NExNet Seg: Neuron Expansion Network for Medical Image Segmentation

Supplementary Material

8. Difference between dPEN, PEN, and T-PEN

The Progressively Expanded Neuron (PEN) [28] uses the Maclaurin series with fixed coefficients to expand neurons, limiting adaptability. Deep PEN (dPEN) [29] introduces dynamic but non-trainable adjustments. T-PEN, proposed here, enhances flexibility by making $\mathbf{w}_k$ and $\mathbf{p}_k$ trainable, optimizing feature extraction for diverse tasks.

9. Approximation with T-PEN

T-PEN enhances feature maps by expanding Φ into layers $\mathbf{S}^{(k)}$ with trainable parameters (Equation (1) of the main manuscript):

$$\mathbf{S}^{(k)} = \begin{cases} g(\mathbf{w}_1 \Phi^{\mathbf{p}_1}) & k = 1 \\ g(\mathbf{S}^{(k-1)} + \mathbf{w}_k \Phi^{\mathbf{p}_k}) & k > 1 \end{cases}$$

The output Y_S is $BN(\text{Concatenate}(\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(K)}))$ with $K = 3$. Figure 10 compares T-PEN's approximation of five functions over epochs 1 to 4000, with each pair of rows (a: 1–2, b: 3–4, c: 5–6, d: 7–8, e: 9–10) showing the first row with T-PEN layers and the second with traditional convolutions:

- (a) $\sin(5x) + \cos(10y)$: T-PEN (row 1) converges by Epoch 4000; convolutions (row 2) are less accurate.
- (b) $\sin(5x)\cos(10y)$: T-PEN (row 3) matches by Epoch 4000; convolutions (row 4) lag.
- (c) $\exp\left(-\frac{x^2+y^2}{2}\right)\sin(5x)$: T-PEN (row 5) refines by Epoch 500, matching by 4000; convolutions (row 6) are less precise.
- (d) $\sin(xy) + \cos(x + y)$: T-PEN (row 7) converges by Epoch 4000; convolutions (row 8) are less effective.
- (e) $|xy|\sin(3(x + y))$: T-PEN (row 9) approximates accurately by Epoch 4000; convolutions (row 10) show reduced accuracy.

T-PEN's superior modeling of nonlinearities compared to traditional convolutions boosts NExNet Seg's segmentation accuracy (Table 1).

10. γ Value in MaSA

The Manhattan Self-Attention (MaSA) [30] uses a decay factor γ to control spatial attention, defined in Equations (4) and (5) of the main manuscript:

$$D_{H_{nm}} = \gamma^{|x_n - x_m|}, \quad D_{W_{nm}} = \gamma^{|y_n - y_m|}$$

A smaller γ (e.g., 0.1) emphasizes local features with steep decay, while a larger γ (e.g., 0.9) enables longer-range dependencies.

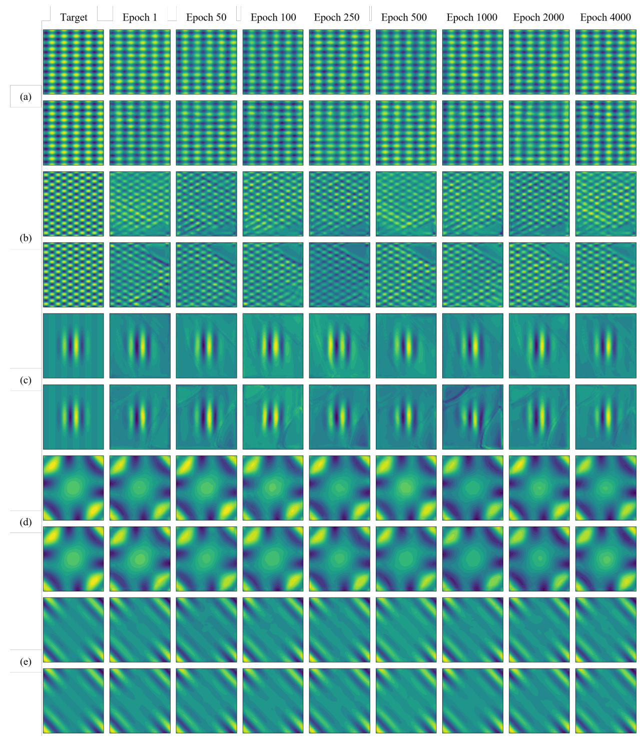

Figure 10. Approximation of target functions over epochs 1 to 4000: (a) rows 1–2 for $\sin(5x) + \cos(10y)$, (b) rows 3–4 for $\sin(5x)\cos(10y)$, (c) rows 5–6 for $\exp\left(-\frac{x^2+y^2}{2}\right)\sin(5x)$, (d) rows 7–8 for $\sin(xy) + \cos(x + y)$, (e) rows 9–10 for $|xy|\sin(3(x + y))$. Each pair's first row uses T-PEN layers; the second uses traditional convolutions.

Figure 11 shows γ's effect on skin lesion feature maps: sparse at 0.1, becoming global at 0.9. Figure 12 analyzes γ's impact on mean activation and standard deviation for ISIC 2016 (skin) and CVC Clinic (polyp) datasets, with variability peaking at $\gamma = 0.9$ (0.007 for skin, higher for polyps). While a higher γ (e.g.@ 0.9) can help capture broader context in both tasks, setting γ around 0.5 strikes a good balance between preserving local detail in skin lesion segmentation and maintaining stable feature maps for polyp segmentation.

11. Additional Results for Statistical Analysis

This section provides further statistical insights into NExNet Seg's performance compared to U-Net, complementing the main manuscript's findings.

These tables demonstrate NExNet Seg's consistent superiority in Dice Coefficient over U-Net (as baseline), with

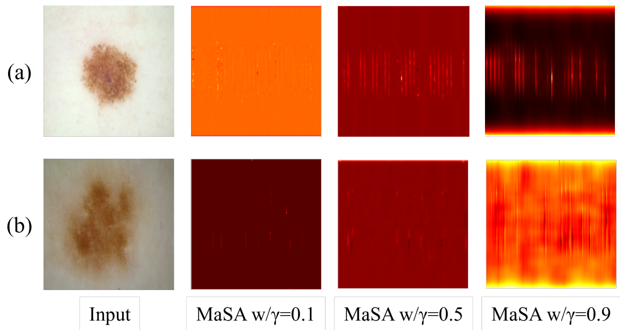

Figure 11. Effect of γ values (0.1, 0.5, 0.9) on MaSA feature maps for skin lesions.

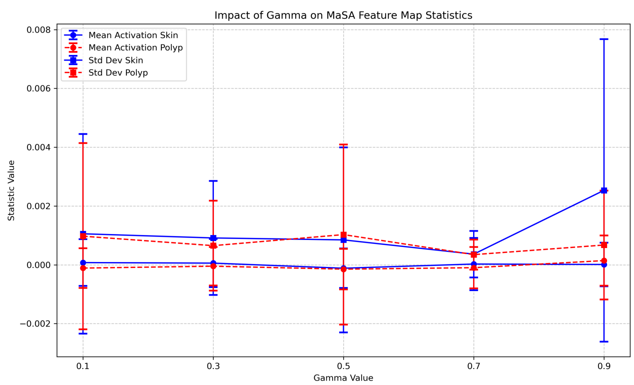

Figure 12. Statistical impact of γ (0.1–0.9) on MaSA feature map statistics for skin (ISIC 2016) and polyp (CVC Clinic) datasets.

Table 5. Performance comparison between NExNet Seg and U-Net in terms of mean Dice Coefficient (DC) and standard deviation (Std). Best results in bold.

Dataset	Method	Dice Coefficient	Std
ISIC 2018	NExNet Seg (Ours)	**0.848**	**0.003**
	U-Net	0.813	0.006
PH2	NExNet Seg	**0.926**	**0.002**
	U-Net	0.873	0.006
Kvasir-Seg	NExNet Seg	**0.939**	**0.003**
	U-Net	0.775	0.008
CVC-Clinic	NExNet Seg	**0.930**	**0.004**
	U-Net	0.856	0.012

statistically significant improvements ($p < 0.05$) on most datasets, except ISIC16 and CVC-Clinic, where the differences are not significant ($p > 0.05$).

12. Detailed Analysis of Limitations

While NExNet Seg performs well in medical image segmentation, several limitations require attention.

Overfitting with T-PEN Layers. The ablation study (Table 2) shows T-PEN layers cause overfitting, with IoU dropping on ISIC17 (0.701 to 0.694) and CVC Clinic (0.865 to 0.838) without MaSA or SSL, likely due to rapid parameter adjustments ($\mathbf{w}_k$, $\mathbf{p}_k$) in Equation (1). This is more pronounced in smaller datasets like ISIC17 (2,000 images) vs. ISIC18 (10,000 images) [5]. Regularization (e.g., dropout, L2) or enhanced augmentation (e.g., color jittering) could improve generalization.

Table 6. Statistical analysis comparing NExNet Seg and U-Net performance across different datasets using Dice coefficient. p-values below 0.05 (in bold) indicate statistically significant improvements.

Dataset	p-value	Std
ISIC16	0.056	0.004
ISIC17	**0.006**	0.006
ISIC18	**0.029**	0.004
PH2	**0.009**	0.003
CVC-Clinic	0.065	0.004
Kvasir	**0.017**	0.004

Dataset-Specific Performance Variations. NExNet Seg lags behind SegFormer (IoU 0.905 vs. 0.876 on CVC Clinic, Table 1), possibly due to MaSA's horizontal-vertical decomposition (Equations (3)–(5)) struggling with irregular polyp shapes. Hybrid attention or T-PEN adjustments ($\mathbf{S}^{(k)}$) and γ fine-tuning could enhance performance.

Generalization to Other Medical Imaging Tasks. Evaluation is limited to skin (ISIC, PH^2) and polyp (Kvasir-Seg, CVC-Clinic) datasets. Generalizing to MRI, CT, or ultrasound may be hindered by T-PEN's nonlinear expansions and MaSA's 2D design.

Hardware Dependency and Clinical Deployment. Training requires an NVIDIA A-100 (40 GB), posing a barrier for resource-limited clinics with low-memory devices (4–8 GB). With 30.4 million parameters and a batch size of 16 over 150 epochs, pruning, quantization, or pre-trained fine-tuning could enhance accessibility.