



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

XTinyU-Net: Training-Free U-Net Scaling via Initialization-Time Sensitivity

Alvin Kimbowa^{1*†}, Moein Heidari^{1*}, David Liu^{1,2}, and Ilker Hacihaliloglu^{1,2,3}

¹ School of Biomedical Engineering, The University of British Columbia, Vancouver, Canada

² Department of Radiology, The University of British Columbia, Vancouver, Canada

³ Department of Medicine, The University of British Columbia, Vancouver, Canada
alvinbk@student.ubc.ca

{moein.heidari, david.liu, ilker.hacihaliloglu}@ubc.ca

Abstract. While U-Net architectures remain the gold standard for medical image segmentation, their deployment in resource-constrained environments demands aggressive model compression. However, finding an optimally efficient configuration is computationally prohibitive, typically requiring exhaustive train-and-evaluate cycles to find the smallest model that maintains peak performance. In this paper, we introduce a training-free selection framework to automatically identify ultralightweight, dataset-specific U-Net configurations directly at initialization. We observe that systematically scaling down U-Net channel width induces a sharp transition from a stable performance plateau to representational capacity collapse. To pinpoint this boundary without training, we propose a Jacobian-based sensitivity metric that scores discrete, width-capped U-Net variants using a small set of unlabeled images. By analyzing the total variation of this sensitivity curve, we isolate the smallest stable configuration, which we denote as XTinyU-Net. Evaluated across six diverse medical datasets within the nnU-Net framework, XTinyU-Net achieves segmentation accuracy comparable to the heavy nnU-Net baseline with $400\times$ – $1600\times$ fewer parameters, and outperforms contemporary lightweight architectures while utilizing $5\times$ – $72\times$ fewer parameters. Code is publicly accessible on GitHub.

Keywords: Light-weight · Efficient · Deep learning · Medical image segmentation

1 Introduction

Medical image segmentation is a foundational primitive for quantitative clinical analysis across diverse modalities [1,15]. While modern deep neural networks achieve remarkable fully-supervised performance, the deployment landscape is rapidly shifting toward decentralized, resource-constrained environments such

* Equal contribution

† Corresponding author

as point-of-care systems and portable devices. To bridge the gap between high accuracy and low computational cost, numerous lightweight architectures have been proposed. These specialized models reduce parameters and floating-point operations via compact blocks, simplified decoders, or token mixing, as exemplified by UNeXt [21], CMUNeXt [20], MedNCA [11], TinyUNet [4], and LeanUNet [8]. Despite this architectural proliferation, the canonical U-Net [18] remains an exceptionally competitive baseline, frequently outperforming specially crafted lightweight models when optimized with standardized pipelines like nnUNet [9].

These lightweight methods demonstrate that strong accuracy can be achieved at low cost, yet they surface an inconvenient reality: efficiency is not a static, one-size-fits-all architectural choice. For a fixed model family, the optimal budget depends heavily on dataset-specific characteristics, including image resolution, anatomical variability, noise, and label complexity. Consequently, practitioners still rely on computationally prohibitive manual sweeps to train and evaluate multiple candidate configurations. This expensive validation process undermines the core promise of lightweight models in settings where rapid iteration is essential.

This paper targets the configuration selection problem directly, asking: *How aggressively can a properly configured U-Net be scaled down before its representational capacity collapses and segmentation performance degrades for a given dataset?* We focus on width scaling because channel dimensions dominate computation in U-Net-style models, and it serves as the primary knob in many lightweight designs. To isolate the effect of model capacity without confounding factors, we instantiate our study within the standardized nnUNet protocol.

To address this selection problem without training, our approach relates to zero-cost neural architecture search (NAS) [14,5,13]. Existing training-free methods predict downstream accuracy using initialization-time statistics, with representative directions including activation-based criteria (e.g., NASWOT [14], SWAP [17]) and gradient-based criteria (e.g., SynFlow [19], SNIP [12]). However, these influential approaches are typically evaluated on classification tasks and macro-level inter-architecture comparisons. In contrast, our intra-family, dense-prediction setting requires reliably evaluating small architectural changes where capacity regime transitions are highly pronounced. Rather than treating the score as a global ranker, we utilize the Jacobian signal as a targeted collapse detector to answer the practical deployment question: how small can the model be before performance deteriorates?

Specifically, our contributions are as follows: **❶** We systematically analyze aggressive U-Net capacity scaling and consistently observe a predictable two-regime behavior: a broad performance plateau followed by an abrupt representational collapse at ultra-small model sizes. **❷** Building on this observation, we introduce a training-free, Jacobian-based framework for selecting dataset-adaptive lightweight configurations directly at initialization. This robust metric instantly isolates the optimal pre-collapse transition configuration, which we designate as XTinyU-Net. **❸** Extensive evaluations across six diverse medical imaging

datasets establish that XTinyU-Net achieves segmentation accuracy comparable to the nnU-Net baseline while requiring up to 1600x fewer parameters, and it consistently outperforms specialized state-of-the-art lightweight architectures while utilizing 5x–72x fewer parameters.

2 Method

We frame the search for an optimal lightweight segmentation network as a problem of detecting capacity collapse within a progressively scaled architecture family. Rather than relying on expensive, iterative train-and-evaluate cycles, we propose a training-free selection mechanism that operates entirely at random initialization. Our approach proceeds in two main steps. First, we construct a discrete, ordered family of candidate architectures by systematically constraining the maximum channel width of a standard baseline. Second, we evaluate the input-output sensitivity of each untrained candidate using a Jacobian-based metric. By analyzing the total variation of this initialization-time sensitivity across the model family, we can pinpoint the critical capacity boundary where the network transitions from a stable performance plateau to an unstable, post-collapse regime. The configuration lying immediately on the stable side of this boundary is selected as our highly efficient target architecture, XTinyU-Net.

2.1 U-Net scaling

We systematically investigate U-Net capacity limits by scaling network width (channels) rather than depth (stages). Reducing depth alters the model’s inductive bias by shrinking the theoretical receptive field and disrupting multi-scale context aggregation. In contrast, width scaling provides a strictly controlled axis for compression: it isolates representational capacity while perfectly preserving the architecture’s foundational spatial topology and downsampling schedule.

Let $C_l^{\text{base}} = 2^l C_0$, where C_0 is the number of channels at the first stage and L is the total number of stages, and let $C_{\max}^{\text{base}} = C_{L-1}^{\text{base}}$ denote the maximum channel dimensionality in the baseline configuration. We define a discrete family of scaled models by imposing a geometric channel cap

$$C_{\max}(i) = \frac{C_{\max}^{\text{base}}}{2^i}, i \in \{0, 1, \dots, \log_2 C_{\max}^{\text{base}}\} \quad (1)$$

such that the channel width at stage l is then given by $C_l(i) = \min(2^l C_0, C_{\max}(i))$.

As i increases, the maximum allowable channel dimensionality decreases, yielding progressively smaller models. For large values of i , the doubling pattern across resolutions saturates and the network approaches a near-uniform channel dimensionality across stages as in [8]. Importantly, the spatial hierarchy, number of stages, and receptive-field growth remain unchanged, and all other architectural and optimization settings are kept fixed. We use $C_{\max}^{\text{base}} = 512$ as in nnU-Net [9] and thus end up with a family of 10 models. We denote the resulting ordered model family by $\{\mathcal{M}_i\}_{i=0}^{N-1}$, where increasing index i corresponds to progressively smaller channel caps and reduced model capacity.

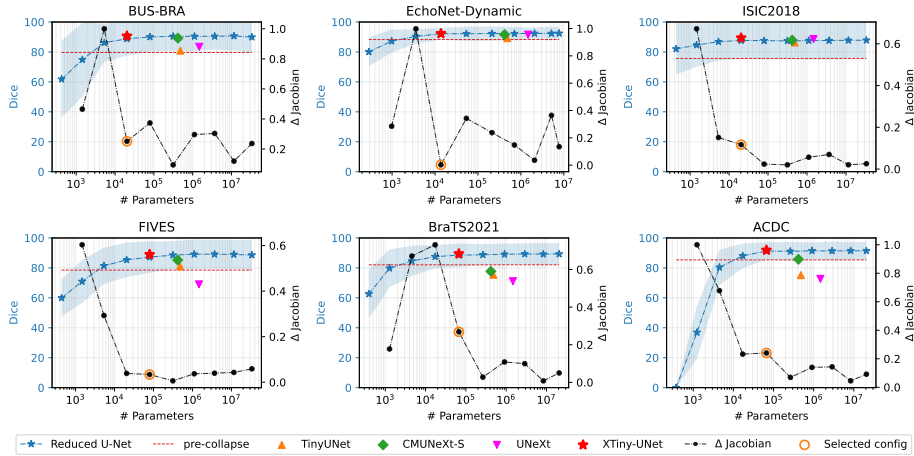


Fig. 1. U-Net width scaling curves (Dice vs parameters) and corresponding initialization-time Jacobian sensitivity across datasets. The estimated collapse boundary separates stable and unstable regimes; XTinyU-Net is selected as the smallest configuration on the stable side.

When training the defined U-Net family across datasets, we consistently observe two distinct performance regimes (measured by Dice, as seen in Fig. 1) as model size decreases. For moderate width reductions, segmentation performance remains relatively stable, forming a broad plateau. Beyond a dataset-dependent threshold, performance deteriorates sharply as models become severely underparameterized. We refer to these regimes as the pre-collapse and post-collapse regions, respectively. Different datasets exhibit different transition points, reflecting varying task complexity and representational requirements. This naturally raises the following question: *"Given a medical imaging dataset, can we estimate the smallest configuration that remains within the pre-collapse regime, at initialization without training the entire model family?"*

2.2 Jacobian-Based Collapse Detection

Our hypothesis is that the transition from the pre-collapse to the post-collapse regime is reflected in the local input–output sensitivity of the network at initialization. As model capacity decreases, sufficiently narrow configurations exhibit increased sensitivity to small perturbations in the input, indicating unstable representational behavior near the capacity boundary.

Input–output sensitivity. To quantify how sensitive a given model M_i (from the model family defined in 2.1) is to a given input $x \in \mathbb{R}^{C \times H \times W}$, where C, H, W are the number of channels, image height, and width, respectively, we take the input-output gradient, $g_i(x)$, of the model at random initialization as

follows; $g_i(x) = \nabla_x \sum_{n,h,w} M_i(x; \theta_i)_{n,h,w}$, where θ_i denotes the model parameters, $M_i(x; \theta_i) \in \mathbb{R}^{N \times H \times W}$, and N is the number of classes. We then define a scalar sensitivity measure $s_i = \|g_i(x)\|_2$. For a set of K images $\{x_k\}_{k=1}^K$, we aggregate the sensitivities by taking the average of the per-image sensitivities; $S_i = \frac{1}{K} \sum_{k=1}^K s_i$.

Cross-configuration normalization. Since the sensitivity magnitudes are arbitrary, with $S_i \in [0, \infty)$, we first normalize the values across the model configurations for meaningful comparison as follows; $\tilde{S}_i = \{S_i - S_{\min}\} / \{S_{\max} - S_{\min}\}$. Empirically, we observe that in the unstable post-collapse regime, the changes in the input-output sensitivity values between consecutive model configurations are larger compared to the changes in the pre-collapse region.

Collapse detection. Thus, we define the transition configuration between the pre- and post-collapse regimes to be the one that maximizes the average total variation (TV) in the Jacobians in the post-collapse regime while minimizing the average total variation (TV) in the pre-collapse regime. Let $\{M_i\}_{i=0}^{N-1}$ be the model configurations ordered from largest capacity ($i = 0$) to smallest capacity ($i = N - 1$). We define the local variation between consecutive model configurations as $d_j = |\tilde{S}_{j+1} - \tilde{S}_j|$, where $j \in \{0, \dots, N - 2\}$. For a candidate split index $k \in \{2, \dots, N - 2\}$, we define the total variation in the post-collapse (small models) and pre-collapse (large models) regions as

$$\text{TV}_{\text{post}}(k) = \frac{1}{N - k} \sum_{j=k-1}^{N-2} d_j, \quad \text{TV}_{\text{pre}}(k) = \frac{1}{k} \sum_{j=0}^{k-1} d_j, \quad (2)$$

respectively. We select the collapse boundary as

$$k^* = \arg \max_k (\text{TV}_{\text{post}}(k) - \text{TV}_{\text{pre}}(k)), \quad (3)$$

and choose M_{k^*-1} as the final pre-collapse configuration of choice. We dub this configuration **XTinyU-Net**. This way, we obtain a lightweight architecture that is as close to the transition point from pre- to post-collapse as possible (corresponding to red stars in Fig. 1).

3 Experiments

Datasets We evaluate our method on six public medical image segmentation datasets spanning diverse modalities. Specifically, we utilize BUS-BRA [7], EchoNet Dynamic [16], and ISIC 2018 [6]. We resize the ISIC data to 256×256 to reduce its resolution due to computational constraints. Furthermore, we include FiVES [10] for retinal fundus vessels, ACDC [3] for cardiac cine MRI structures, and BraTS [2] for brain MRI tumor sub-regions. We maintain the default splits of the datasets to ensure fairness across evaluations.

Table 1. Results on BUS-BRA, EchoNet Dynamic, and ISIC 2018. We report Params, FLOPs, Dice, and HD95 (mean $\pm$ std over three runs). Best is bold and second best is underlined. nnU-Net is shown in gray as the upper bound.

Config.	Params (k) $\downarrow$	FLOPS (G) $\downarrow$	BUS-BRA		EchoNet Dynamic		ISIC2018	
			Dice $\uparrow$	HD95 $\downarrow$	Dice $\uparrow$	HD95 $\downarrow$	Dice $\uparrow$	HD95 $\downarrow$
UNeXt [21]	1471.94	4.59	83.43 $\pm$ 13.78	20.19 $\pm$ 20.77	<u>91.52$\pm$3.91</u>	12.09 $\pm$ 13.99	88.58 $\pm$ 12.86	<u>20.64$\pm$23.33</u>
CMUNeXt-S [20]	<u>417.50</u>	<u>2.18</u>	<u>89.22$\pm$10.32</u>	12.47$\pm$15.89	90.81 $\pm$ 5.54	<u>11.65$\pm$12.08</u>	<u>88.67$\pm$13.53</u>	21.02 $\pm$ 26.20
TinyU-Net [4]	481.11	3.32	80.88 $\pm$ 17.91	27.97 $\pm$ 27.30	89.57 $\pm$ 5.74	18.35 $\pm$ 18.30	86.37 $\pm$ 15.48	24.25 $\pm$ 26.35
XTinyU-Net	20.39	0.92	90.64$\pm$7.57	<u>14.74$\pm$20.43</u>	92.31$\pm$3.94	3.78$\pm$2.52	89.39$\pm$12.38	18.19$\pm$23.56
nnU-Net [9]	33,472.40	62.40	89.90 $\pm$ 10.25	15.25 $\pm$ 21.05	92.45 $\pm$ 4.16	3.75 $\pm$ 3.90	87.84 $\pm$ 12.19	19.63 $\pm$ 23.32

Implementation details We instantiate our approach within the nnU-Net training and evaluation protocol in order to keep preprocessing, augmentation, patch size, and optimization fixed across all configurations. This ensures that performance differences primarily reflect capacity changes rather than variations in the training recipe. All models are trained using the default nnU-Net settings for each dataset. We report Dice and the 95th percentile Hausdorff distance (HD95), and present results as mean and standard deviation over three runs. We compare against recent efficient segmentation architectures targeting low-compute settings, including UNeXt, CMUNeXt-S, and TinyU-Net. We train the architectures using their open source code and following their respective publications. When also include nnU-Net as an upper-bound reference under the same protocol.

4 Results and Discussion

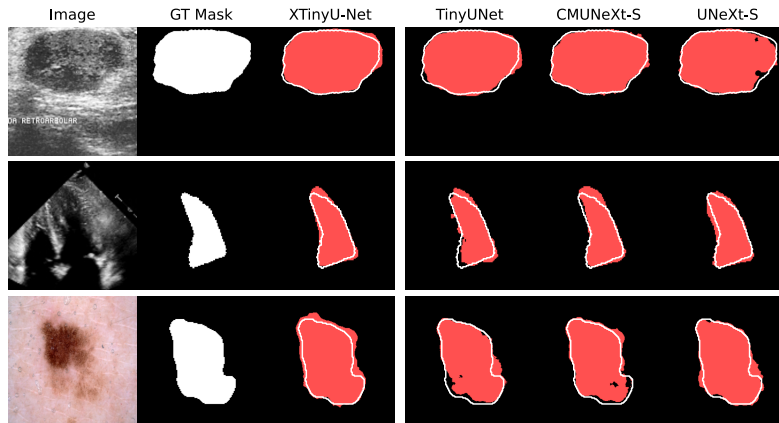
4.1 Quantitative Results

U-Net scaling. Across all six datasets, width scaling reveals a consistent two-regime pattern. As model capacity decreases, Dice remains relatively stable over a broad range of channel caps before degrading sharply at very small widths. This data-dependent pattern is visible in Fig. 1.

XTinyU-Net. XTinyU-Net consistently selects a configuration near the empirical collapse boundary while remaining on the stable side of the transition. As shown in Tables 1 and 2, XTinyU-Net achieves performance comparable to nnU-Net across datasets despite using over 400x–1600x fewer parameters. For instance, on BUS-BRA, XTinyU-Net achieves 90.64% Dice compared to 89.90% for nnU-Net while reducing parameters from 33,472k to 20.39k (over 1600 $\times$ fewer parameters) and FLOPs from 62.40G to 0.92G ($\sim 68\times$ reduction). Similar trends are observed on EchoNet-Dynamic and ACDC, where Dice remains within a narrow margin of the baseline despite substantial computational savings.

Table 2. Results on FiVES, BraTS2020, and ACDC. We report Params, FLOPs, Dice, and HD95 (mean $\pm$ std). Best is bold and second best is underlined.

Config.	Params (k) $\downarrow$	FLOPS (G) $\downarrow$	FiVES		BraTS2021		ACDC	
			Dice $\uparrow$	HD95 $\downarrow$	Dice $\uparrow$	HD95 $\downarrow$	Dice $\uparrow$	HD95 $\downarrow$
UNeXt [21]	1471.67	4.59	68.98 $\pm$ 7.14	13.25 $\pm$ 10.71	71.08 $\pm$ 17.81	26.71 $\pm$ 16.26	72.00 $\pm$ 14.28	8.38 $\pm$ 6.52
CMUNeXt-S [20]	<u>417.60</u>	<u>2.19</u>	85.27 $\pm$ 9.80	5.99 $\pm$ 12.97	77.68 $\pm$ 37.43	12.33 $\pm$ 12.54	84.40 $\pm$ 20.58	4.06 $\pm$ 4.91
TinyU-Net [4]	481.31	3.34	84.72 $\pm$ 8.07	6.99 $\pm$ 12.63	75.43 $\pm$ 17.75	17.48 $\pm$ 20.02	71.28 $\pm$ 22.29	14.35 $\pm$ 17.19
XTinyU-Net	79.55	1.96	88.97$\pm$10.42	5.17$\pm$14.80	89.47 $\pm$ 11.30	4.15$\pm$5.27	91.86$\pm$6.06	2.53$\pm$8.98
nnU-Net [9]	33,472.40	33.36	88.57 $\pm$ 10.03	5.08 $\pm$ 13.13	91.19 $\pm$ 11.17	2.89 $\pm$ 3.33	91.37 $\pm$ 6.01	2.27 $\pm$ 6.65

**Fig. 2.** Qualitative results on BUS-BRA (top row), EchoNet-Dynamic (middle row), and ISIC 2018 (bottom row). XTinyU-Net achieves high quality segmentation performance.

Baseline comparison. Compared to recent efficient segmentation models (UNeXt, CMUNeXt-S, TinyU-Net), XTinyU-Net achieves competitive or superior Dice while using over 5x-72x fewer parameters. These results suggest that dataset-adaptive width selection can be as effective as architecture redesign for achieving favorable performance-efficiency trade-offs.

4.2 Qualitative Results

Representative segmentation outputs are shown in Figs. 2 and 3. Despite having significantly more parameters, baseline light-weight architectures exhibit fragmented masks, missing regions, and unstable boundaries. In contrast, XTinyU-Net generates cleaner and more complete segmentations that closely resemble ground-truth masks.

4.3 Ablation Study

As illustrated in Fig. 4, we evaluate the robustness of our Jacobian-based collapse detection and compare it against existing zero-cost NAS metrics.

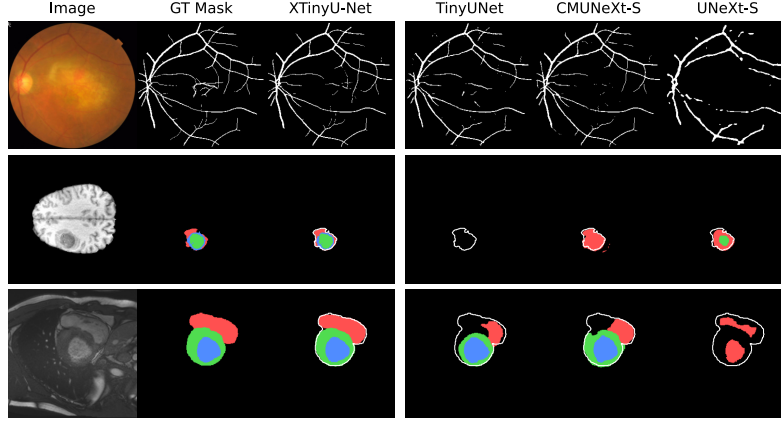


Fig. 3. Qualitative results on FiVES (top row), BraTS2021 (middle row), and ACDC (bottom row). XTinyU-Net achieves high quality segmentation performance.

Number of samples & Initialization: Our method exhibits robustness to variations in the number of samples used (K), consistently identifying the same model configuration even with a single sample image. The method also demonstrates resilience to random initialization seeds, with the selected architectures tightly clustered around two adjacent configurations. Crucially, these selected models achieve competitive segmentation performance while requiring 5x to 72x fewer parameters than existing lightweight networks.

NAS Metric Comparison: We observe that established NAS metrics (e.g., NASWOT, SWAP) scale monotonically with model capacity. This inherently continuous increase renders existing metrics incapable of detecting the sharp pre- to post-collapse transition boundary in U-Net architectures. Conversely, our sensitivity measure exhibits a distinct peak that successfully isolates this critical boundary.

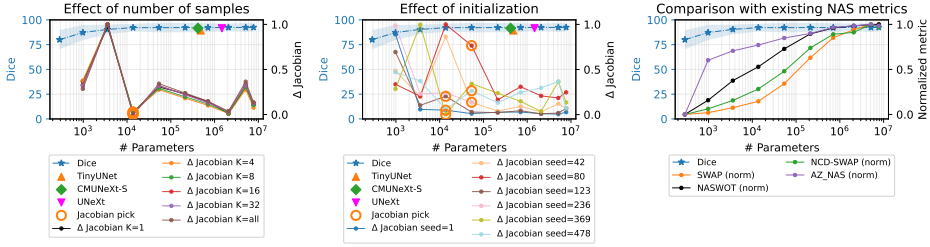


Fig. 4. Ablation study on EchoNet-Dynamic showing robustness of the proposed Jacobian-based collapse detection method to number of samples used (K), random initialization, and comparison with existing NAS metrics.

5 Conclusion

We propose a training-free, Jacobian-based collapse detection method to identify optimal, dataset-specific U-Net configurations directly at initialization. Evaluated across six medical datasets, our resulting model, XTinyU-Net, matches the segmentation performance of nnU-Net while requiring up to 1600x fewer parameters. Furthermore, it outperforms specialized lightweight architectures using 5x to 72x fewer parameters. Ultimately, we demonstrate that a properly width-scaled standard U-Net provides a superior performance-efficiency trade-off compared to custom lightweight designs.

Acknowledgments. This work was supported by the Canadian Foundation for Innovation–John R. Evans Leaders Fund (CFI-JELF) program [Grant ID 42816, AWD-023869 CFI], the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2023-03575, AWD-024385] / Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) [RGPIN-2023-03575, AWD-024385], the Canadian Consortium of Clinical Trial Training (CANTRAIN) platform, Michael Smith Health Research British Columbia, and the Minimally Invasive Image Guided Procedure Lab (MIIPs) at the University of British Columbia School of Biomedical Engineering.

Disclosure of Interests. The authors have no competing interests in the paper.

References

1. Azad, R., Heidokoohi, A., Baseri, S., et al.: Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024)
2. Baid, U., Ghodasara, S., Mohan, S., Bilello, et al.: The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification (Jul 2021). <https://doi.org/10.48550/arXiv.2107.02314>, <https://ui.adsabs.harvard.edu/abs/2021arXiv210702314B>, aDS Bibcode: 2021arXiv210702314B
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging* **37**(11), 2514–2525 (2018)
4. Chen, J., Chen, R., Wang, W., Cheng, J., Zhang, L., Chen, L.: TinyU-Net: Lighter yet Better U-Net with Cascaded Multi-Receptive Fields. In: *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. vol. LNCS 15009. Springer Nature Switzerland (October 2024). https://doi.org/10.1007/978-3-031-72114-4_60
5. Chen, W., Gong, X., Wang, Z.: Neural architecture search on imagenet in four gpu hours: A theoretically inspired perspective. In: *International Conference on Learning Representations* (2021), https://openreview.net/forum?id=6z_BEpN6Y0

6. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)
7. Gómez-Flores, W., Gregorio-Calas, M.J., Coelho de Albuquerque Pereira, W.: Busbra: A breast ultrasound dataset for assessing computer-aided diagnosis systems. *Medical physics* **51**(4), 3110–3123 (2024)
8. Hassler, T., Åkerholm, I., Nordström, M., Balletti, G., Goksel, O.: Lean unet: A compact model for image segmentation (2025). <https://doi.org/10.48550/arXiv.2512.03834>, <https://arxiv.org/abs/2512.03834>
9. 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). <https://doi.org/10.1038/s41592-020-01008-z>
10. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: Fives: A fundus image dataset for artificial intelligence based vessel segmentation. *Scientific data* **9**(1), 475 (2022)
11. Kalkhof, J., Ihm, N., Köhler, T., Gregori, B., Mukhopadhyay, A.: Med-nca: Bio-inspired medical image segmentation. *Medical Image Analysis* **103**, 103601 (2025). <https://doi.org/10.1016/j.media.2025.103601>
12. Lee, N., Ajanthan, T., Torr, P.H.S.: Snip: Single-shot network pruning based on connection sensitivity. In: *International Conference on Learning Representations* (2019). <https://doi.org/10.48550/arXiv.1810.02340>, <https://arxiv.org/abs/1810.02340>
13. Lin, M., Wang, P., Sun, Z., Chen, H., Sun, X., Qian, Q., Li, H., Jin, R.: Zen-nas: A zero-shot nas for high-performance image recognition. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 347–356 (October 2021)
14. Mellor, J., Turner, J., Storkey, A., Crowley, E.J.: Neural architecture search without training. In: *Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 139, pp. 7588–7598. PMLR (2021), <https://proceedings.mlr.press/v139/mellor21a.html>
15. Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P.: Foundation models for generalist medical artificial intelligence. *Nature* **616**(7956), 259–265 (2023)
16. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. *Nature* **580**(7802), 252–256 (2020)
17. Peng, Y., Song, A., Fayek, H.M., Ciesielski, V., Chang, X.: Swap-nas: Sample-wise activation patterns for ultra-fast nas. In: *International Conference on Learning Representations* (2024), <https://openreview.net/forum?id=tveiUXU2aa>
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
19. Tanaka, H., Kunin, D., Yamins, D.L.K., Ganguli, S.: Pruning neural networks without any data by iteratively conserving synaptic flow. In: *Advances in Neural Information Processing Systems* (2020). <https://doi.org/10.48550/arXiv.2006.05467>, <https://arxiv.org/abs/2006.05467>

20. Tang, F., Ding, J., Quan, Q., Wang, L., Ning, C., Zhou, S.K.: Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2024). <https://doi.org/10.1109/ISBI56570.2024.10635609>
21. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 23–33. Springer Nature Switzerland (2022). https://doi.org/10.1007/978-3-031-16443-9_3