Deconver for Stroke Lesion Segmentation

Niels Vyncke[1], Shahryar Noei[2], Giuseppe Jurman[2], Aleksandra Pižurica[1], and Pooya Ashtari[1]

[1] Department of Telecommunications and Information Processing (TELIN), Ghent University, B-9000 Ghent, Belgium
pooya.ashtari@ugent.be (corresponding author)
[2] Data Science for Health Unit, Fondazione Bruno Kessler, Via Sommarive 18, Povo, 38123 Trento, Italy

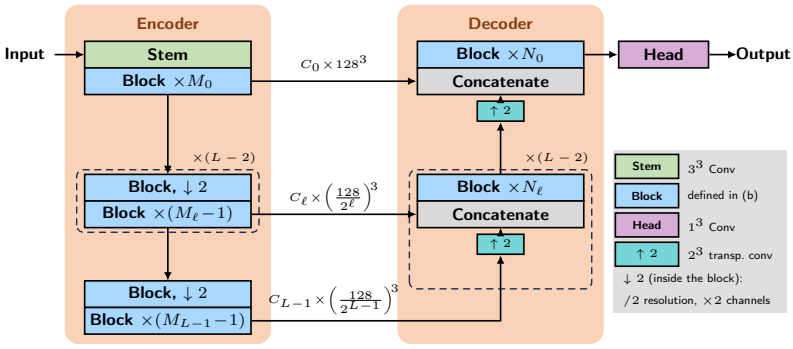
Abstract. ISLES'26 poses ischemic stroke lesion segmentation on native-space T1-weighted MRI acquired at more than 60 centers and covering acute, sub-acute, and chronic lesions. We evaluate Deconver, a U-shaped network descended from Factorizer and built on a core block that replaces self-attention with a nonnegative deconvolution (NDC) operator, against nnU-Net, SegResNet, and SwinUNETR. Preprocessing, augmentation, optimization, folds, and inference are identical across models, so differences reflect architecture rather than pipeline engineering. In five-fold cross-validation on the ISLES'26 training cohort, Deconver reaches the highest mean Dice of the four, 63.6, ahead of SegResNet (63.0), nnU-Net (62.5), and SwinUNETR (62.4), at about 3× fewer parameters than either large baseline; a 10.6M-parameter variant matches SegResNet at over 7× fewer parameters, and a 6.5M variant still exceeds nnU-Net and SwinUNETR. Ablating encoder depth, initial width, and NDC kernel size identifies width as the dominant factor, finds no benefit from enlarging the NDC kernel beyond 3^3, and shows diminishing returns from further scaling. Ensembling Deconver with SegResNet, our final challenge submission, raises Dice to 64.4. Code and configurations: https://github.com/pashtari/deconver-isles26.

Keywords: U-Net · Vision Transformer · Stroke Lesion · Image Segmentation · Deconvolution · Deconver

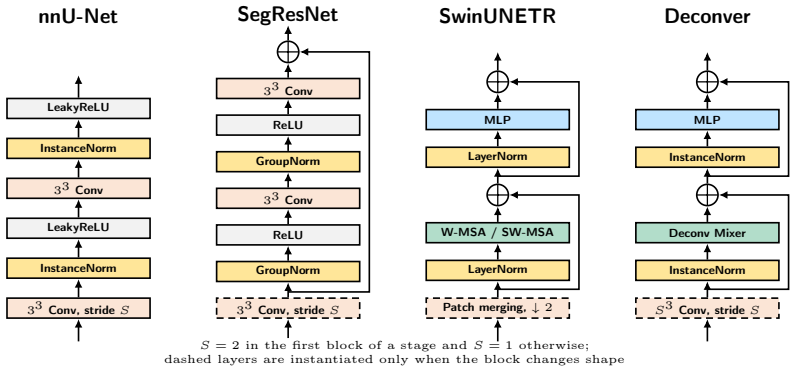
1 Introduction

Ischemic stroke is a leading cause of long-term neurological disability, and delineating infarcted tissue on MRI underpins the quantification of tissue damage and the development of imaging biomarkers of outcome. Manual annotation is slow and subject to inter-observer variability; automated segmentation reaches expert-level agreement on acute diffusion-weighted MRI [14], but remains difficult because lesions vary widely in size, location, morphology, and contrast, and that variability compounds across disease stages and sites.

The ISLES series has benchmarked stroke lesion segmentation since 2015; ISLES'22 provided multi-sequence acute-stage MRI with expert annotations [6].



(a) Shared L-level encoder–decoder backbone. Level ℓ carries C_ℓ channels and contains M_ℓ encoder and N_ℓ decoder blocks.



(b) Computational block of each architecture, with data flowing upwards.

Fig. 1. Evaluated architectures. Downsampling (↓ 2: half the resolution, twice the channels) happens *inside* the first block of each encoder stage, whereas upsampling (↑ 2) is a separate transposed convolution on the decoder transition. SwinUNETR instead downsamples by patch merging outside the block and uses convolutional decoder blocks.

ISLES'26 [13] changes the problem along three axes at once: a single structural modality—*T1-weighted* MRI—so the diffusion contrasts on which acute delineation usually relies are absent; a cohort spanning acute, sub-acute, and chronic lesions, whose T1-weighted appearance changes markedly with time since onset; and scans from more than 60 centers at native resolution and orientation. Generalization, rather than performance under harmonized acquisition, is the operative difficulty.

Most segmentation networks proposed since U-Net [12] keep its encoder–decoder topology with skip connections and differ mainly in the block applied at each resolution level (Fig. 1). We study *Deconver* [1], which keeps the transformer-style block layout but replaces self-attention with a Deconv Mixer built around a nonnegative deconvolution (NDC) layer. Deconver descends from Factorizer [2], which first showed that a nonnegative matrix factorization layer can stand in for self-attention in a U-shaped segmentation network and, as team

SWAN, ranked third in ISLES'22 [14]. Deconver keeps that nonnegativity prior but recasts the context operator as a deconvolution, capturing spatial dependencies while restoring high-frequency detail far more cheaply than attention. Our baselines span the dominant alternatives: nnU-Net [7] (plain convolutions), SegResNet [11] (residual convolutions), and SwinUNETR [4] (a hierarchical Swin Transformer [8] encoder whose shifted-window self-attention reaches beyond a convolution's local receptive field, at higher cost).

Published comparisons are hard to interpret: each method comes with its own preprocessing, augmentation, and optimization recipe, and often a different model size, so differences attributed to the block confound it with pipeline engineering and capacity. We therefore evaluate all four families under one fixed pipeline. Our contributions are threefold. (i) We report the first evaluation of Deconver on ISLES'26, in a *controlled* setting that isolates the effect of the core block. (ii) Of the four models compared, Deconver attains the best mean Dice using about a third of the parameters of the strongest baselines, and a 6.46M-parameter variant already outperforms nnU-Net and SwinUNETR. (iii) An ablation over encoder blocks, initial width, and NDC kernel size shows that width dominates, that a 3^3 deconvolution kernel suffices, and that returns from further scaling are small; ensembling Deconver with SegResNet, which errs differently, adds a further 0.85 Dice points and forms our challenge submission.

2 Methods

2.1 Data

We use the ISLES'26 training cohort [13]: native-space T1-weighted MRI with binary expert annotations of ischemic infarct, acquired at more than 60 centers and released without resampling or spatial normalization, so voxel spacing, field of view, and orientation vary. Lesions span the acute, sub-acute, and chronic stages.

2.2 Architectures and Configurations

All four networks instantiate the encoder–decoder template of Fig. 1a: an L-level encoder in which each stage halves the resolution and doubles the channel count, a symmetric decoder that upsamples with transposed convolutions, and skip connections concatenating encoder features into the decoder. Level ℓ carries C_ℓ channels and contains M_ℓ encoder and N_ℓ decoder blocks, with C_0 the initial width; Deconver caps its width at 512, so $C_\ell = \min(2^\ell C_0, 512)$. The families differ in the block repeated at each level (Fig. 1b).

Two deviations deserve emphasis. We adopt the nnU-Net *network topology* but not its self-configuring pipeline, so its row in Table 1 characterizes the plain convolutional block rather than the framework, whose principal strength is the automated configuration we hold fixed. We use SwinUNETR-V2 [5], which prepends a residual convolutional block to each Swin stage; its depth is fixed by

the Swin hierarchy, so we report feature size as C_0 and count only Swin blocks in $[M_\ell]$.

Each family keeps its customary normalization—instance for nnU-Net and Deconver, group with eight groups for SegResNet, and layer normalization in SwinUNETR-V2's transformer blocks—and each baseline is reported at one representative configuration (Table 1). For Deconver we additionally sweep the encoder blocks per level $[M_\ell]$, the initial width C_0, and the NDC kernel size k (Table 2). Decoder depth is N_ℓ=1 at every level and all convolutional kernels are 3^3 unless stated otherwise.

2.3 Deconver

A Deconver block [1] keeps the two-sub-module layout of a vision-transformer block but substitutes a Deconv Mixer for self-attention and instance normalization (IN) for layer normalization, which suits small 3D batch sizes:

$$\mathcal{Z} = \mathrm{DeconvMixer}(\mathrm{IN}(\mathcal{X})) + \mathcal{X}, \qquad \mathcal{Y} = \mathrm{MLP}(\mathrm{IN}(\mathcal{Z})) + \mathcal{Z}, \tag{1}$$

where the MLP is two pointwise convolutions separated by a GELU. The mixer projects, enforces nonnegativity, applies the NDC layer, and projects back:

$$\mathrm{DeconvMixer}(\mathcal{X}) = \mathrm{PWConv}\big(\mathrm{NDC}(\mathrm{ReLU}(\mathrm{PWConv}(\mathcal{X})))\big). \tag{2}$$

The NDC layer treats its nonnegative input $\mathcal{X}$ as a blurred observation and recovers a latent source by minimizing $\|\mathcal{X} - \mathcal{S} * \mathcal{V}\|_{\mathrm{F}}^2$ over $\mathcal{S} \geq 0$, with $\mathcal{V} \geq 0$ a learnable filter and $*$ multi-channel cross-correlation with a k^3 kernel. From a learned nonnegative initialization $\mathcal{S}^{(0)}$, one multiplicative Richardson–Lucy step gives

$$\mathcal{S}^{(1)} = \mathcal{S}^{(0)} \odot \frac{\mathcal{X} * \mathcal{V}^- + \epsilon}{\big(\mathcal{S}^{(0)} * \mathcal{V}\big) * \mathcal{V}^- + \epsilon}, \tag{3}$$

with $\odot$ the elementwise product, $\mathcal{V}^-$ the adjoint (transposed and spatially flipped) filter, and ϵ=10^{-8}. The update provably does not increase the reconstruction error, preserves nonnegativity, and is differentiable, so $\mathcal{V}$—shared with $\mathcal{V}^-$—is trained end to end. It runs over G channel groups and expands the source by a ratio R; we use depthwise groups, R=4, and one iteration throughout.

2.4 Preprocessing and Data Augmentation

All images were processed with a common pipeline implemented in MONAI [3]: cropping to the intensity-based foreground bounding box with a 10-voxel margin, reorientation to RAS, and resampling to an isotropic spacing of $1 \times 1 \times 1\,\mathrm{mm}^3$ (linear interpolation for images, nearest-neighbor for masks). Intensities were normalized per channel using the mean and standard deviation of the non-zero voxels, and volumes smaller than the network input were zero-padded.

During training, $128 \times 128 \times 128$ voxel patches were sampled from the preprocessed volumes. Each augmentation was applied independently with probability

Table 1. Five-fold cross-validation on ISLES'26; notation as in Sect. 2. nnU-Net and SegResNet Dice average two runs. Best Dice in bold, second best underlined.

Model	$[M_\ell]$	$[N_\ell]$	C_0	Params (M)	Dice (%) ↑
nnU-Net	$[1, 1, 1, 1, 1]$	$[1, 1, 1, 1]$	32	22.57	62.54
SegResNet	$[1, 2, 2, 4]$	$[1, 1, 1]$	64	75.17	63.00
SwinUNETR-V2	$[2, 2, 2, 2]$	$[1, 1, 1, 1, 1]$	48	72.76	62.41
Deconver	$[1, 2, 2, 4]$	$[1, 1, 1]$	64	25.46	<u>63.59</u>
Ensemble (SegResNet + Deconver)				100.63	**64.44**

0.2: random affine transformations (rotations up to 0.26 rad, ≈15°, about each axis; isotropic scaling within ±20%), Gaussian noise (σ=0.1), Gaussian smoothing ($\sigma \in [0.5, 1.0]$), intensity scaling from $[0.7, 1.3]$, and intensity shifts up to 0.1; random flips along each axis used probability 0.5. Validation used no augmentation.

2.5 Training and Evaluation

All models were trained for 300 epochs with a batch size of two, optimized with AdamW [9] at an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}, under a warm-up cosine schedule (linear increase over three epochs, then cosine annealing). The objective summed a soft Dice loss [10] and cross-entropy, and was held fixed across all architectures and configurations.

We used five-fold cross-validation with folds stratified by lesion volume at the subject level. The same fold assignment was used for every architecture and configuration, so all comparisons are paired across identical data splits. Predictions were obtained by sliding-window inference with $128 \times 128 \times 128$ windows, and we report the mean Dice coefficient over the validation cases of all five folds.

Our final challenge submission ensembles the two strongest architectures, averaging the softmax maps of SegResNet and Deconver voxelwise before the argmax. As every model is trained per fold, the deployed ensemble holds ten networks (five folds × two architectures); the cross-validated Dice in Table 1 averages, per validation case, only the two networks whose fold excluded it.

3 Results and Discussion

Comparison with the baselines. Deconver attains the highest mean Dice of the four single models in Table 1, 63.59, ahead of SegResNet (63.00), nnU-Net (62.54), and SwinUNETR-V2 (62.41). The 0.59-point margin over the strongest baseline comes at 2.95× fewer parameters (25.46M against 75.17M); against SwinUNETR-V2 the gap widens to 1.18 points, again at 2.9× fewer. Only nnU-Net is comparable in size (22.57M), and it trails by 1.05 points. Accuracy and capacity are thus not aligned: the largest model is not the most accurate, nor the smallest the least accurate. Substituting nonnegative deconvolution for residual

Table 2. Deconver ablation on ISLES'26, ordered by parameter count. Best Dice in bold, second best underlined.

$[M_\ell]$	$[N_\ell]$	C_0	k	Params (M)	Dice (%) ↑
$[1, 2, 2, 4]$	$[1, 1, 1]$	32	3	6.46	62.76
$[1, 1, 1, 1, 1]$	$[1, 1, 1, 1]$	32	3	10.55	62.99
$[1, 1, 1, 1, 1]$	$[1, 1, 1, 1]$	32	5	11.13	62.85
$[1, 1, 1, 1, 1]$	$[1, 1, 1, 1]$	64	3	24.23	63.40
$[1, 1, 1, 1, 1, 1]$	$[1, 1, 1, 1, 1]$	32	3	24.30	62.91
$[1, 2, 2, 4]$	$[1, 1, 1]$	64	3	25.46	**63.59**
$[1, 2, 2, 4, 4]$	$[1, 1, 1, 1]$	64	3	52.77	<u>63.56</u>

convolution or shifted-window attention saves parameters outright; the accuracy margin is smaller, and the SegResNet comparison is not normalization-matched.

Ablation. *Width* is by far the most effective knob in Table 2: doubling C_0 from 32 to 64 raises Dice by 0.83 at $[M_\ell]$=$[1, 2, 2, 4]$ ($62.76 \rightarrow 63.59$) and by 0.41 at $[1, 1, 1, 1, 1]$ ($62.99 \rightarrow 63.40$). *Depth* is not: a sixth level at C_0=32 costs 2.3× the parameters for −0.08 Dice, and a fifth level at C_0=64 doubles them for −0.03 Dice. *Enlarging the NDC kernel does not help*: k=5 loses 0.14 Dice against k=3 while adding parameters, so a 3^3 deconvolution already spans the useful support. The three C_0=64 rows, spanning 24.23–52.77M, differ by only 0.19 Dice, so we report the 25.46M model rather than the largest.

Efficiency. The compact variants are more striking still: 10.55M matches SegResNet (62.99 against 63.00) at 7.1× fewer parameters, and 6.46M still beats nnU-Net and SwinUNETR-V2 at 3.5× and 11.3× fewer; every row of Table 2 outperforms both.

Ensemble. Averaging the Deconver and SegResNet probability maps reaches 64.44 Dice, 0.85 points above the better of the two and most of the 1.18-point spread between the best and worst single models. No scaling bought as much: all of Table 2 spans 0.83 points across an eightfold parameter range. The two blocks evidently err partly independently—something an ensemble exploits and a larger network cannot.

4 Conclusion

Under one fixed pipeline on ISLES'26, Deconver gives the best mean Dice of the four models compared (63.59 against 63.00, 62.54, 62.41) on roughly a third of the parameters of the two large baselines; a 10.55M variant already matches SegResNet. The ablation locates the gain in width rather than depth or kernel size, with small returns from further scaling, and combining the two best blocks pays more than enlarging either—the Deconver–SegResNet ensemble (64.44 Dice) is our challenge submission. Lesion-wise metrics, per-fold tests, and center-held-out splits come next.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ashtari, P., Noei, S., Haredasht, F.N., Chen, J.H., Jurman, G., Pižurica, A., Van Huffel, S.: Deconver: A deconvolutional network for medical image segmentation. IEEE Journal of Biomedical and Health Informatics (2025)
2. Ashtari, P., Sima, D.M., De Lathauwer, L., Sappey-Marinier, D., Maes, F., Van Huffel, S.: Factorizer: A scalable interpretable approach to context modeling for medical image segmentation. Medical Image Analysis **84**, 102706 (2023). https://doi.org/10.1016/j.media.2022.102706
3. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
4. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2022)
5. He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., Xu, D.: SwinUNETR-V2: Stronger Swin transformers with stagewise convolutions for 3D medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 416–426. Springer (2023)
6. Hernandez Petzsche, M.R., de la Rosa, E., Hanning, U., Wiest, R., Valenzuela, W., Reyes, M., Meyer, M.I., Liew, S.L., Kofler, F., Ezhov, I., et al.: ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Scientific Data **9**(1), 762 (2022)
7. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
8. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
11. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: International MICCAI Brainlesion Workshop. pp. 311–320. Springer (2019)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241 (2015)
13. de la Rosa, E., Menze, B., Deseoe, J., Wiestler, B., Kirschke, J., Yalcin, Ç., Cabezas, M., Rorden, C., Khan, M., Liew, S.L., Reyes, M., Wiest, R., Su, R.: Ischemic Stroke Lesion Segmentation Challenge (ISLES) 2026. Zenodo (2026). https://doi.org/10.5281/zenodo.19856506
14. de la Rosa, E., Reyes, M., Liew, S.L., Hutton, A., Wiest, R., Kaesmacher, J., Hanning, U., Hakim, A., Zubal, R., Valenzuela, W., Robben, D., Sima, D.M., Anania, V., Brys, A., Meakin, J.A., Mickan, A., Broocks, G., Heitkamp, C., Gao, S., Liang, K., Zhang, Z., Rahman Siddiquee, M.M., Myronenko, A., Ashtari, P., Van Huffel, S., Jeong, H., Yoon, C., Kim, C., Huo, J., Ourselin, S., Sparks, R.,

Clèrigues, A., Oliver, A., Lladó, X., Chalcroft, L., Pappas, I., Bertels, J., Heylen, E., Moreau, J., Hatami, N., Frindel, C., Qayyum, A., Mazher, M., Puig, D., Lin, S.C., Juan, C.J., Hu, T., Boone, L., Goubran, M., Liu, Y.J., Wegener, S., Kofler, F., Ezhov, I., Shit, S., Hernandez Petzsche, M.R., Müller, M., Menze, B., Kirschke, J.S., Wiestler, B.: DeepISLES: a clinically validated ischemic stroke segmentation model from the ISLES'22 challenge. Nature Communications **16**(1), 7357 (2025). https://doi.org/10.1038/s41467-025-62373-x