U-Bench: A Comprehensive Understanding of U-Net through 100-Variant Benchmarking

NEURIPS 2025

Fenghe Tang, Chengqi Dong, Wenxin Ma, Zikang Xu, Heqin Zhu, Zihang Jiang, Rongsheng Wang, Yuhao Wang, Chenxu Wu, Yingtai Li, Shaohua Kevin Zhou*

Medical Imaging, Robotic, Analytic Computing & Learning

My homepage

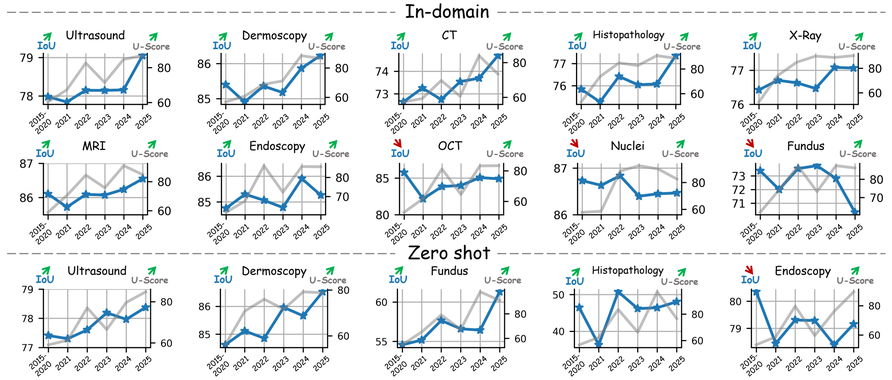


Fig: Performance trends of SOTA models over the past decade



Finding 1: In-domain performance has marginal gains in segmentation accuracy, while zero-shot improves more pronouncedly.

Finding 2: Increasing attention on efficiency.

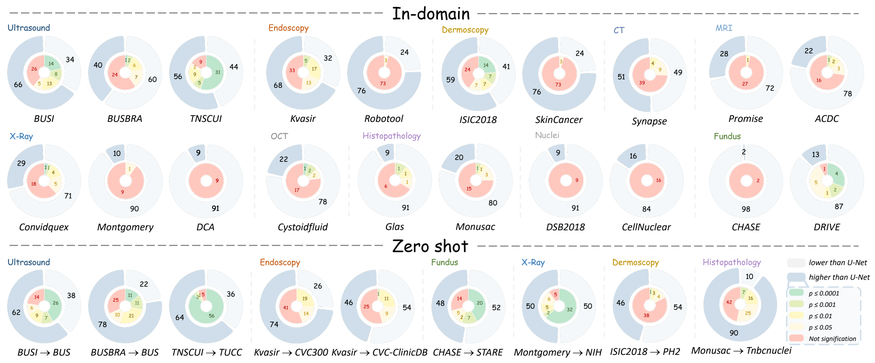


Fig: Significance analysis on the IoU against U-Net across 28 datasets across 10 modalities.



Finding 3: Although some in-domain improvements exist on average, few reach statistical significance, whereas the zero-shot improvements remain significant.

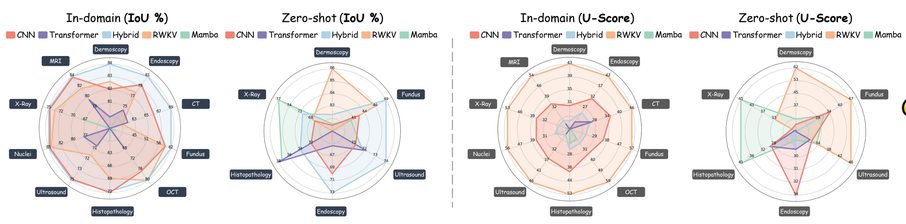
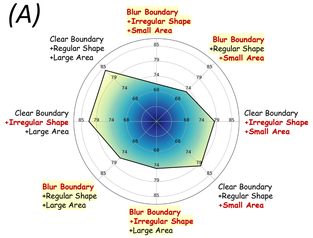


Fig: Architectures Analysis:





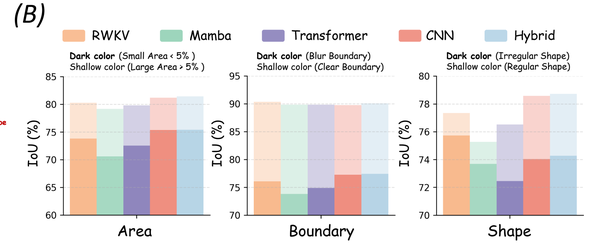
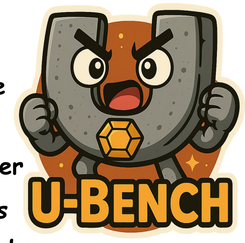


Fig: Data Characteristics Analysis:

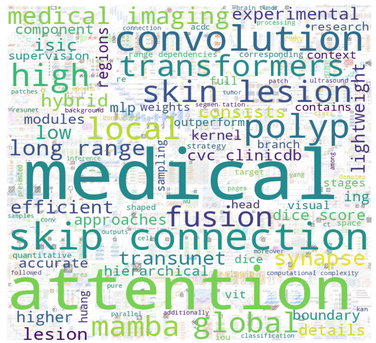


Support 🤗 **:**

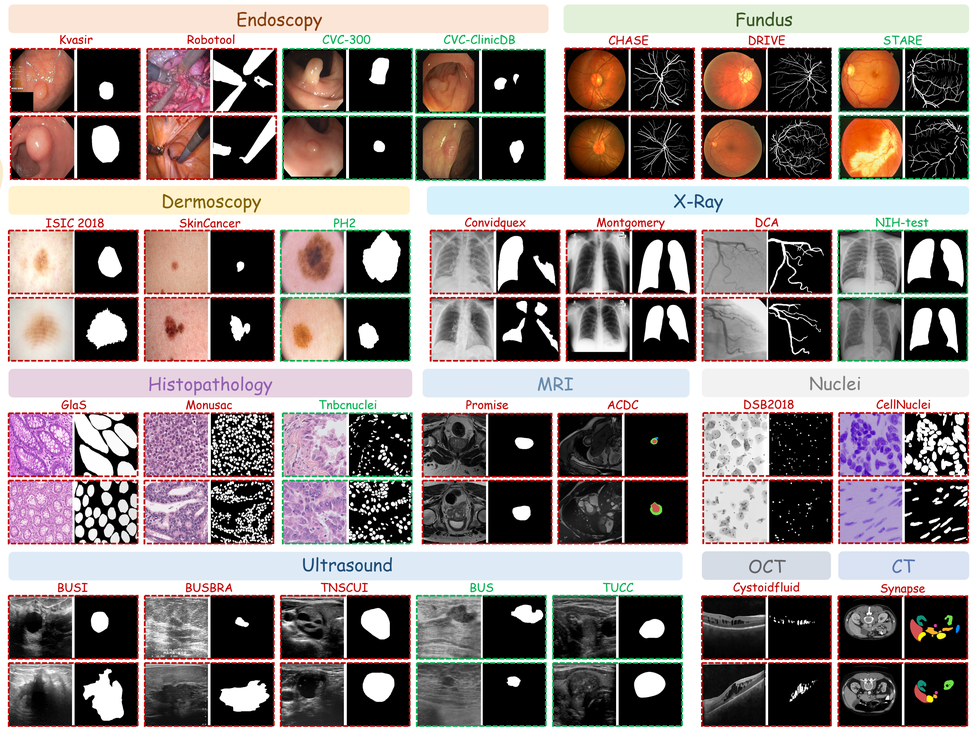
- ✔ Zero-shot evaluation
- ✔ Statistical Analysis
- ✔ Edge-aware Metric
- Model Recommender
- 🤗 **28** *datasets*
- 🤗 **100** *networks*



(A) Summary of ***U-Bench***

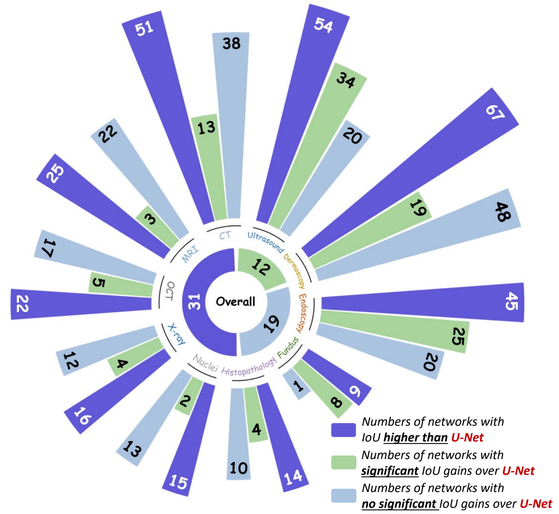


(B) Word cloud of 100 u-shape variants in Model Zoo

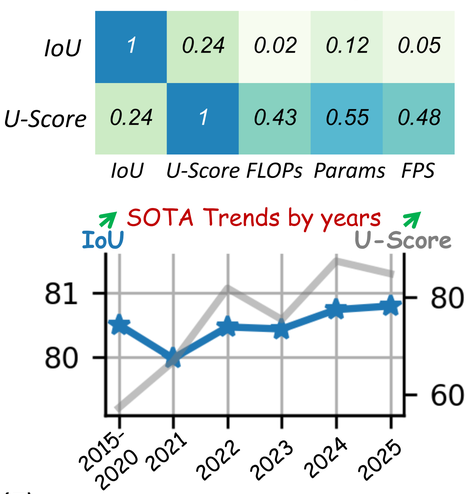


*(C) Overview of the 28 datasets (**In-domain** / **Zero-shot**) across modalities in Data Zoo*

(D) Few variants include zero-shot or significance evaluations



*(E) **Higher** / **Significance** / **no Significance** than U-Net across different modalities*



(F) Correlation and Trend of U-Score v.s. IoU

Generalization
Zero-shot Evaluation

Robustness
Statistical Significance Evaluation

Efficiency
Trade-off Metric: U-Score

Dataset Characteristics
foreground scale | Boundary sharpness | Shape complexity | Modality

Multi-Architecture
CNN | Mamba | RWKV | Transformer | Hybrid