



Endoscopic Artifact Inpainting for Improved Endoscopic Image Segmentation

Zhangyuan Yu*, Chenlin Du*, Hongrui Liang, Xiuqi Zheng,
Zeyao Ma, Mingjun Wu, Mingwu Ao, and Qicheng Lao*
Contact: raven-july@bupt.edu.cn



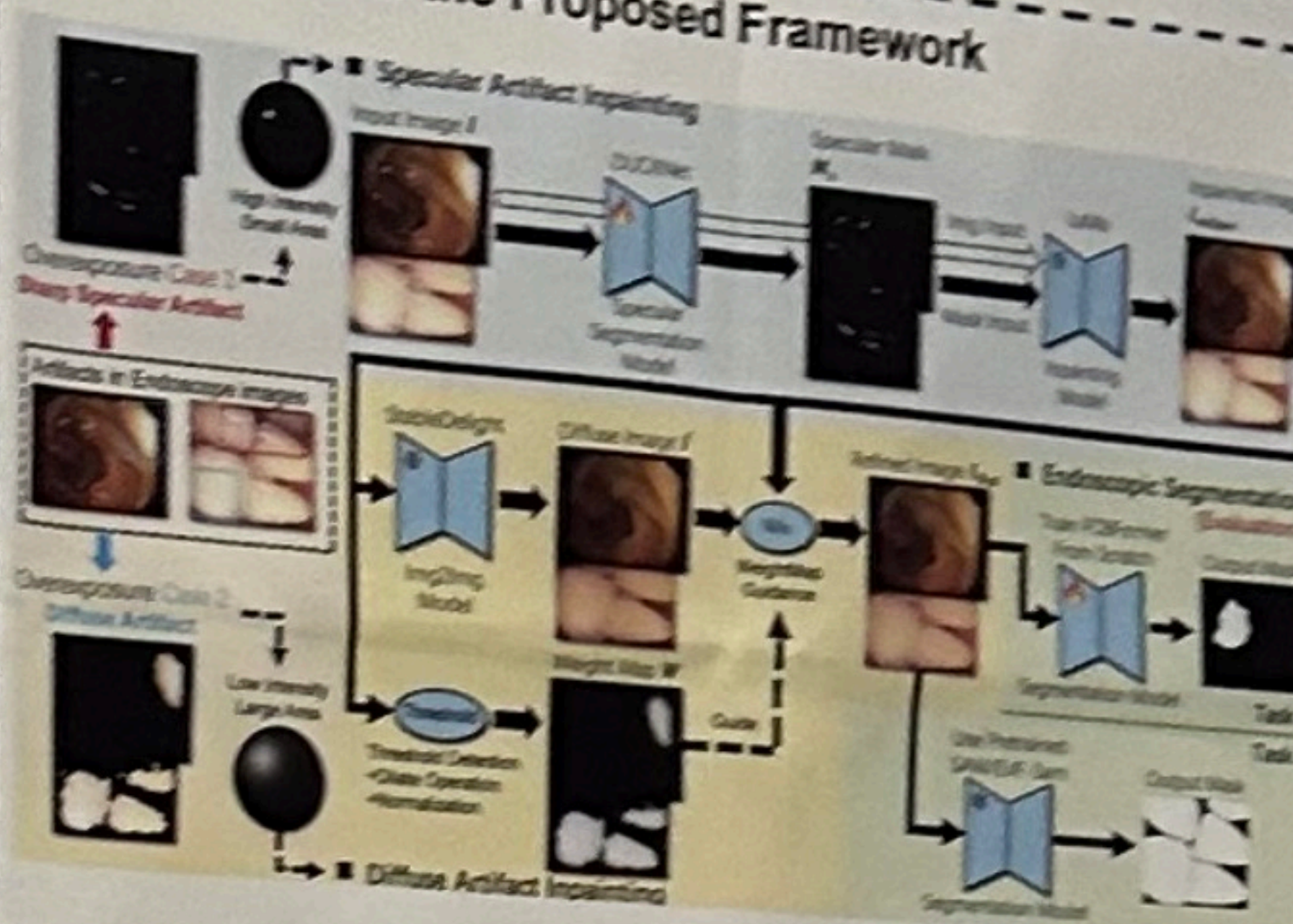
Introduction

Endoscopic imaging plays a crucial role in modern diagnostics and minimally invasive procedures. However, artifacts caused by specular and diffuse reflections present significant challenges, particularly in tasks such as endoscopic image segmentation. Existing methods tackling endoscopic artifacts typically address only one type of reflection, failing to fully account for the non-Lambertian reflectance of endoscopic tissue structures.

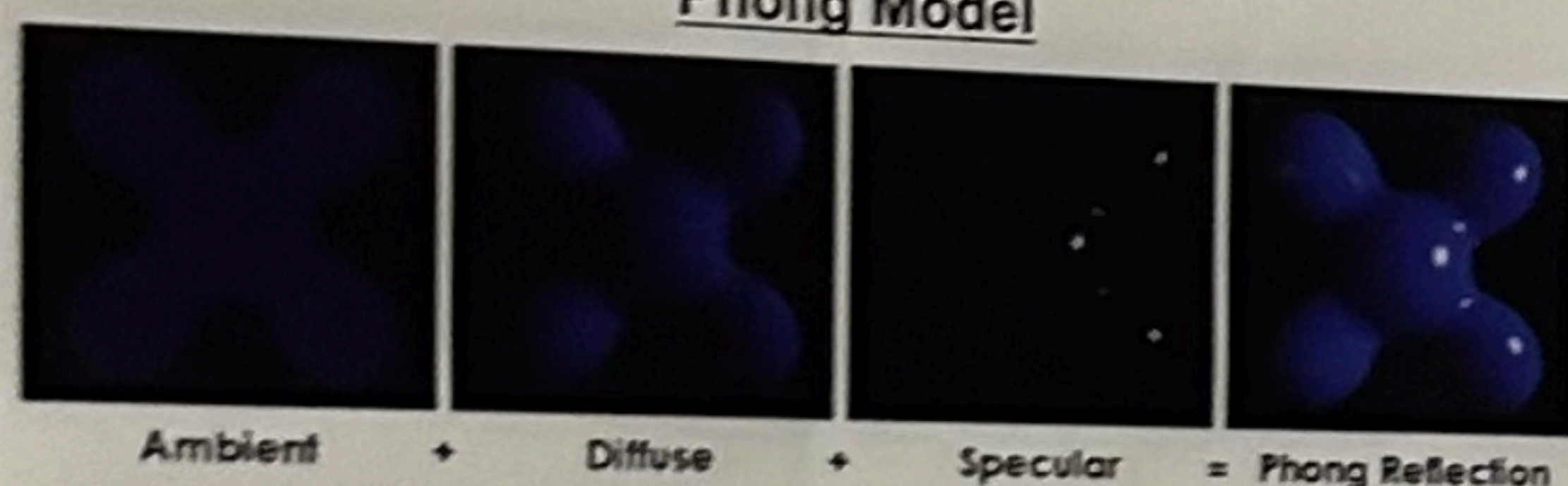
Therefore, inspired by the simplified Phong model for endoscopy, we propose a novel artifact inpainting framework. Our work can be summarized as follows:

- Propose a two-stage artifact inpainting framework. The first stage suppresses specular artifacts, while the second stage focuses on inpainting diffuse artifacts. Additionally, we introduce a weight map to control the handling of diffuse artifacts, ensuring a more precise enhancement.
- Conduct thorough experiments to compare with SOTA models. The proposed framework can offer better artifact inpainting quality and robustly improve the segmentation performance of endoscopic images.
- Conduct various ablation studies to evaluate the effectiveness of key components and hyperparameter settings in our proposed framework.

Illustration of the Proposed Framework



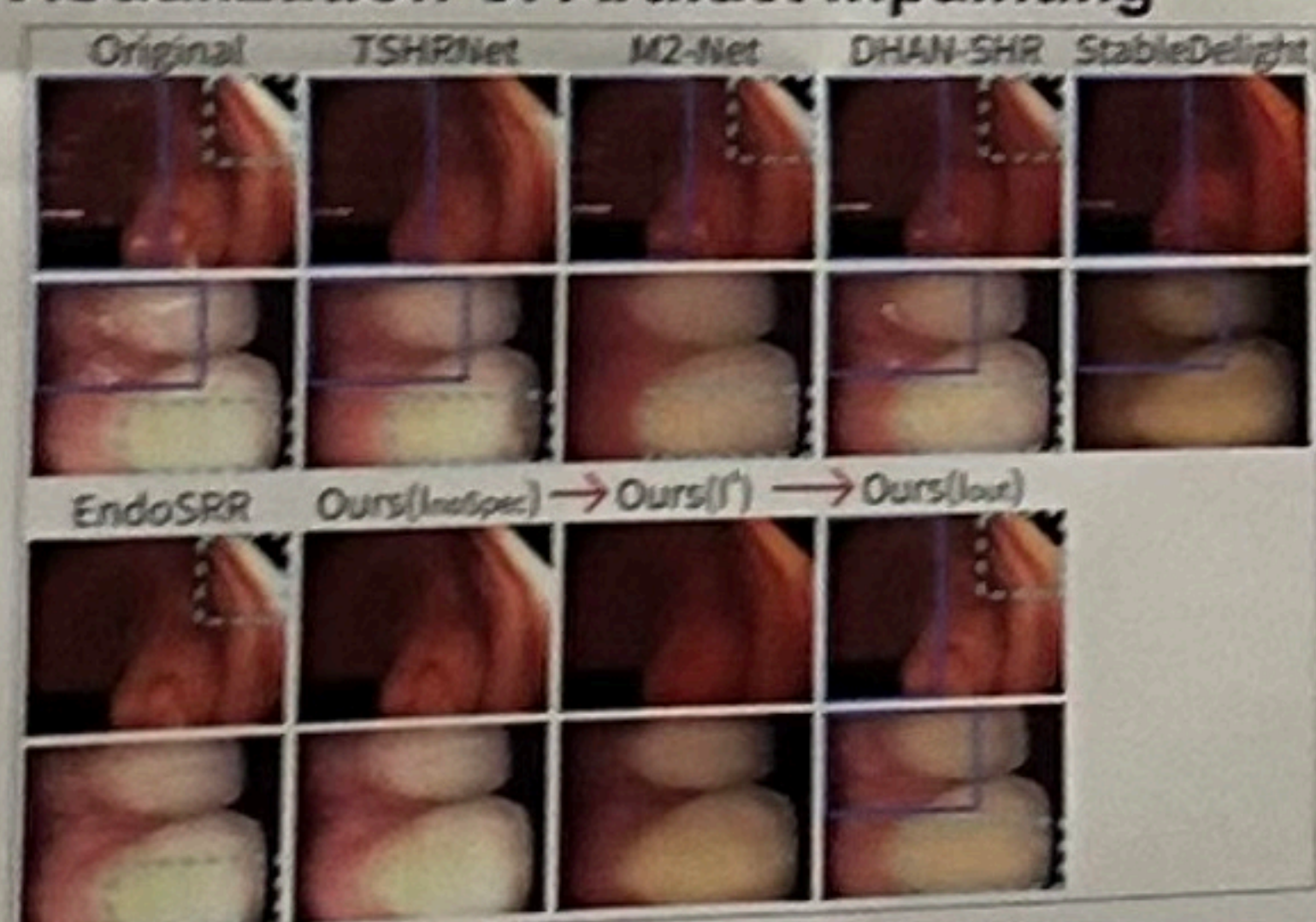
Phong Model



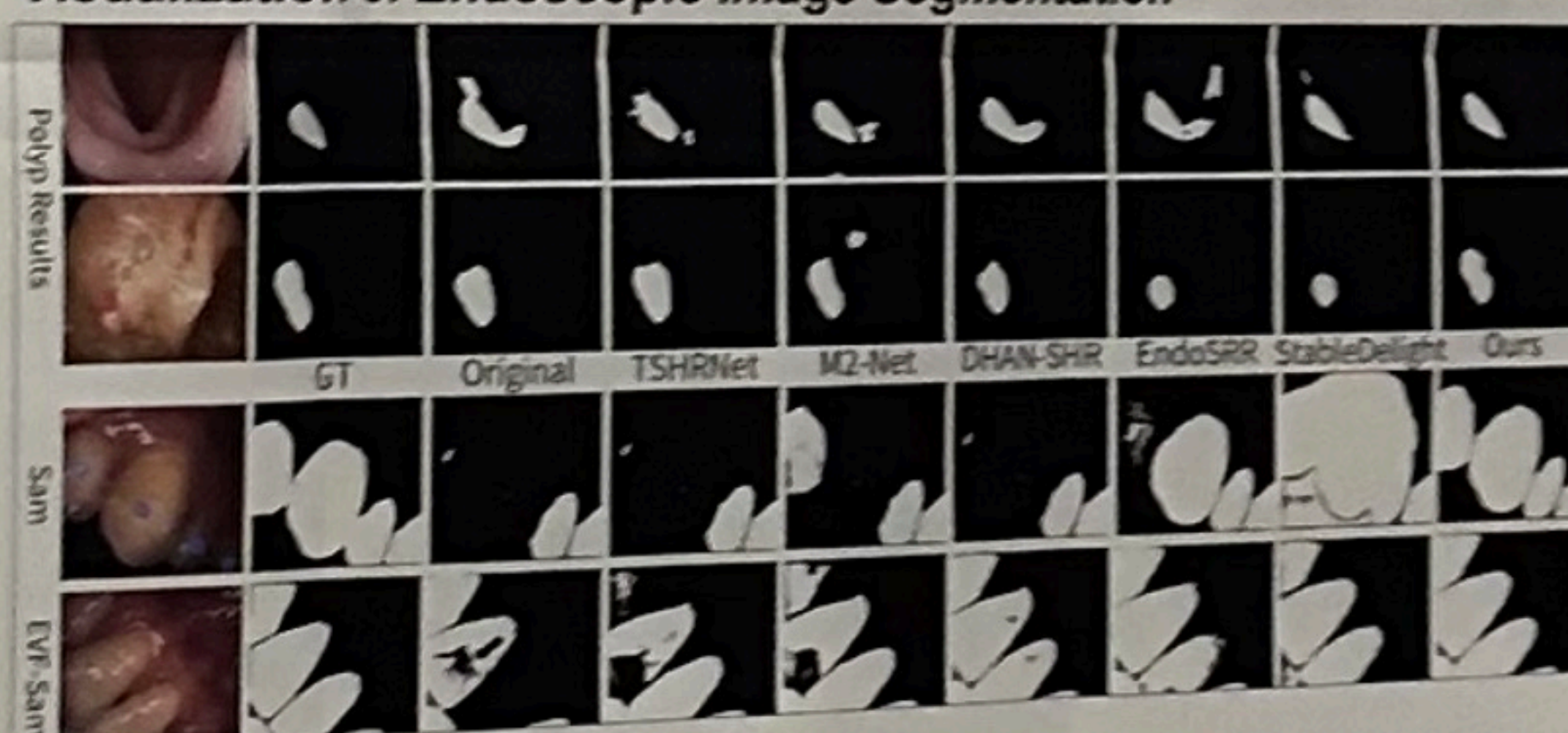
Process Pipeline



Visualization of Artifact Inpainting



Visualization of Endoscopic Image Segmentation



Quantitative comparison for segmentation performance enhancement

| Methods | Polyp Datasets | | | | | | | | Tooth Dataset | | | |
|--------------------|----------------|------------|-------------|------|--------------|-------------|-------|------|---------------|------|-------|------|
| | CVC-ClinicDB | Kvasir-SEG | CVC-ColonDB | ETIS | Teeth(Point) | Teeth(Text) | mDice | mIoU | mDice | mIoU | mDice | mIoU |
| Original Image | 93.2 | 88.8 | 92.0 | 87.0 | 77.4 | 69.0 | 79.3 | 71.3 | 51.5 | 39.3 | 84.1 | 74.0 |
| TSHRNet [10] | 93.0 | 88.4 | 91.4 | 86.2 | 79.0 | 69.7 | 73.6 | 65.2 | 73.5 | 62.1 | 80.6 | 70.0 |
| M2-Net [11] | 93.6 | 89.4 | 92.5 | 87.4 | 80.9 | 73.0 | 78.2 | 70.9 | 75.7 | 65.6 | 84.6 | 74.9 |
| DHAN-SHR [27] | 93.6 | 88.5 | 92.1 | 87.0 | 79.4 | 70.2 | 77.7 | 68.6 | 65.1 | 52.9 | 85.1 | 75.4 |
| EndoSRR [7] | 94.1 | 89.2 | 92.7 | 87.6 | 79.6 | 70.0 | 79.5 | 71.4 | 95.8 | 92.5 | 86.3 | 77.3 |
| StableDelight [13] | 93.4 | 89.0 | 91.3 | 86.0 | 76.2 | 69.0 | 69.2 | 62.4 | 91.4 | 85.6 | 77.9 | 66.8 |
| Ours | 95.0 | 90.6 | 93.3 | 88.3 | 81.5 | 73.4 | 81.0 | 73.2 | 96.1 | 93.0 | 86.8 | 78.0 |

Ablation on key components for artifact inpainting

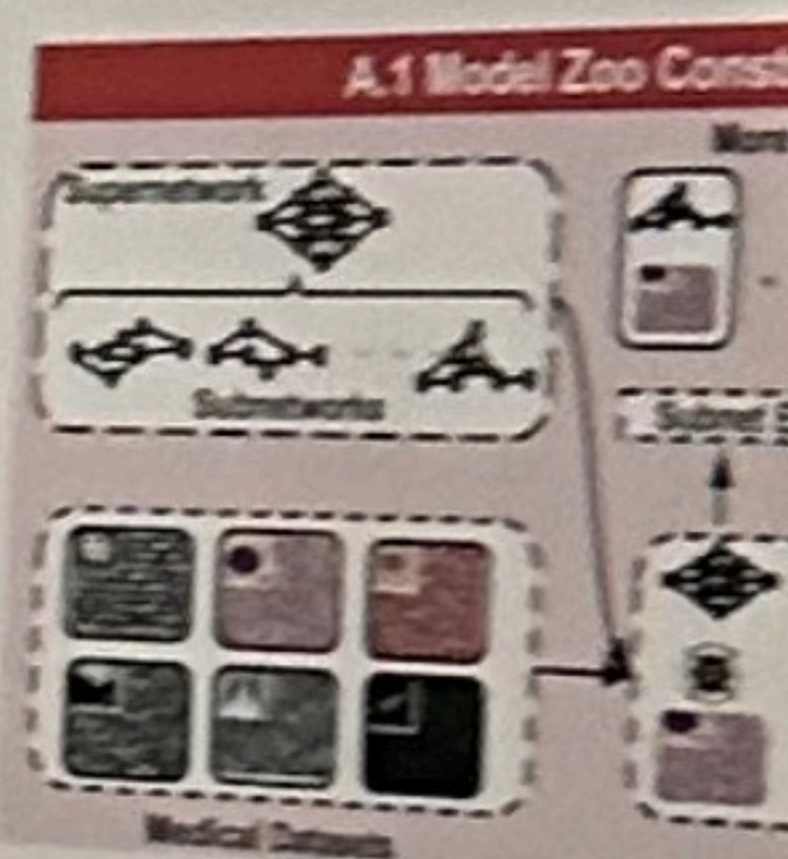
| Modules | | Datasets | | | | | | | | Teeth(Point) | | | | Teeth(Text) | | | |
|----------------|----------|--------------|------------|-------------|------|--------------|-------------|-------|------|--------------|------|-------|------|-------------|------|-------|------|
| | | CVC-ClinicDB | Kvasir-SEG | CVC-ColonDB | ETIS | Teeth(Point) | Teeth(Text) | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU |
| Specular | Diffuse | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU | mDice | mIoU |
| detect Inpaint | SD Guide | 93.2 | 88.8 | 92.0 | 87.0 | 77.4 | 69.0 | 79.3 | 71.3 | 51.5 | 39.3 | 84.1 | 74.0 | 73.5 | 62.1 | 80.6 | 70.0 |
| × | × | 94.5 | 89.9 | 92.3 | 87.2 | 78.4 | 69.5 | 76.5 | 68.1 | 49.5 | 37.4 | 83.6 | 73.2 | 75.7 | 65.6 | 84.6 | 74.9 |
| × | × | 94.6 | 90.0 | 93.0 | 88.0 | 81.8 | 73.8 | 80.0 | 72.2 | 95.8 | 92.7 | 86.5 | 77.5 | 65.1 | 52.9 | 85.1 | 75.4 |
| × | × | 94.4 | 89.6 | 91.1 | 86.4 | 76.8 | 69.7 | 77.7 | 70.5 | 94.5 | 90.5 | 83.4 | 73.6 | 91.4 | 85.6 | 77.9 | 66.8 |
| × | × | 95.0 | 90.6 | 93.3 | 88.3 | 81.5 | 73.4 | 81.0 | 73.2 | 96.1 | 93.0 | 86.8 | 78.0 | 96.1 | 93.0 | 86.8 | 78.0 |



B248 (Thu-PM)



- Adapting DL models to medical challenges: Architecture initialization.
- Transfer learning from natural but often ineffective due to differences between natural and medical data.
- Neural Architecture Search automated design but lacks a systematic way for selecting effective weights.



- A. Training Phase: Meta-Space Construction
- 1. Create a model zoo with 720K pairs efficiently using Supernet.
- 2. Encode models and datasets into latent space.
- 3. Refine the latent space with a co-evolutionary algorithm.

III. Experiments

- Protocol: Cross-validation strategy built using all datasets except the one

| Method | Paramm | — | Blood |
|-----------|--------|------|-------|
| DeepNet22 | 92.6 | 95.2 | 91.6 |
| ResNet18 | 92.1 | 95.2 | 91.6 |
| NASNet | 92.9 | 94.3 | 91.5 |
| PrunedNAS | 92.7 | 94.1 | 91.4 |
| MedNAS | 92.8 | 96.2 | 95.8 |
| MedNAS | 92.8 | 96.2 | 95.8 |

Faster Convergence + Better Main Result