Weak-Mamba-UNet

Marijan Sorić
Žan Stanonik

NTNU – Computer Vision and Deep Learning – TDT4265

SOTA Paper Challenge

Article presentation

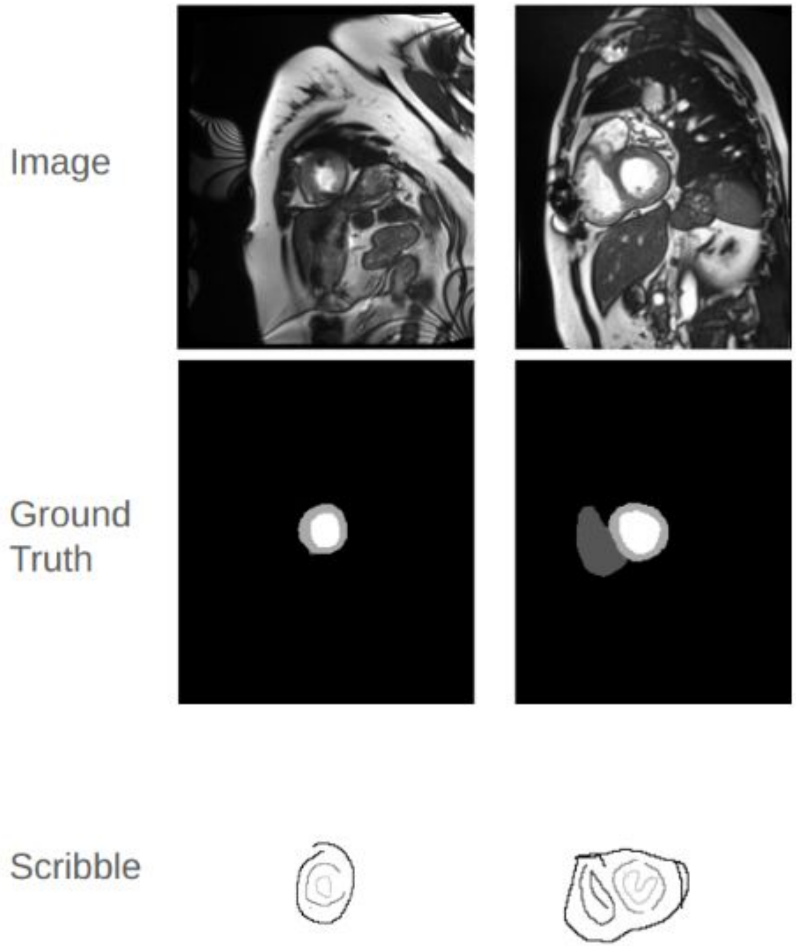
- Medical Image Segmentation
- Preprint: 16 Feb 2024

Weak-Mamba-UNet: Visual Mamba Makes CNN and ViT Work Better for Scribble-based Medical Image Segmentation

Ziyang Wang[1] and Chao Ma[2]

[1] Department of Computer Science, University of Oxford, UK
[2] Mianyang Visual Object Detection and Recognition Engineering Center, China
ziyang.wang@cs.ox.ac.uk
https://github.com/ziyangwang007/Mamba-UNet

Introduction



- Method: CNN, ViT & Visual Mamba
- Combines deep learning with the efficiency of WSL
- Architectures benefit from each other

Results

Benchmarks

- Outperforms single architectures
- Showcases advantages of:
  - limited supervision
  - limited resources

Table 2. Ablation Studies on Different Combinations of Segmentation Backbone Networks with the Same WSL Framework.

Network	Dice↑	Acc↑	Pre↑	Sen↑	Spe↑	HD↓	ASD↓
3×UNet	0.9141	0.9959	0.8958	0.9383	0.9927	8.0566	2.8806
3×SwinUNet	0.7446	0.9791	0.6555	0.9142	0.9815	121.4224	51.4317
3×MambaUNet	0.9128	0.9958	0.8931	<u>0.9395</u>	<u>0.9932</u>	8.3386	2.7928
UNet+SwinUNet+MambaUNet(Ours)	<u>0.9171</u>	<u>0.9963</u>	<u>0.9095</u>	0.9309	0.9920	<u>3.9597</u>	<u>0.8810</u>

The main idea

- Three distinct architectures

$$\mathbf{Y}_{\text{pseudo}} = \alpha \times f_{\text{cnn}}(\mathbf{X};\theta) + \beta \times f_{\text{vit}}(\mathbf{X};\theta) + \gamma \times f_{\text{mamba}}(\mathbf{X};\theta)$$

$$\alpha + \beta + \gamma = 1$$

- Multi-view cross-SL

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{3}(\mathcal{L}^{i}_{\text{pce}} + \mathcal{L}^{i}_{\text{dice}})$$

- Overall loss
  - Scribble-based

$$\mathcal{L}_{\text{pce}} = -\sum_{i \in \Omega_L}\sum_{k} y_{\text{s}}[i,k]\log(y_{\text{p}}[i,k])$$

(Pixels annotated with scribbles)

  - Dense-signal pseudo label

$$\mathcal{L}_{\text{dice}} = \text{Dice}\big(\text{argmax}(f(\boldsymbol{X};\theta), \boldsymbol{Y}_{\text{pseudo}})\big)$$

Architecture

Overview of the model

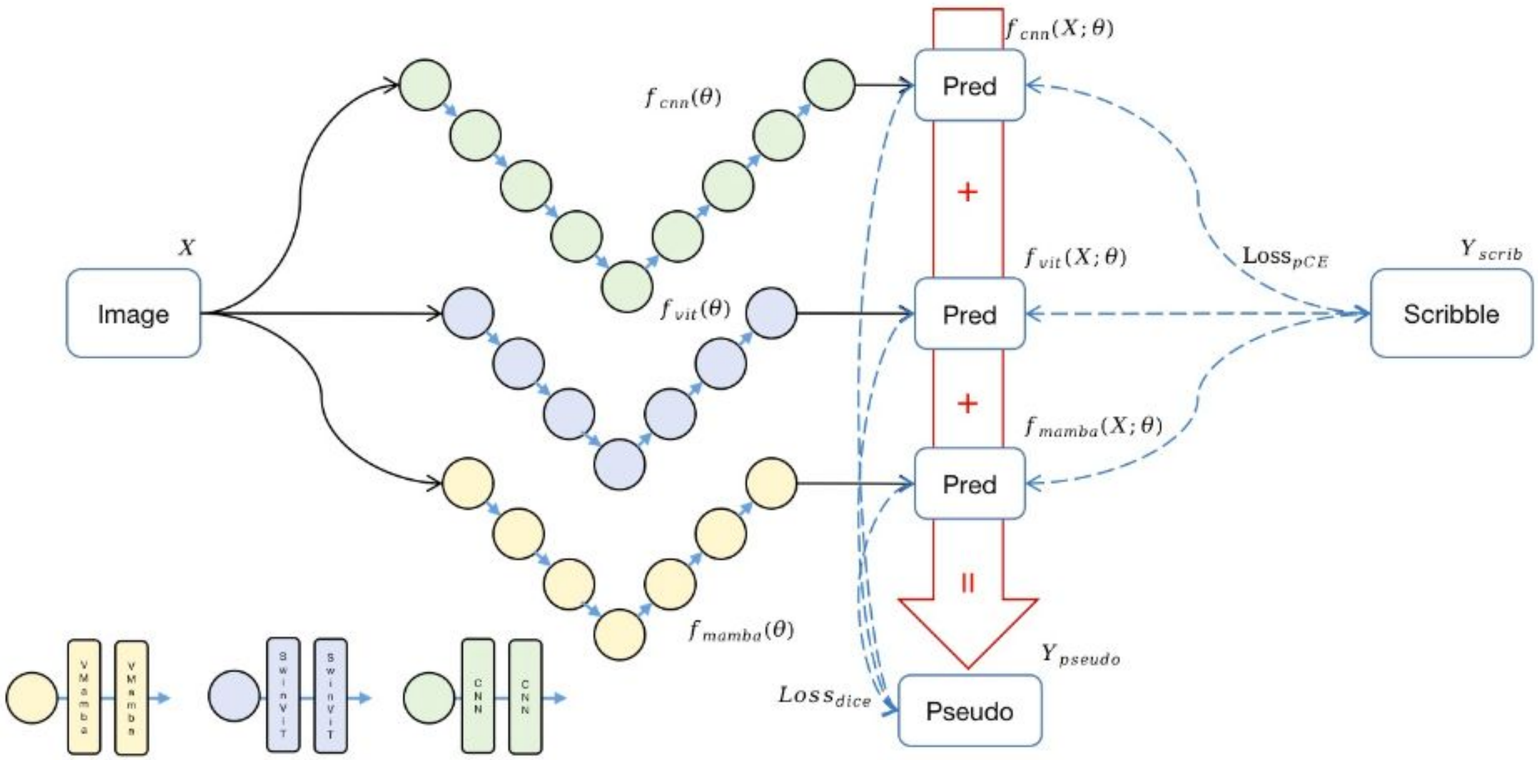


Fig. 2. Semi-Mamba-UNet: The Framework of Contrastive Cross-Supervised Visual Mamba-based UNet for Semi-Supervised Medical Image Segmentation.

CNN-based UNet

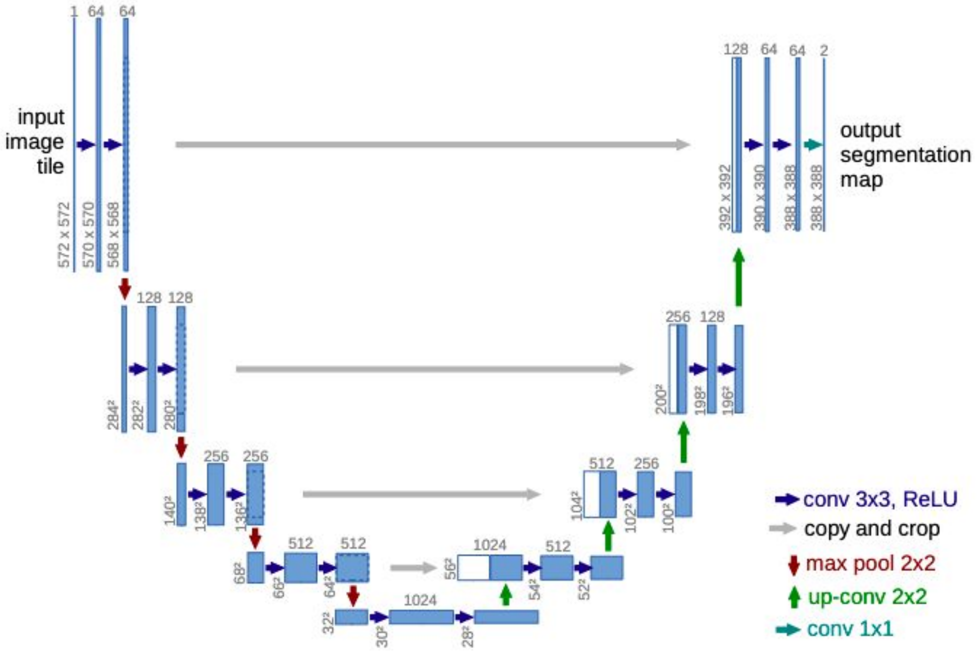


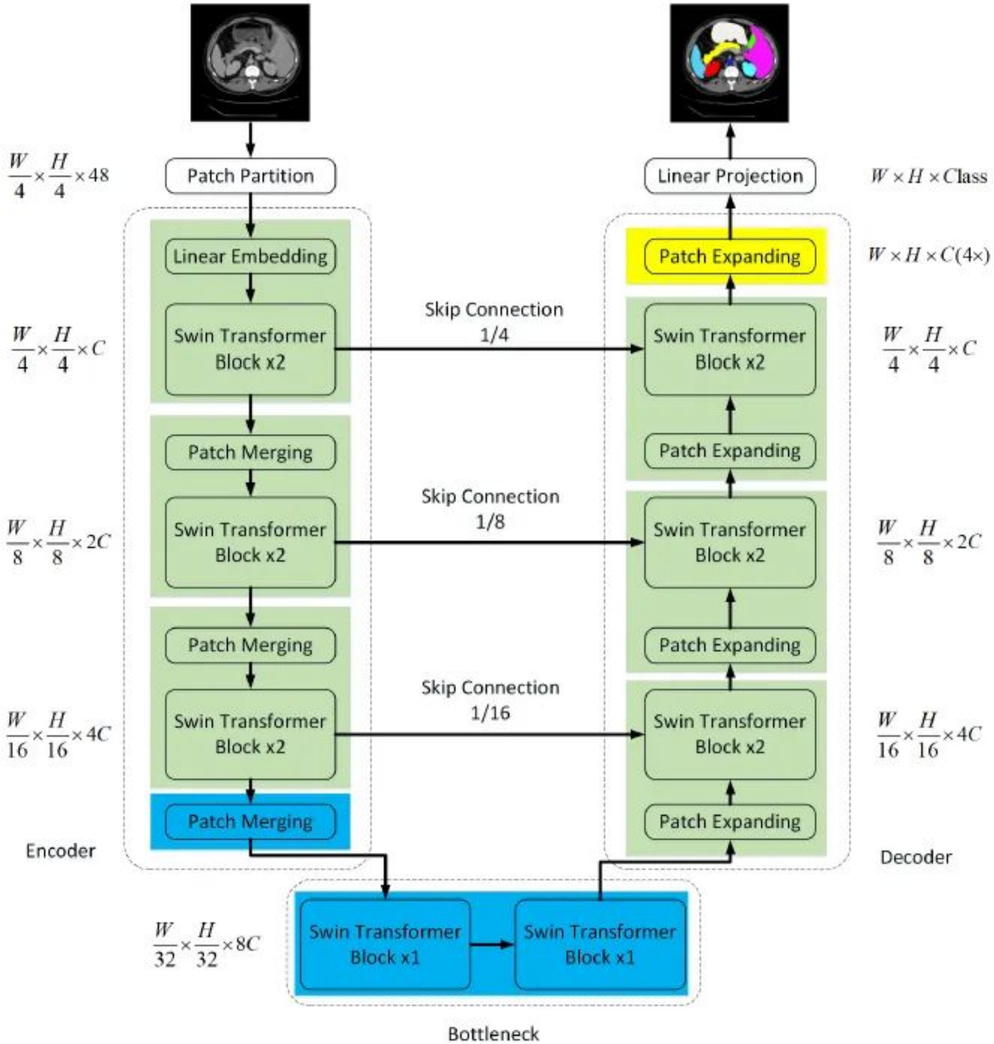
Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

- Encoder: What?
- Decoder: Where?
- Skip connection

SwinUNet

- Variant of UNet
- Swin captures long-range dependencies
- Image is broken down to patches
- UNet decoder-encoder condenses information
- Ideal for medical image segmentation

SwinUNet

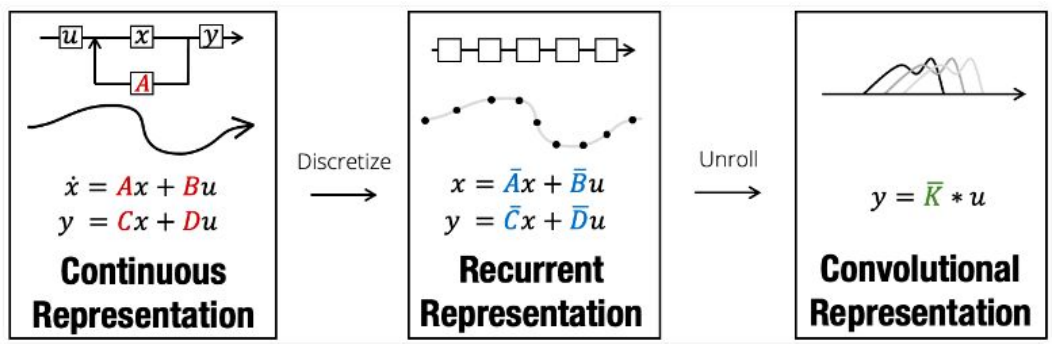


Patch Partition
Linear Projection
$\frac{W}{4} \times \frac{H}{4} \times 48$
$W \times H \times Class$
Linear Embedding
Patch Expanding
$W \times H \times C(4\times)$
Swin Transformer Block x2
Skip Connection 1/4
$\frac{W}{4} \times \frac{H}{4} \times C$
Patch Merging
Skip Connection 1/8
$\frac{W}{8} \times \frac{H}{8} \times 2C$
Skip Connection 1/16
$\frac{W}{16} \times \frac{H}{16} \times 4C$
Encoder
Decoder
$\frac{W}{32} \times \frac{H}{32} \times 8C$
Swin Transformer Block x1
Bottleneck

Mamba explained

RNN → SSM → S4 →S6 *(Mamba)*

- State Space Models
- Linear computation



$$h'(t) = Ah(t) + Bx(t) \quad \text{(1a)}$$
$$y(t) = Ch(t) \quad \text{(1b)}$$

$$h_t = \overline{A}h_{t-1} + \overline{B}x_t \quad \text{(2a)}$$
$$y_t = Ch_t \quad \text{(2b)}$$

$$\overline{K} = (C\overline{B}, C\overline{AB}, \dots, C\overline{A}^k\overline{B}, \dots) \quad \text{(3a)}$$
$$y = x * \overline{K} \quad \text{(3b)}$$

Algorithm 1 SSM (S4)

Input: $x : (\mathtt{B}, \mathtt{L}, \mathtt{D})$
Output: $y : (\mathtt{B}, \mathtt{L}, \mathtt{D})$
1: $A : (\mathtt{D}, \mathtt{N}) \leftarrow \mathsf{Parameter}$
▷ Represents structured $N \times N$ matrix
2: $B : (\mathtt{D}, \mathtt{N}) \leftarrow \mathsf{Parameter}$
3: $C : (\mathtt{D}, \mathtt{N}) \leftarrow \mathsf{Parameter}$
4: $\Delta : (\mathtt{D}) \leftarrow \tau_\Delta(\mathsf{Parameter})$
5: $\overline{A}, \overline{B} : (\mathtt{D}, \mathtt{N}) \leftarrow \mathsf{discretize}(\Delta, A, B)$
6: $y \leftarrow \mathsf{SSM}(\overline{A}, \overline{B}, C)(x)$
▷ Time-invariant: recurrence or convolution
7: **return** y

Algorithm 2 SSM + Selection (S6)

Input: $x : (\mathtt{B}, \mathtt{L}, \mathtt{D})$
Output: $y : (\mathtt{B}, \mathtt{L}, \mathtt{D})$
1: $A : (\mathtt{D}, \mathtt{N}) \leftarrow \mathsf{Parameter}$
▷ Represents structured $N \times N$ matrix
2: $B : (\mathtt{B}, \mathtt{L}, \mathtt{N}) \leftarrow s_B(x)$
3: $C : (\mathtt{B}, \mathtt{L}, \mathtt{N}) \leftarrow s_C(x)$
4: $\Delta : (\mathtt{B}, \mathtt{L}, \mathtt{D}) \leftarrow \tau_\Delta(\mathsf{Parameter} + s_\Delta(x))$
5: $\overline{A}, \overline{B} : (\mathtt{B}, \mathtt{L}, \mathtt{D}, \mathtt{N}) \leftarrow \mathsf{discretize}(\Delta, A, B)$
6: $y \leftarrow \mathsf{SSM}(\overline{A}, \overline{B}, C)(x)$
▷ Time-varying: recurrence (*scan*) only
7: **return** y

Mamba-based MambaUNet

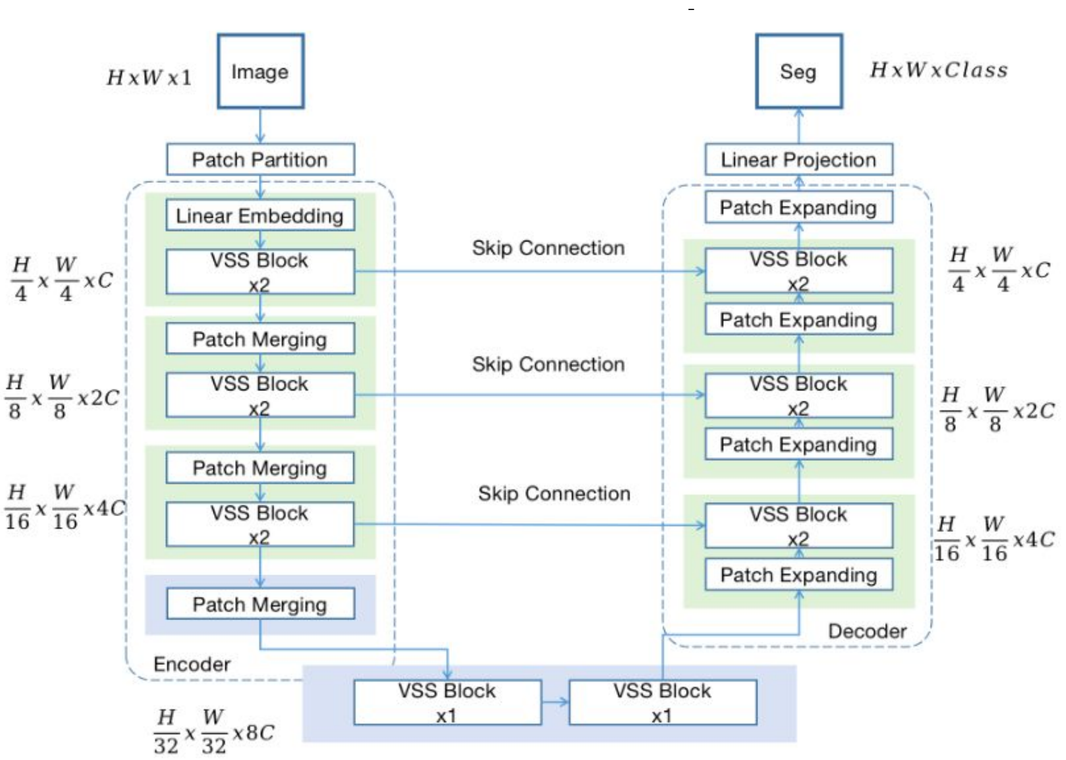
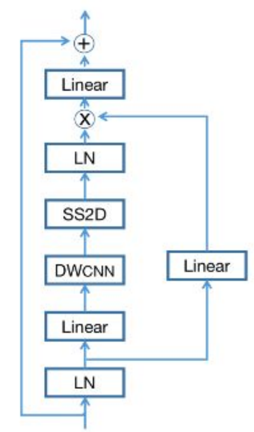


Fig. 2. The architecture of Mamba-UNet, which is composed of encoder, bottleneck, decoder and skip connections. The encoder, bottleneck and decoder are all constructed based on Visual Mamba block.

- Encoder-Decoder architecture
- Efficient long-range dependency modelling



VSS Block

Conclusion

- Reduces the cost & resources required for annotation
- Different algorithms completed each other
- Simpler form of annotation
- Can be applied to a variety of ML analysis tasks

Sources

Papers:

- Weak-Mamba-UNet: Visual Mamba Makes CNN and ViT Work Better for Scribble-based Medical Image Segmentation
- U-Net: Convolutional Networks for Biomedical Image Segmentation
- Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation
- Mamba: Linear-Time Sequence Modeling with Selective State Spaces
- Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation

Blogs:

- Introduction to State Space Models (SSM) – *Hugging Face*