NANTES - FRANCE
ICME
JUNE 30 - JULY 4
2025
IEEE International Conference on Multimedia and Expo
JOURNEY TO THE CENTER OF MACHINE IMAGINATION

ICME 2025

# KAN-SAM: Kolmogorov-Arnold Network Guided Segment Anything Model for RGB-T Salient Object Detection

Xingyuan Li, Ruichao Hou*, Tongwei Ren, Gangshan Wu

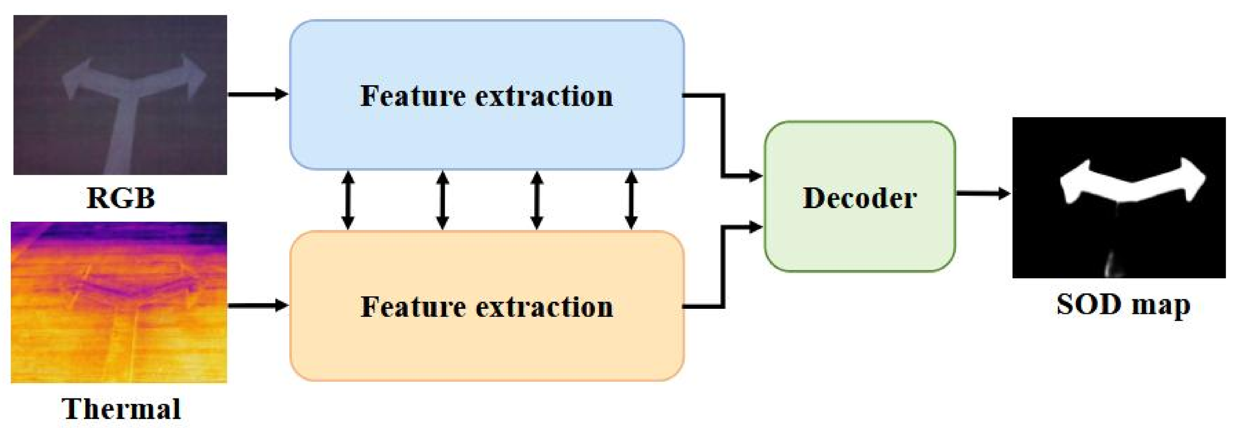
State Key Laboratory for Novel Software Technology, Nanjing University

Nanjing 210023, China

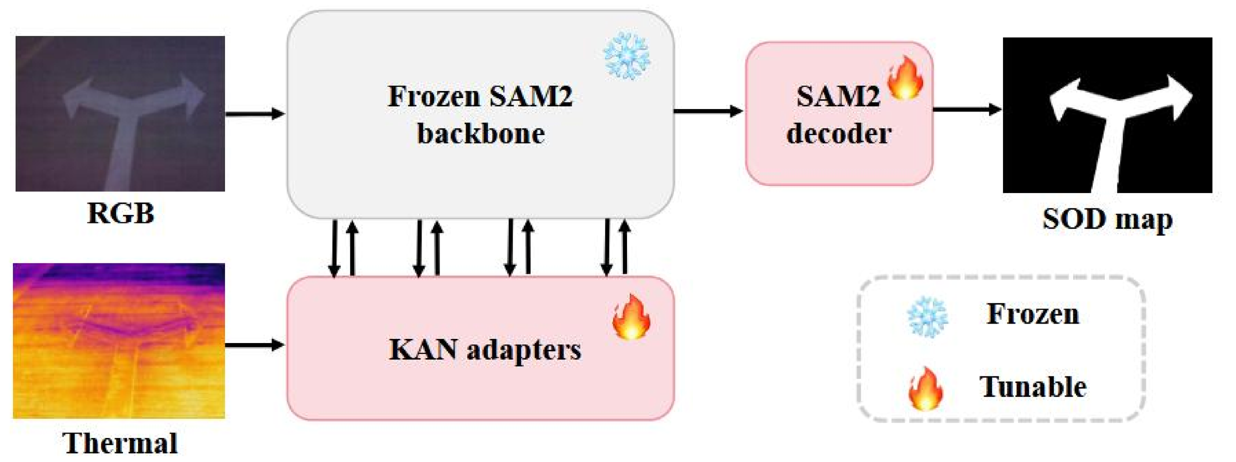
## Introduction

**RGB-thermal salient object detection (RGB-T SOD)** integrates thermal information with RGB images to improve detection performance under challenging conditions such as low illumination, cluttered environments, and complex backgrounds.

We propose a novel prompt learning-based RGB-T SOD method that uses KAN adapters to combine RGB and thermal features in SAM2 for RGB-T SOD.



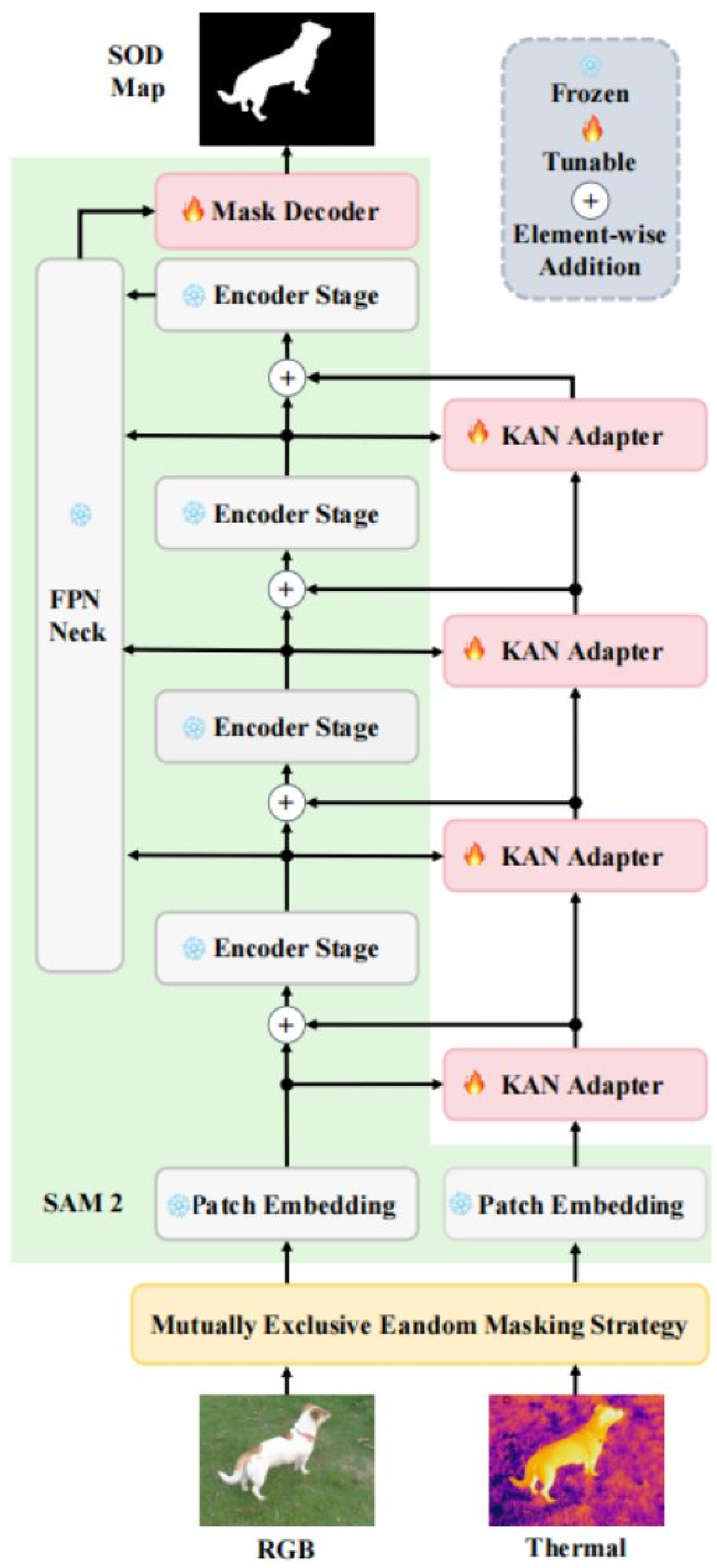

(a) Existing RGB-T SOD methods




(b) Our method

## Method

Given a set of RGB and corresponding thermal images, we introduce a mutually exclusive random masking strategy at the input level. The masked images are then converted into embedding features using the patch embedding module. The RGB features are processed through the frozen SAM2 Hiera backbone, while the KAN adapters progressively introduce thermal features into the RGB features at each stage. The multi-stage prompting mechanism improves the robustness of multi-modal representations. The multi-scale features extracted in this manner are further refined through the frozen Feature Pyramid Network (FPN) neck for deeper fusion. Finally, the fused features are sent to a tunable mask decoder, which generates the final saliency maps.




## Experiments

**Dataset:** **VT821, VT1000, VT5000**

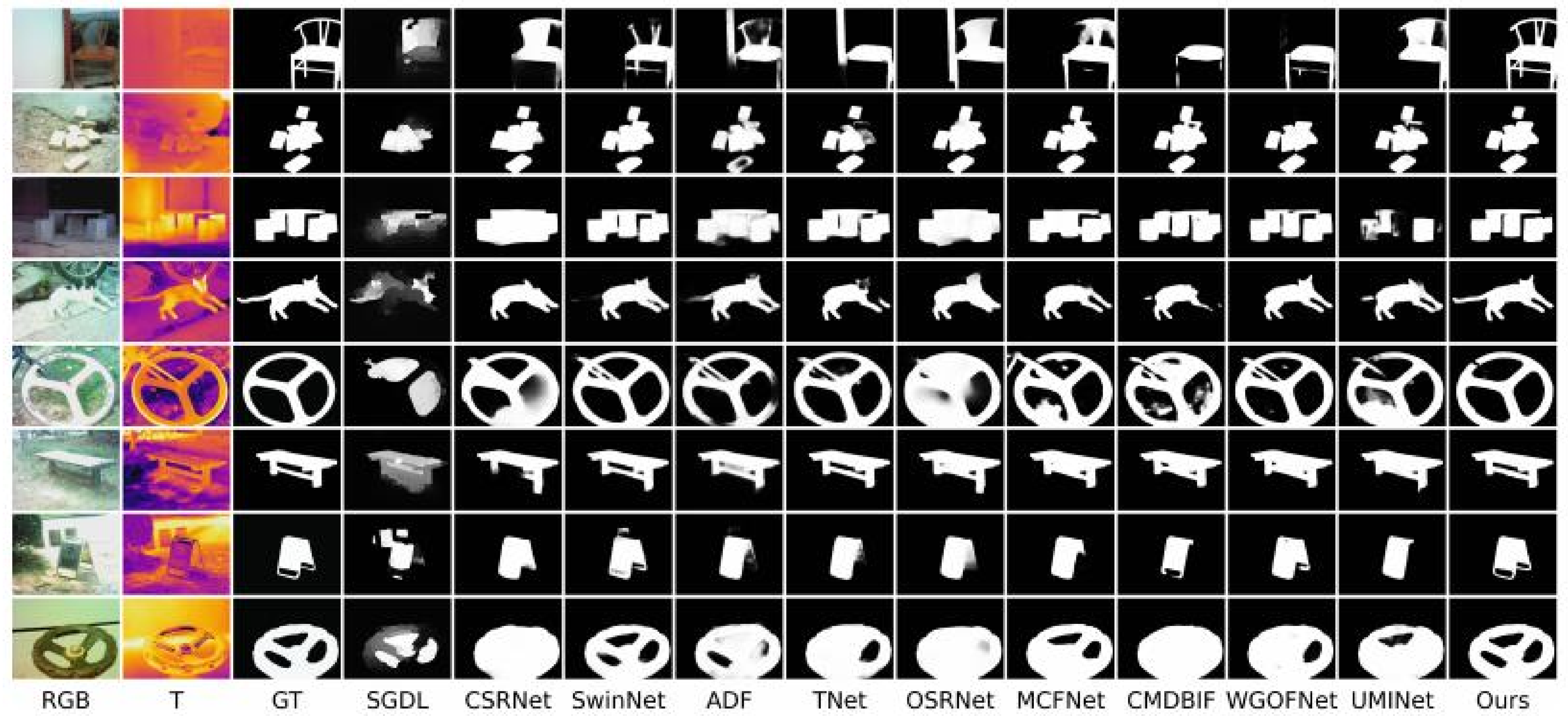
- Training: 2500
- Test: 821, 1000, 2500

**Metrics:** Favg, Fmax, Fw, MAE, Em, and Sm

**Comparison with the SOTA:** KAN-SAM consistently outperforms existing methods across multiple evaluation metrics.

**Qualitative Analysis:** KAN-SAM demonstrates superior visual performance compared to existing methods.

| Methods | VT5000 | | | | | | VT1000 | | | | | | VT821 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $F_{avg}\uparrow$ | $F_{max}\uparrow$ | $F_w\uparrow$ | $MAE\downarrow$ | $E_m\uparrow$ | $S_m\uparrow$ | $F_{avg}\uparrow$ | $F_{max}\uparrow$ | $F_w\uparrow$ | $MAE\downarrow$ | $E_m\uparrow$ | $S_m\uparrow$ | $F_{avg}\uparrow$ | $F_{max}\uparrow$ | $F_w\uparrow$ | $MAE\downarrow$ | $E_m\uparrow$ | $S_m\uparrow$ |
| SGDL [28] | 0.672 | 0.737 | 0.558 | 0.089 | 0.824 | 0.750 | 0.764 | 0.807 | 0.652 | 0.090 | 0.856 | 0.787 | 0.731 | 0.780 | 0.583 | 0.085 | 0.846 | 0.764 |
| CSRNet [3] | 0.811 | 0.857 | 0.796 | 0.042 | 0.905 | 0.868 | 0.877 | 0.918 | 0.878 | 0.024 | 0.925 | 0.918 | 0.831 | 0.880 | 0.821 | 0.038 | 0.909 | 0.885 |
| CGFNet [11] | 0.851 | 0.887 | 0.831 | 0.035 | 0.922 | 0.883 | 0.906 | 0.936 | 0.900 | 0.023 | 0.944 | 0.923 | 0.845 | 0.885 | 0.829 | 0.038 | 0.912 | 0.881 |
| SwinNet [14] | 0.865 | 0.915 | 0.846 | 0.026 | 0.942 | 0.912 | 0.896 | 0.948 | 0.894 | 0.018 | 0.947 | 0.938 | 0.847 | 0.903 | 0.818 | 0.030 | 0.926 | 0.904 |
| ADF [29] | 0.778 | 0.863 | 0.722 | 0.048 | 0.891 | 0.864 | 0.847 | 0.923 | 0.804 | 0.034 | 0.921 | 0.910 | 0.717 | 0.804 | 0.627 | 0.077 | 0.843 | 0.810 |
| TNet [30] | 0.846 | 0.895 | 0.840 | 0.033 | 0.927 | 0.895 | 0.889 | 0.937 | 0.895 | 0.021 | 0.937 | 0.929 | 0.842 | 0.904 | 0.841 | 0.030 | 0.919 | 0.899 |
| OSRNet [31] | 0.823 | 0.866 | 0.807 | 0.040 | 0.908 | 0.875 | 0.892 | 0.929 | 0.891 | 0.022 | 0.935 | 0.926 | 0.814 | 0.862 | 0.801 | 0.043 | 0.896 | 0.875 |
| ACMANet [32] | 0.858 | 0.890 | 0.823 | 0.033 | 0.932 | 0.887 | 0.904 | 0.933 | 0.889 | 0.021 | 0.945 | 0.927 | 0.837 | 0.873 | 0.807 | 0.035 | 0.914 | 0.883 |
| MCFNet [33] | 0.848 | 0.886 | 0.836 | 0.033 | 0.924 | 0.887 | 0.902 | 0.939 | 0.906 | 0.019 | 0.944 | 0.932 | 0.844 | 0.889 | 0.835 | 0.029 | 0.918 | 0.891 |
| CMDBIF [34] | 0.868 | 0.892 | 0.846 | 0.032 | 0.933 | 0.886 | 0.914 | 0.931 | 0.909 | 0.019 | 0.952 | 0.927 | 0.856 | 0.887 | 0.837 | 0.032 | 0.923 | 0.882 |
| MITF-Net [35] | 0.880 | 0.899 | 0.870 | 0.025 | 0.943 | 0.910 | 0.915 | 0.938 | 0.906 | 0.016 | 0.949 | 0.938 | 0.865 | 0.891 | 0.853 | 0.027 | 0.927 | 0.905 |
| CAVER [36] | 0.856 | 0.897 | 0.849 | 0.028 | 0.935 | 0.899 | 0.906 | 0.945 | 0.912 | 0.016 | 0.949 | 0.938 | 0.854 | 0.897 | 0.846 | 0.026 | 0.928 | 0.897 |
| ADNet [37] | 0.893 | 0.924 | 0.884 | 0.022 | 0.953 | 0.924 | 0.916 | 0.952 | 0.920 | 0.015 | 0.952 | 0.944 | 0.869 | 0.915 | 0.860 | 0.024 | 0.930 | 0.915 |
| WGOFNet [38] | 0.883 | 0.912 | 0.873 | 0.025 | 0.945 | 0.911 | 0.919 | 0.946 | 0.922 | 0.016 | 0.951 | 0.940 | 0.875 | 0.911 | 0.868 | 0.025 | 0.934 | 0.908 |
| UMINet [39] | 0.831 | 0.877 | 0.820 | 0.035 | 0.919 | 0.882 | 0.892 | 0.935 | 0.896 | 0.021 | 0.941 | 0.926 | 0.791 | 0.849 | 0.782 | 0.054 | 0.879 | 0.858 |
| Ours | 0.909 | 0.931 | 0.905 | 0.020 | 0.957 | 0.927 | 0.930 | 0.947 | 0.934 | 0.013 | 0.958 | 0.946 | 0.883 | 0.911 | 0.880 | 0.025 | 0.932 | 0.915 |




NANJING UNIVERSITY
lixy@smail.nju.edu.cn
rchou@nju.edu.cn
MAGUS
MediA recoGnition and UnderStanding