp boundaries from illumination artifacts such as specular reflections and shadow transitions. We further conduct an ablation study comparing three variants: (i) the full KAN-AINet with both KAN-IMM and KAN-BAM, (ii) KAN-AINet with KAN-IMM only, and (iii) KAN-AINet with KAN-BAM only to evaluate each module's contribution.

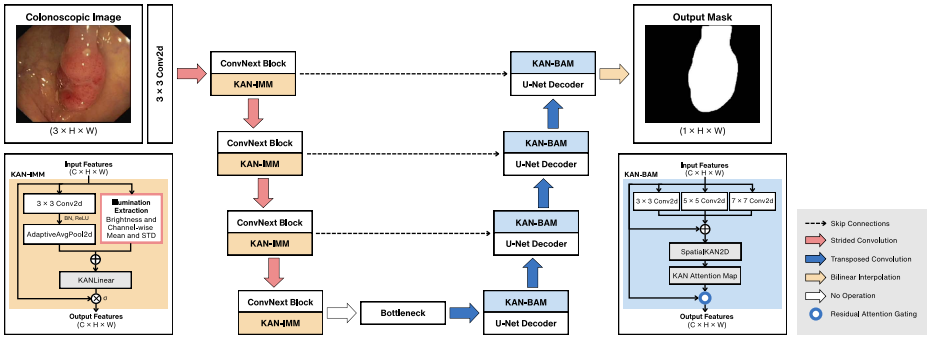

Fig. 1. Overview of KAN-AINet architecture. The encoder integrates KAN-IMM for adaptive illumination modulation. The decoder uses KAN-BAM for boundary-aware feature refinement via skip connections.

2.1 KAN Utility Blocks

KAN utility blocks, consisting of KANLinear and SpatialKAN2D, act as operators for our KAN modules. These components incorporate intermediate features from the ConvNeXt–U-Net backbone and process them through KAN to produce adaptive activations.

KANLinear KANLinear represents a single computational unit of KAN, where each connection learns its own activation function. The output is computed as:

$$y_i = \sum_{j=1}^{n_{\text{in}}} \left[w_{i,j} \cdot \text{silu}(x_j) + \sum_{k=0}^{G+K-1} c_{i,j,k} \cdot B_{k,K}\left(\frac{x_j - x_{\min}}{x_{\max} - x_{\min}} \cdot 2 - 1 \right) \right] \quad (1)$$

The first term provides a residual linear pathway through base weights $w_{i,j}$ combined with the Sigmoid Linear Unit (SiLU) activation, defined as silu. The second term models nonlinear transformations via learnable spline coefficients $c_{i,j,k}$ over cubic B-spline basis functions ($K = 3$, grid size $G = 5$).

SpatialKAN2D SpatialKAN2D extends KANLinear to 2D feature maps, enabling seamless integration with conventional segmentation backbones. Given an input feature map $F \in \mathbb{R}^{C_{\text{in}} \times H \times W}$, each spatial location (h, w) contains a channel feature vector $f(h, w) \in \mathbb{R}^{C_{\text{in}}}$ encoding local appearance information. SpatialKAN2D applies a shared KANLinear transformation independently to each spatial vector:

$$F' = \text{SpatialKAN2D}(F), \quad \text{where} \quad f'(h, w) = \text{KANLinear}(f(h, w)). \quad (2)$$

KANLinear and SpatialKAN2D constitute key contributions of this work. To the best of our knowledge, this is the first implementation of KAN designed to bridge fully connected and convolutional KAN formulations. Unlike prior approaches that either flatten spatial features or replace convolutional kernels with learned activations, SpatialKAN2D preserves spatial structure by applying a shared KANLinear at each location while modeling complex channel-wise interactions. This design makes it more spatially aware than standard KAN linear layers and more parameter-efficient than fully convolutional KAN variants.

2.2 Encoder: KAN Illumination Modulation Module (KAN-IMM) with ConvNeXt encoders

KAN-IMM modulates encoder features after each ConvNeXt stage by learning per-channel scaling coefficients from both local feature information and illumination cues. Specifically, it extracts two compact 1D descriptors from the input feature map: a local descriptor, obtained by applying Conv-BN-ReLU followed by global average pooling to capture spatial patterns, and an illumination descriptor, computed from scalar statistics that summarize the overall lighting

condition. These descriptors are concatenated and passed through a KANLinear layer to generate per-channel scaling weights. The input feature channels are then rescaled through channel-wise multiplication, enabling the model to adaptively correct illumination-dependent features and focus on actual polyp morphology rather than lighting variations. (see **Fig. 1**).

2.3 Decoder: KAN Boundary Attention Module (KAN-BAM) with U-Net decoders

KAN-BAM detects boundaries from decoder features through multi-scale edge detection. We apply depthwise convolutions with kernel sizes 3×3, 5×5, and 7×7 to capture from fine edges to structural boundaries. We concatenate these multi-scale edge features with the original decoder features and pass them through SpatialKAN2D, whose learnable spline activations distinguish between illumination artifacts and polyp boundaries. A lightweight attention head then produces a single-channel boundary map with sigmoid activation.

2.4 Loss Function

We use a weighted combination of Dice (1.0) and Focal (0.5) losses. Dice optimizes prediction–ground truth overlap, while Focal emphasizes hard regions such as boundaries and illumination artifacts.

3 Experiment and Results

3.1 Datasets

We followed the official ESPNet training and external validation protocol [22] to ensure fair comparison with state-of-the-art methods. The training set included 4,000 colonoscopy images from Kvasir-SEG (804 images) [9], PolypGen C1–C5 (1,182 images) [1], and the SUN-SEG training split (2,014 images) [10,11,16]. The combined dataset was divided into training, validation, and internal test sets using a 70:15:15 split. Our exact data split is publicly available in the repository for reproducibility. External validation evaluated generalization on both seen and unseen datasets. The seen set, Kvasir-Sessile (196 images), originated from the same clinical center as part of the training data but was excluded from training. The unseen sets included CVC-ColonDB (345 images) [2], ETIS-LaribPolypDB (193 images) [21], and PolypGen-C6 (88 images) [1], which differed in acquisition devices and imaging conditions.

3.2 Evaluation Metrics

We evaluated segmentation using seven metrics [5,6,14,18]. Mean Intersection over Union (mIoU) and mean Dice coefficient (mDice) measure region overlap.

Structure measure (S_α) and weighted F-measure (F_β^w) assess structural similarity and location-aware precision–recall. Mean Absolute Error (MAE) quantifies pixel-wise prediction confidence. Two boundary metrics, the 95th-percentile Hausdorff Distance (HD95) and Average Surface Distance (ASD), measure the worst-case and mean boundary alignment error, respectively.

3.3 Implementation Details

All experiments were implemented in PyTorch on a single NVIDIA GeForce RTX 3090 GPU. We used an ImageNet-pretrained ConvNeXt backbone, while all KAN modules were trained from scratch. Images and masks were resized to 512×512. Training was conducted for 100 epochs with a batch size of 8 using deep supervision. We optimized with AdamW (initial learning rate 1×10^{-4}) and cosine annealing, and selected the best model by validation Dice.

Geometric augmentations included horizontal ($p = 0.5$) and vertical ($p = 0.3$) flips, 90° rotations ($p = 0.5$), `ShiftScaleRotate` (shift 0.1, rotation 15°, $p = 0.5$), and random scaling (80%–120%, $p = 0.1$). Motion blur (kernel 3–9, $p = 0.2$) and coarse dropout (up to 8 holes, 8×8 to 32×32, $p = 0.3$) were applied to simulate endoscopic motion and occlusion. Illumination augmentations were not applied to any model in order to isolate the effect of the proposed KAN-IMM module and avoid confounding improvements from augmentation-based illumination variability.

3.4 Result

Quantitative Results We compared KAN-AINet with nnU-Net and ESPNet under identical training settings (see **Table 1**); all overlap-metric gains were reported as relative improvements. Relative to nnU-Net, KAN-AINet improved mDice/mIoU by 36.67%/51.48% on the seen Kvasir-Sessile dataset and 42.27%/55.51% on the combined unseen datasets. Compared with ESPNet, KAN-AINet achieved comparable mDice and mIoU on Kvasir-Sessile while slightly improving S_α (+0.22%) and F_β^w (+0.22%), and reducing HD95 and ASD by 15.1% and 15.7%, respectively. On the unseen datasets, KAN-AINet improved mDice and mIoU by 2.20% and 3.05%, with the largest gains on ETIS-LaribPolypDB (+3.24% mDice) and PolypGen-C6 (+7.26% mDice), demonstrating strong cross-dataset generalization. Its main advantage was boundary accuracy, reducing HD95 and ASD by 55.6% and 61.8% relative to ESPNet. Although ESPNet remains competitive on S_α and F_β^w and performed better on CVC-ColonDB, KAN-AINet achieves these results with 151M parameters and 54.15 FPS at 384×384 resolution (18.47 ms mean latency), compared with ESPNet's 297M parameters and 12.2 FPS (81.95 ms), exceeding the clinical real-time threshold (≥ 25 FPS).

Illumination Robustness Analysis We used stratified analysis on 822 images from four external datasets. Images were partitioned into three equal brightness

Table 1. Performance comparison on seen and unseen datasets. The best results are in **bold**, second-best underlined. "No KAN" refers to KAN-AINet without KAN module. The bottom block shows average scores across all unseen datasets.

Dataset	Method	mDice↑	mIoU↑	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	MAE↓	$HD_{95} \downarrow$	ASD↓
Kvasir-Sessile (seen)	nnU-Net	0.649	0.540	0.775	0.834	0.072	122.34	48.03
	ESPNet	0.887	0.818	0.920	0.916	0.020	36.76	8.78
	No KAN	0.856	0.785	0.909	0.911	0.024	54.10	20.50
	KAN-IMM only	**0.890**	**0.824**	**0.925**	**0.920**	**0.018**	35.20	11.40
	KAN-BAM only	0.876	0.807	0.919	0.919	0.021	43.20	13.00
	KAN-AINet (Ours)	0.887	0.818	0.922	0.918	0.019	**31.20**	**7.40**
CVC-ColonDB (unseen)	nnU-Net	0.608	0.505	0.763	0.777	0.055	129.77	55.57
	ESPNet	**0.816**	**0.738**	**0.888**	**0.859**	**0.032**	**47.55**	**21.70**
	No KAN	0.731	0.657	0.830	0.759	0.044	120.50	89.20
	KAN-IMM only	0.775	0.704	0.856	0.799	0.040	82.90	56.70
	KAN-BAM only	0.747	0.675	0.846	0.782	0.041	105.00	77.80
	KAN-AINet (Ours)	0.784	0.711	0.862	0.811	0.039	78.30	52.00
ETIS-Larib-PolypDB (unseen)	nnU-Net	0.532	0.465	0.757	0.599	0.023	396.53	322.89
	ESPNet	0.832	0.761	0.902	**0.899**	0.031	98.62	40.04
	No KAN	0.809	0.744	0.894	0.860	0.011	94.50	66.10
	KAN-IMM only	0.800	0.739	0.889	0.840	0.011	90.70	68.20
	KAN-BAM only	0.804	0.740	0.892	0.854	0.011	87.70	60.70
	KAN-AINet (Ours)	**0.859**	**0.795**	**0.919**	0.896	**0.007**	**48.40**	**29.10**
PolypGen-C6 (unseen)	nnU-Net	0.637	0.552	0.805	0.725	0.050	419.75	241.59
	ESPNet	0.813	0.760	0.905	**0.899**	**0.012**	336.64	218.55
	No KAN	0.837	0.787	0.903	0.869	0.014	89.90	65.50
	KAN-IMM only	0.856	0.806	0.906	0.873	0.019	108.30	41.80
	KAN-BAM only	0.839	0.790	0.904	0.863	0.016	135.40	71.60
	KAN-AINet (Ours)	**0.872**	**0.822**	**0.914**	0.892	0.015	**87.60**	**26.10**
Average across all unseen datasets								
	nnU-Net	0.589	0.499	0.767	0.715	0.044	252.78	164.14
	ESPNet	0.820	0.753	**0.898**	**0.886**	0.025	160.94	93.43
	No KAN	0.792	0.729	0.876	0.829	0.023	101.63	73.60
	KAN-IMM only	0.810	0.750	0.884	0.837	0.023	93.97	55.57
	KAN-BAM only	0.797	0.735	0.881	0.833	0.023	109.37	70.03
	KAN-AINet (Ours)	**0.838**	**0.776**	**0.898**	0.866	**0.020**	**71.43**	**35.73**

groups. We then measured KAN-AINet's performance advantage compared to its no-KAN counterpart. KAN-AINet achieved statistically significant p-value and Dice improvements across all three partitions: dark, medium, and bright for 4.98%, 3.47%, and 5.06%, respectively. Improvements were larger under extreme illumination (dark and bright) compared to medium, indicating enhanced robustness to challenging lighting conditions.

Additionally, Brown–Forsythe variance testing demonstrated significantly lower overall prediction variance for KAN-AINet compared to without KAN

($F = 11.28$, $p < 0.001$; overall $\sigma_1^2/\sigma_2^2 = 0.68$). Subgroup analyses showed consistently reduced variance across illumination partitions (σ_1^2/σ_2^2 – dark: 0.76, medium: 0.51, bright: 0.66), indicating more stable and robust predictions under varying illumination conditions.

Feature Map Analysis We analyzed feature maps from KAN-AINet and its baseline variant (without KAN blocks), both trained separately from scratch (see **Fig. 2**). In Encoders 2 and 4 (Columns 2-3), KAN-IMM concentrated activations on the polyp region while suppressing lighting artifacts and mucosal edges. The baseline model treated these artifacts as salient features, producing diffuse activation maps. However, KAN-IMM occasionally oversuppressed polyp regions in Encoder 4, requiring recovery in the decoder pathway. KAN-BAM addressed this in Decoder 3 (Column 4) by refining boundaries and recovering suppressed polyp features. KAN-AINet produced sharper, higher-confidence edge maps at true polyp boundaries compared to the baseline. In regions where the baseline generated false positives, KAN-AINet correctly reduced confidence and prevents incorrect segmentation.

This analysis revealed a two-stage strategy in KAN-AINet: KAN-IMM filters illumination artifacts during encoding, followed by KAN-BAM, which recovered polyp structures and refined boundaries during decoding. This sequential modulation enabled the model to focus on task-relevant features while suppressing illumination-related confounders.

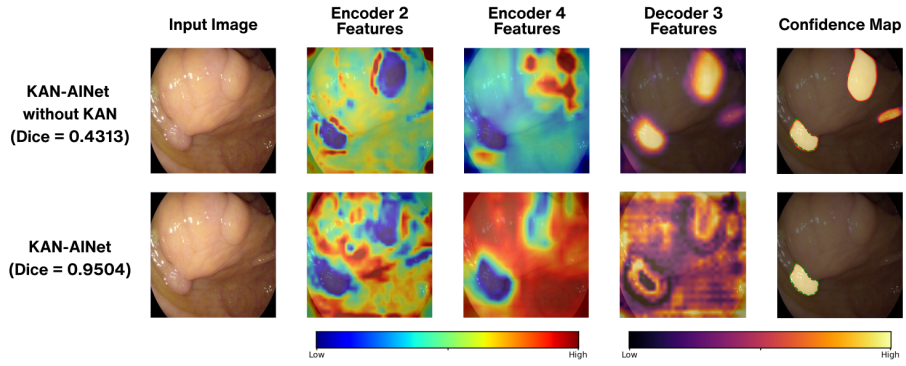

Fig. 2. Qualitative comparison of model performance: (i) KAN-AINet vs. (ii) KAN-AINet without KAN blocks.

3.5 Ablation Study: KAN Block Removal

On the unseen datasets, KAN-AINet outperformed all ablation variants across every evaluation metric. Aggregated over the unseen datasets, it improved mDice

by 5.81%, mIoU by 6.45%, S_α by 2.51%, and F_β^w by 4.46% over the No-KAN baseline, along with a 13.04% reduction in MAE. Boundary delineation accuracy improved substantially, reducing HD95 by 29.72% and ASD by 51.45% relative to the No-KAN baseline. KAN-IMM alone sometimes oversuppressed illumination changes, while KAN-BAM alone struggled with boundary accuracy under strong lighting interference. When combined, KAN-IMM adjusted illumination during encoding and KAN-BAM refined structural details during decoding, complementing each other and improving generalization on unseen datasets.

Acknowledgments. This work was supported by the Faculty of Medicine Ramathibodi Hospital, Mahidol University; the Department of Biomedical Engineering, Faculty of Engineering, Mahidol University; and the University of Alabama at Birmingham. This research also received funding from the National Science, Research and Innovation Fund (NSRF) through the Research and Innovation Acceleration Agency for Competitiveness and Area Development (RCAD) under the Program Management Unit for Frontier Brainpower and Future Industries (grant number B13F690044).

Disclosure of Interests. The authors declare that they have no conflicts of interest.

References

1. Ali, S., Jha, D., Ghatwary, N., Realdon, S., Cannizzaro, R., Salem, O.E., Lamarque, D., Daul, C., Riegler, M.A., Anonsen, K.V., et al.: A multi-centre polyp detection and segmentation dataset for generalisability assessment. Scientific Data **10**(1), 75 (2023)
2. Bernal, J., Sánchez, J., Vilariño, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognition **45**(9), 3166–3182 (2012). `https://doi.org/https://doi.org/10.1016/j.patcog.2012.03.002`, best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011)
3. Bo, D., Wenhai, W., Deng-Ping, F., Jinpeng, L., Huazhu, F., Ling, S.: Polyp-pvt: Polyp segmentation with pyramidvision transformers. CAAI AIR (2023)
4. Bui, N.T., Hoang, D.H., Nguyen, Q.T., Tran, M.T., Le, N.: Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 7985–7994 (January 2024)
5. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88**(2), 303–338 (2010)
6. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 4548–4557 (2017)
7. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 263–273. Springer (2020)
8. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., Maier-Hein, K.H.: nnu-net: Self-adapting framework for u-net-based medical image segmentation (2018), `https://arxiv.org/abs/1809.10486`

9. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: International Conference on Multimedia Modeling. pp. 451–462. Springer (2020)
10. Ji, G.P., Chou, Y.C., Fan, D.P., Chen, G., Jha, D., Fu, H., Shao, L.: Progressively normalized self-attention network for video polyp segmentation. In: MICCAI. pp. 142–152 (2021)
11. Ji, G.P., Xiao, G., Chou, Y.C., Fan, D.P., Zhao, K., Chen, G., Fu, H., Van Gool, L.: Video polyp segmentation: A deep learning perspective. Machine Intelligence Research (2022)
12. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
13. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljacic, M., Hou, T.Y., Tegmark, M.: KAN: Kolmogorov–arnold networks. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=Ozo7qJ5vZi
14. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 248–255 (2014)
15. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
16. Misawa, M., Kudo, S.e., Mori, Y., Hotta, K., Ohtsuka, K., Matsuda, T., Saito, S., Kudo, T., Baba, T., Ishida, F., et al.: Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointestinal endoscopy **93**(4), 960–967 (2021)
17. Morgan, E., Arnold, M., Gini, A., Lorenzoni, V., Cabasag, C.J., Laversanne, M., Vignat, J., Ferlay, J., Murphy, N., Bray, F.: Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from globocan. Gut **72**(2), 338–344 (2023). https://doi.org/10.1136/gutjnl-2022-327736, https://gut.bmj.com/content/72/2/338
18. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 733–740 (2012)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
20. Schmidt-Hieber, J.: The kolmogorov–arnold representation theorem revisited. Neural Networks **137**, 119–126 (2021). https://doi.org/https://doi.org/10.1016/j.neunet.2021.01.020
21. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International journal of computer assisted radiology and surgery **9** (09 2013). https://doi.org/10.1007/s11548-013-0926-3
22. Toman, R., Subramanian, V., Ali, S.: ESPNet: Edge-Aware Feature Shrinkage Pyramid for Polyp Segmentation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15970. Springer Nature Switzerland (September 2025)

23. Wei, J., Hu, Y., Zhang, R., Li, Z., Zhou, S., Cui, S.: Shallow attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 699–708. Springer (2021)
24. Zhang, J., Chen, G., Li, Z., Zhang, P., Li, X., Gan, D., Cao, X., Du, H., Zhang, J., Zhang, L., Ye, Y.: Colonoscopic screening is associated with reduced colorectal cancer incidence and mortality: a systematic review and meta-analysis. J Cancer **11**, 5953–5970 (2020). https://doi.org/10.7150/jca.46661, https://www.jcancer.org/v11p5953.htm
25. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging (2019)