
FMG-Net and W-Net: Multigrid Inspired Deep Learning Architectures For Medical Imaging Segmentation

Adrian Celaya
Rice University
Houston, TX 77005

Beatrice Riviere
Rice University
Houston, TX 77005

David Fuentes
The University of Texas MD Anderson Cancer Center
Houston, TX 77003

Abstract

Accurate medical imaging segmentation is critical for precise and effective medical interventions. However, despite the success of convolutional neural networks (CNNs) in medical image segmentation, they still face challenges in handling fine-scale features and variations in image scales. These challenges are particularly evident in complex and challenging segmentation tasks, such as the BraTS multi-label brain tumor segmentation challenge. In this task, accurately segmenting the various tumor sub-components, which vary significantly in size and shape, remains a significant challenge, with even state-of-the-art methods producing substantial errors. Therefore, we propose two architectures, FMG-Net and W-Net, that incorporate the principles of geometric multigrid methods for solving linear systems of equations into CNNs to address these challenges. Our experiments on the BraTS 2020 dataset demonstrate that both FMG-Net and W-Net outperform the widely used U-Net architecture regarding tumor subcomponent segmentation accuracy and training efficiency. These findings highlight the potential of incorporating the principles of multigrid methods into CNNs to improve the accuracy and efficiency of medical imaging segmentation.

1 Introduction

Medical imaging segmentation is an essential task in medical imaging analysis that involves dividing an image into regions of interest based on anatomical or pathological features. It is crucial in various clinical applications, including disease diagnosis, treatment planning, and surgical guidance. Accurate segmentation is critical for precise and effective medical interventions and can significantly impact patient outcomes [1, 2].

Deep learning methods, specifically convolutional neural networks (CNNs), have emerged as powerful tools for medical imaging segmentation [3–6]. CNNs have remarkably succeeded in accurately segmenting anatomical structures and pathological regions from medical images. However, despite these advancements, CNNs still face challenges in preserving fine structures and handling significant variations in image scales [7–9]. These challenges are particularly evident in complex and challenging segmentation tasks, such as the BraTS multi-label brain tumor segmentation challenge. In this task, accurately segmenting the various tumor sub-components, which vary significantly in size and shape, remains a significant challenge, with even state-of-the-art methods producing substantial errors [10–12].

Geometric multigrid methods (GMMs) are well-established techniques used in scientific computing to efficiently solve problems with multiple scales [13–15]. These methods use a hierarchy of grids with decreasing resolutions to efficiently capture large and fine-scale features. By capitalizing on these principles, we proposed two novel deep learning architectures, FMG-Net and W-Net, which incorporate the well-established ideas of the full and W-cycle multigrid methods into CNNs for medical imaging segmentation. These architectures aim to address the limitations of CNNs in handling fine-scale features and variations in image scales by incorporating the principles of these GMMs into the network architecture. Additionally, we aim to take advantage of the convergence properties of the full and W-cycle networks to reduce the number of iterations (i.e., epochs) required to achieve satisfactory results. The proposed architectures show promising results in accurately segmenting brain tumors while converging to lower loss function values in fewer epochs than the popular U-Net architecture, demonstrating their potential to improve medical imaging segmentation and ultimately enhance patient care.

1.1 Background and Previous Work

GMMs are a class of iterative methods that arise in scientific and engineering applications for solving linear systems of equations of the form $Au = f$, where the linear system A and the unknown variable u relate to a geometric grid involving multiple resolutions [16, 17]. Intrinsic to these methods is the construction of a series of grids of appropriate resolution. At each resolution, the constructed grid allows for the specific frequencies to be resolved, without aliasing. Like GMMs, many common CNN architectures for imaging tasks rely on manipulating images (or image features) at multiple scales because natural images encapsulate data at multiple resolutions. As a result, most CNN architectures – including many popular state-of-the-art methods such as nnUNet [5] – follow a pattern of downsampling and upsampling, following the intuition of the original U-Net paper [3] that popularized this approach. This common idea of hierarchical grids seen in GMMs and CNNs has guided the design of recent CNN architectures, resulting in smaller networks (in terms of the number of parameters) that achieve similar performance to current methods. A brief survey of these methods is provided below.

In their 2019 paper [18], He and Xu initially explored the similarities between multigrid methods and CNNs. Their approach was to leverage the principles of multigrid methods to create a unified framework, which they called MgNet. This framework simultaneously recovered some CNNs for image classification and GMMs, leading to a better understanding of the functions of various convolution operations and pooling used in CNNs. This analysis resulted in the development of modified CNNs with fewer weights and hyperparameters, which exhibited competitive and sometimes better performance when applied to CIFAR-10 and CIFAR-100 datasets. However, the applications of this approach were limited to image classification models, and no extension was made to segmentation.

Celaya et al. utilize the nearly identical structures of the V-cycle multigrid method and the U-Net segmentation architecture to propose the PocketNet paradigm for deep learning models, a straightforward modification to existing architectures that substantially reduces the number of parameters and maintains the same performance as the original architecture [19]. This modification questions the long-held assumption that doubling the number of features after each downsampling operation (i.e., pooling or convolution) is necessary for convolutional neural networks [20–22, 4, 3]. However, their results apply the concepts of GMMs to modify existing architectures similar to the V-cycle multigrid method (i.e., U-Net and nnUNet). Two other well-known GMMs are the full multigrid (FMG) and W-cycles, which construct a more complex hierarchy of grids than the V-cycle and are not utilized in their work.

1.2 Novel Contributions

In this paper, we make the following novel contributions:

1. We propose the FMG-Net and W-Net architectures for 3D medical imaging segmentation. These architectures are based on the well know FMG and W-cycle multigrid methods for solving linear systems of equations.
2. We show that our proposed architectures achieve superior segmentation performance vs. the popular U-Net architecture and that each architecture converges to lower loss values in fewer epochs.

2 Materials and Methods

2.1 FMG-Net and W-Net

The V-cycle, W-cycle, and FMG-cycle are different implementations of GMMs, with varying complexities and convergence properties. The V-cycle is the simplest and a commonly used multigrid method, where a sequence of smoothing and restriction operations are performed on the system at multiple grid levels, followed by a sequence of prolongation operations with corrections (i.e., skip connections) to interpolate the solution back to the finest grid. The W-cycle and FMG-cycle are more complex and computationally intensive, but they offer superior convergence properties by performing additional smoothing, interpolation, and correction operations at intermediate grid levels [16, 17]. Figure 1 illustrates the faster convergence of the FMG and W-cycles vs. the V-cycle for solving a linear system of equations resulting from the finite difference discretization of the Poisson problem $-\Delta u = f$ with zero Dirichlet boundary conditions.

Given that the U-Net architecture and the V-cycle are nearly identical, we propose the FMG-Net and W-Net architectures which mirror the structure of the FMG and W-cycles, respectively. We hypothesize that these architectures, with more complex interactions between features at intermediate grid levels than the U-Net, can achieve superior segmentation performance and faster convergence to lower loss function values. Figure 2 sketches the U-Net, FMG-Net, and W-Net architectures for various network depths. Note that we define the depth of each network to be the number of different resolution grids used by each architecture. The following rules govern the skip connections in each of these networks:

1. If we transfer features to a coarser grid (i.e., are on an encoder branch), then we pass all of the features to all subsequent upsampling operations at the current grid level.
2. If we encounter a “peak” (i.e., a transfer to a finer grid immediately followed by a transfer to a coarser grid), then we pass the features from the peak only to the next set of features at the same grid level.

Figure 3 illustrates these skip connection patterns for the U-Net, FMG-Net, and W-Net for the network depth set to three. In addition, each convolutional block (i.e. conv. block) consists of two convolutions, each followed by batch normalization and a ReLU non-linearity. Finally, we use the PocketNet paradigm proposed by Celaya et al. for the FMG-Net and W-Net, and *do not* double the number of features after downsampling to a coarser grid.

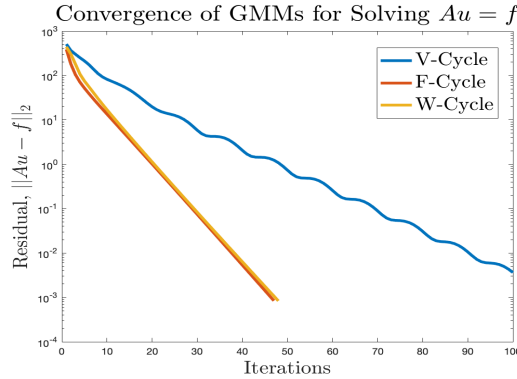


Figure 1: Convergence of V-cycle, FMG-cycle, and W-cycle GMMs. The linear system $Au = f$ results from the finite difference discretization of the Poisson problem. In this figure, we see that the FMG and W-cycles converge to lower residual values in fewer iterations than the V-cycle.

2.2 Data

We test our proposed network architectures on multi-label tumor segmentation in the publicly available (via CC BY 3.0 license) MICCAI Brain Tumor Segmentation (BraTS) Challenge 2020 dataset [23–25]. The BraTS training set contains 369 anonymized multimodal scans from 19

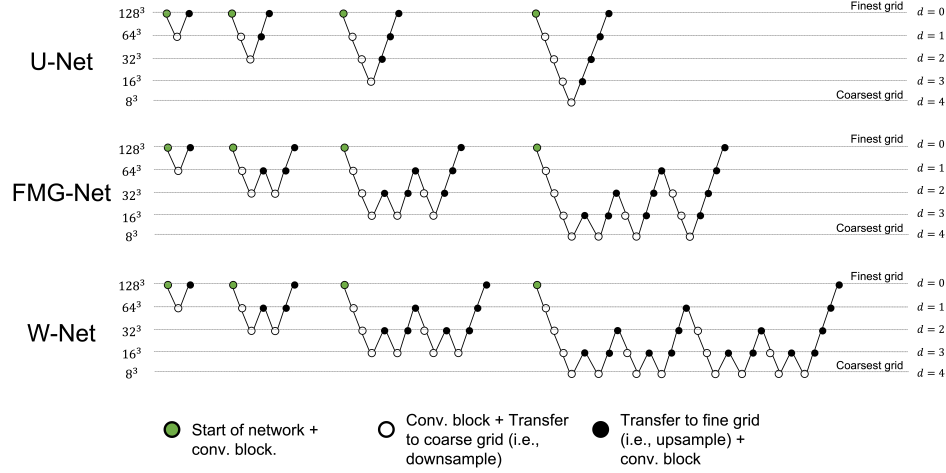


Figure 2: Sketches of U-Net, W-Net, and FMG-Net for different depths.

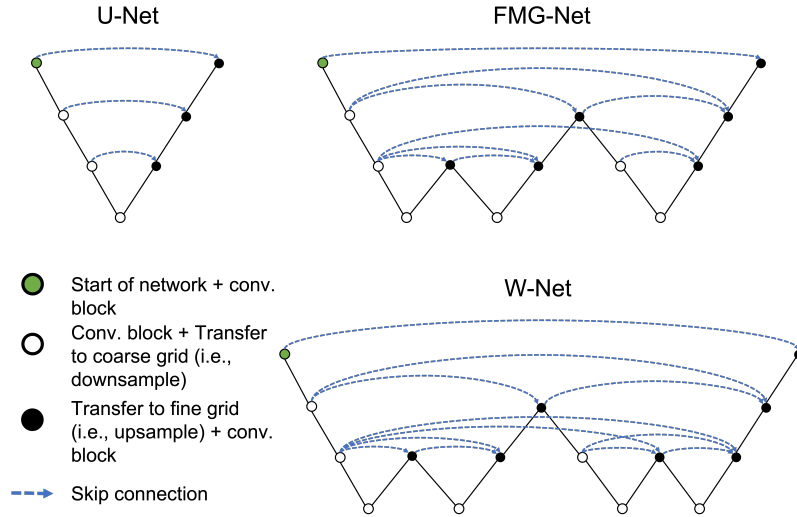


Figure 3: Sketches of U-Net, W-Net, and FMG-Net with skip connections.

institutions. Each set of scans includes a T1-weighted, post-contrast T1-weighted, T2-weighted, and T2 Fluid Attenuated Inversion Recovery volume along with a multi-label ground truth segmentation. The annotations include the GD-enhancing tumor (ET - label 4), the peritumoral edema (ED - label 2), and the necrotic and non-enhancing tumor core (NCR/NET - label 1). The final segmentation classes are the whole tumor (WT - labels 1, 2, and 4), tumor core (TC - labels 2 and 4), and ET. All volumes are provided at an isotropic voxel resolution of $1 \times 1 \times 1 \text{ mm}^3$, co-registered to one another, and skull stripped, with a size of $240 \times 240 \times 155$. We crop each image according to the brainmask (i.e., non-zero voxels) and apply z-score intensity normalization on only non-zero voxels for pre-processing. The BraTS training dataset is available for download at <https://www.med.upenn.edu/cbica/brats2020/registration.html>.

2.3 Metrics

We used three common medical imaging segmentation metrics to evaluate the performance of the proposed FMG-Net and W-Net architectures - the Dice coefficient, 95th percentile Hausdorff distance, and average surface distance. These metrics are implemented using the SimpleITK Python package [26–28].

Dice Coefficient - The Dice coefficient is a widely used metric for evaluating the similarity between the predicted segmentation and ground truth [29]. It measures the overlap between the two sets and ranges from 0 to 1, where a value of 1 indicates a perfect match between the predicted and ground truth segmentations.

95th Percentile Hausdorff Distance - While the Dice score is a commonly used metric for comparing two segmentation masks, it is not sensitive to local differences, as it represents a global measure of overlap. Therefore, we compute a complimentary metric, the 95th percentile Hausdorff distance (HD95), which is a distance metric that measures the maximum of the minimum distances between the predicted segmentation and the ground truth at the 95th percentile. The HD95 is a non-negative real number measured in millimeters, with a value of 0mm indicating a perfect prediction.

Average Surface Distance - The average surface distance (ASD) measures the average distance between the predicted segmentation and the ground truth along the surface of the object being segmented. It is a more fine-grained metric than the HD95 since it captures the surface-level details of the segmentation and can provide insight into the quality of the segmentation. The ASD is a non-negative real number measured in millimeters with a perfect prediction achieving a value of 0mm.

2.4 Training and Testing Protocols

We compare the segmentation performance for the popular U-Net architecture vs. the proposed FMG-Net and W-Net architectures described in Section 2.1. We select the Dice loss with cross-entropy for each architecture as a loss function [5]. We initialize the first layer in each network with 32 feature maps and use the Adam optimizer [30]. The learning rate is set to 0.0003. A five-fold cross-validation scheme is used to train and evaluate each loss function. That is, We train on 80% of the images for each fold and generate test-time predictions for the excluded 20%. We iterate this process until we have test-time predictions for all of the images in our dataset. During training, we select a batch size of two, use a patch size of $128 \times 128 \times 128$, and apply the same random augmentation described in [5]. To evaluate the validity of a predicted segmentation mask, we use the metrics described in Section 2.3. Our models are implemented in Python using TensorFlow (v2.11.0) and trained on an NVIDIA Quadro RTX 8000 GPU [31, 32]. All network weights are initialized using the default TensorFlow initializers. For reproducibility, we set all random seeds to 42. All other hyperparameters are left at their default values. The code for each network architecture is available at <https://github.com/aecelaya/mg-nets>.

3 Results

3.1 Accuracy

Using the methods described in Section 2, we train a 3D U-Net, FMG-Net, and W-Net and compare their segmentation performance. Table 1 shows the results for each architecture for depths 3, 4, and 5, respectively. In each case, we see that FMG-Net and W-Net generally outperform the U-Net architecture with respect to our selected metrics. Figure 4 illustrates several predictions from each of the three networks we test. Here, we see that, even for more difficult cases, the FMG-Net and W-Net can produce visually superior predictions than the U-Net.

3.2 Loss Curve Convergence

We also compare the loss curves (training and validation) for varying depths for the U-Net, FMG-Net, and W-Net architectures. Figure 5 shows the loss curves for each network for depths 3 through 5. Although each network converges to similar loss values, the FMG-Net and W-Net architectures generally do so in fewer epochs. This faster convergence (in terms of epochs) is especially true for the validation loss curves.

4 Discussion

The results presented above show that our proposed FMG and W-Net, which incorporate the design of two well-known corresponding GMM schemes, can achieve superior segmentation accuracy

Depth	Architecture	# Param.	Class	Dice	Hausdorff 95 (mm)	Avg. Surface (mm)
3	U-Net	5,608,036	WT	0.8833 (0.1363)	8.4089 (25.772)	2.9406 (19.634)
			TC	0.8240 (0.2156)	12.265 (41.892)	6.5104 (38.793)
			ET	0.7386 (0.2846)	37.747 (102.73)	32.701 (103.15)
	FMG-Net	1,108,356	WT	0.8977 (0.1195)	6.5106 (34.221)	4.1923 (33.483)
			TC	0.8422 (0.1940)	8.5238 (39.318)	5.6793 (38.910)
			ET	0.7647 (0.2650)	26.615 (87.858)	23.501 (88.249)
	W-Net	1,366,372	WT	0.8969 (0.1236)	7.6357 (43.081)	6.0217 (43.033)
			TC	0.8354 (0.1954)	9.0972 (43.199)	6.4143 (43.014)
			ET	0.7591 (0.2678)	28.205 (91.320)	25.201 (91.791)
4	U-Net	22,589,796	WT	0.8980 (0.1261)	4.6759 (27.798)	3.0767 (27.346)
			TC	0.8398 (0.2009)	10.046 (47.302)	7.4168 (47.058)
			ET	0.7549 (0.2731)	32.065 (98.065)	29.197 (98.572)
	FMG-Net	1,862,980	WT	0.9025 (0.1138)	4.3865 (21.538)	2.4595 (20.088)
			TC	0.8378 (0.2023)	8.3782 (39.088)	5.6000 (38.693)
			ET	0.7610 (0.2695)	27.331 (89.607)	24.363 (90.033)
	W-Net	3,235,684	WT	0.9001 (0.1094)	6.1075 (33.910)	4.0880 (33.441)
			TC	0.8472 (0.1840)	7.0097 (33.791)	4.3385 (33.429)
			ET	0.7661 (0.2632)	27.764 (91.349)	25.088 (91.814)
5	U-Net	90,500,964	WT	0.8968 (0.1129)	4.7741 (21.625)	2.2959 (19.467)
			TC	0.8363 (0.1933)	6.6805 (28.699)	3.5758 (27.392)
			ET	0.7510 (0.2729)	30.795 (94.914)	27.361 (95.237)
	FMG-Net	2,847,652	WT	0.8988 (0.0984)	5.2557 (22.106)	2.3330 (19.505)
			TC	0.8356 (0.1948)	6.0468 (21.959)	2.9306 (20.289)
			ET	0.7525 (0.2707)	28.966 (91.423)	25.663 (91.856)
	W-Net	7,886,692	WT	0.8993 (0.0963)	4.3079 (2.5746)	2.1407 (19.376)
			TC	0.8262 (0.1995)	7.9477 (34.216)	4.6273 (33.430)
			ET	0.7499 (0.2662)	28.893 (91.221)	25.371 (91.757)

Table 1: Mean and standard deviation for each metric for the U-Net, FMG-Net, and W-Net architectures with varying depths. For each depth, we highlight the best metric values in bold text. In each case, we see that FMG and W-Nets generally outperform the U-Net architecture regarding our selected metrics.

compared to the widely used U-Net architecture on the challenging BraTS dataset. Indeed, the proposed architectures seem to help address current CNNs’ limitations in handling multi-scale features. This ability to improve upon handling large and fine-scale features is evident in the FMG and W-Net’s superior segmentation accuracy for the various tumor subcomponents (which vary significantly in size and shape) in the BraTS dataset compared to the U-Net architecture. Figure 4 makes this clear visually, with the FMG and W-Net architectures resolving finer scale features in smaller, more difficult-to-segment tumors. However, one of the objectives of future work will be to add more comparisons to other architectures like nnUNet and HRNet (non-U-shaped) [33]. Additionally, incorporating an adaptive framework like the nnUNet (i.e., depth and network shape based on the dataset) into our architectures may improve our segmentation accuracy.

It is clear from Table 1 that for depth equal to three, the FMG-Net outperforms the U-Net and W-Net. It is also important to note that even in that case, the W-Net yields more accurate results than the U-Net. However, the picture is less clear regarding the FMG vs. the W-Net architecture for depths four and five. In these cases, the FMG and W-Net still generally outperform the U-Net, but it is unclear whether one is more accurate. Additionally, there is no trend in the accuracy for these networks vs. the depth. This behavior is not only consistent with multigrid theory, which does not guarantee faster convergence to a solution as a function of depth [16, 15], but also with the results in [34], which show that for the BraTS dataset, the depth of the U-Net architecture does not significantly affect accuracy. However, this may not be true for other datasets like the Liver and Tumor Segmentation or Medical Segmentation Decathlon datasets [35, 7]. Further testing on these is needed to validate our results further, determine which cases to use the FMG or W-Net architectures, and assess the effect of the depth parameter on accuracy.

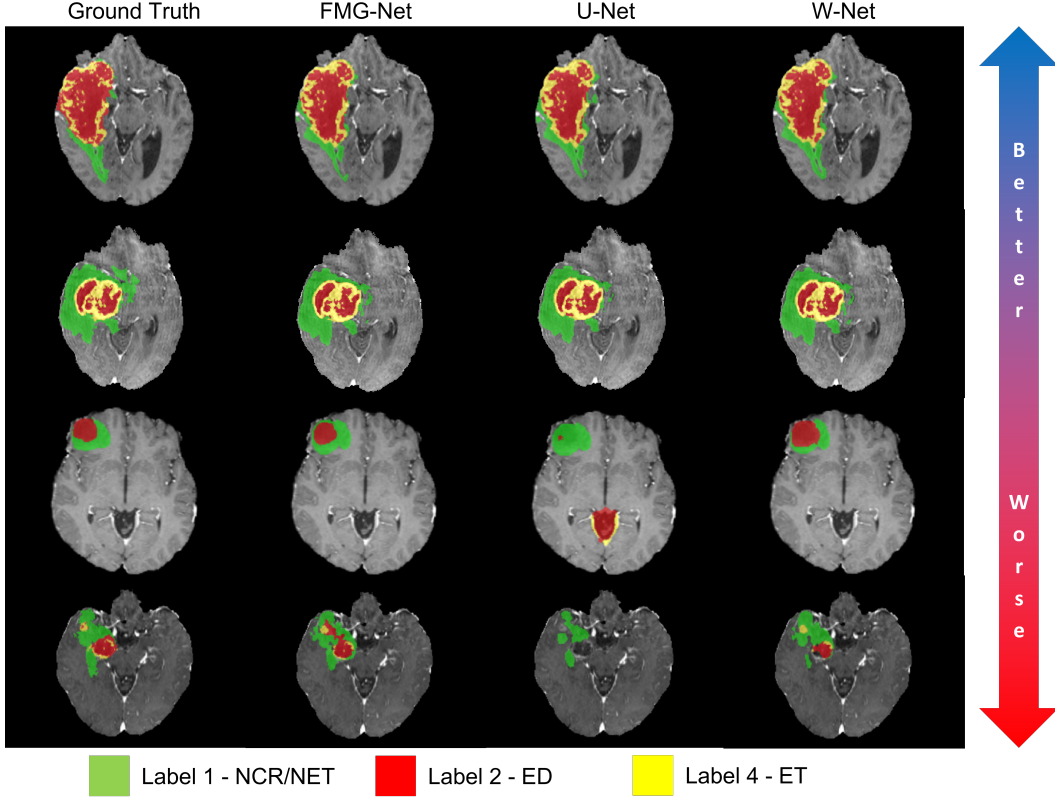


Figure 4: From left to right, ground truth and predictions from the FMG-Net, U-Net, and W-Net for several cases. Here, we see that, even for more difficult cases, the FMG-Net and W-Net can produce visually superior predictions than the U-Net.

Because the FMG and W-Net architectures use substantially fewer parameters than the U-Net, they offer advantages regarding speed-ups in the average time per training step and memory usage. The FMG-Net offers the highest training speed-ups vs. the U-Net, with reductions in the average time per training step ranging from 7% to 20%, depending on depth. The W-Net offers more modest speed-ups, with reductions in the average time per training step ranging from 2% to 6%, depending on depth. The W-Net architecture not only has more parameters than the FMG-Net but also requires more upsampling layers, downsampling layers, and skip connections (i.e., concatenations) than the FMG-Net, which adds to the computational cost of the architecture. We see similar results from the point of view of GPU memory usage, with the FMG-Net offering the most savings (roughly 20% vs. the U-Net) and the W-Net giving more modest savings (roughly 8% vs. the U-Net). While further work is needed to determine which architecture results in higher accuracy, it is clear that the FMG-Net produces the highest savings in terms of time and memory requirements. Practitioners considering the use of our proposed architectures should also consider this advantage in computational cost.

In addition to computational cost savings offered by our architectures, we also see that the loss curves for the FMG and W-Net achieve lower loss values in fewer epochs than the U-Net. Indeed, as the depth increases, the separation between the loss curves for the U-Net and our proposed architectures becomes more apparent. Especially in the validation curves, our architectures achieve lower loss values in fewer epochs than the U-Net. This convergence behavior is consistent with the convergence for the V-cycle vs. the FMG and W-cycles seen in Figure 1.

The FMG-Net and W-Net achieve higher accuracy than the U-Net architecture while using substantially fewer parameters. One possible explanation is that our proposed architectures rely on more complex interactions between features at multiple resolutions. These results suggest that the success of CNNs for medical imaging segmentation may have less to do with the number of parameters in each architecture but rather the richness of the interactions of features at different grid levels. Further

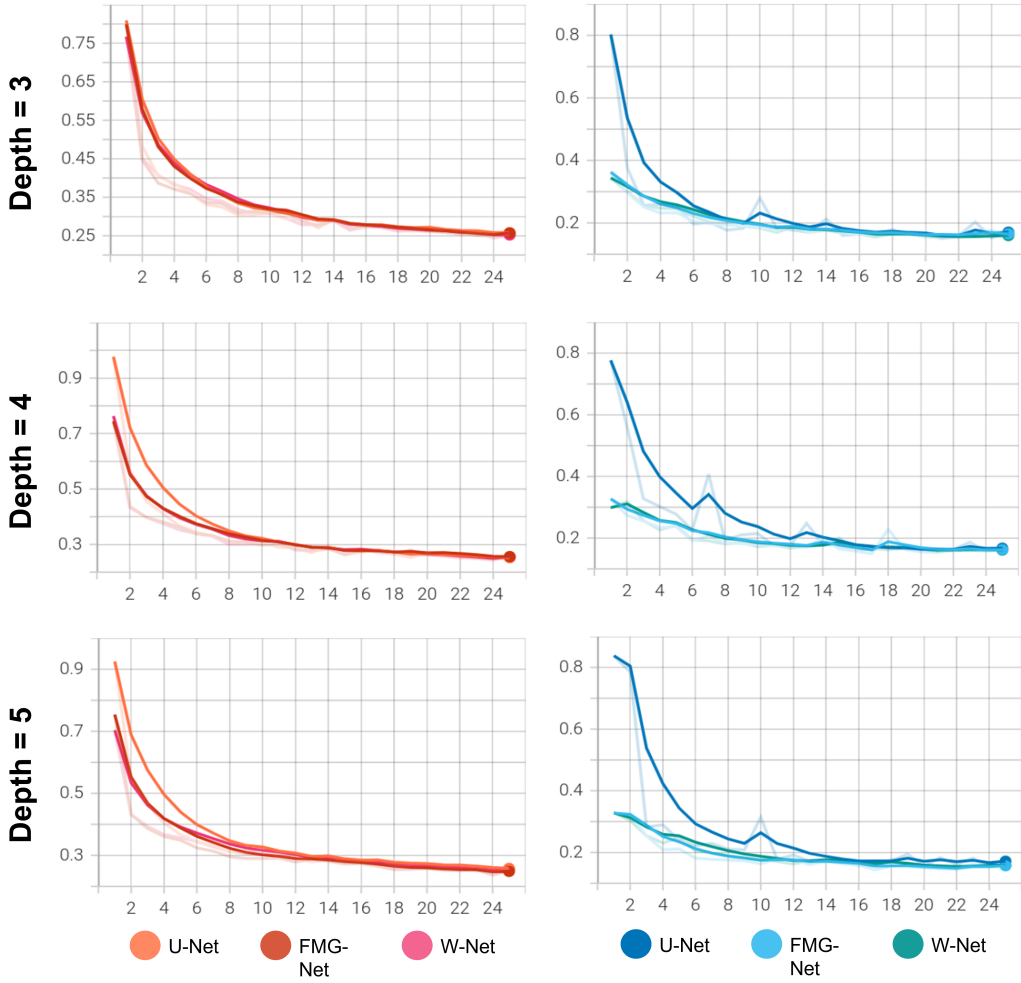


Figure 5: (Left) Training loss curves for depths 3 through 5 for the U-Net, FMG-Net, and W-Net architectures for the first 25 training epochs. (Right) Validation loss curves for depths 3 through 5 for the U-Net, FMG-Net, and W-Net architectures for the first 25 training epochs. As the depth increases, the separation between the loss curves for the U-Net and our proposed architectures becomes more apparent. Especially in the validation curves, our architectures achieve lower loss values in fewer epochs than the U-Net.

testing on different datasets is necessary to validate our findings’ generalizability and explore the potential of FMG-Net and W-Net in other medical imaging applications. Overall, our study highlights the promising potential of incorporating GMM principles into CNNs to improve the accuracy and efficiency of medical imaging segmentation.

Acknowledgments and Disclosure of Funding

The Department of Defense supports Adrian Celaya through the National Defense Science & Engineering Graduate Fellowship Program. David Fuentes is partially supported by R21CA249373. Beatrice Riviere is partially supported by NSF-DMS2111459. This research was partially supported by the Tumor Measurement Initiative through the MD Anderson Strategic Research Initiative Development (STRIDE), NSF-2111147, and NSF-2111459.

References

- [1] N. Sharma and et al., “Automated medical image segmentation techniques,” *Journal of Medical Physics*, vol. 35, pp. 3–14, 1 2010.
- [2] E. Ermiş and et al., “Fully automated brain resection cavity delineation for radiation target volume definition in glioblastoma patients using deep learning,” *Radiation Oncology*, vol. 15, pp. 1–10, 5 2020.
- [3] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [4] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, 2016.
- [5] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods* 2020 18:2, vol. 18, pp. 203–211, 12 2020.
- [6] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-net and its variants for medical image segmentation: A review of theory and applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
- [7] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, et al., “The medical segmentation decathlon,” *Nature communications*, vol. 13, no. 1, p. 4128, 2022.
- [8] T. Kitrungsrotsakul, X.-H. Han, Y. Iwamoto, A. H. Foruzan, L. Lin, and Y.-W. Chen, “Robust hepatic vessel segmentation using multi deep convolution network,” in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10137, pp. 269–274, SPIE, 2017.
- [9] L. A. Lim and H. Y. Keles, “Foreground segmentation using convolutional neural networks for multiscale feature encoding,” *Pattern Recognition Letters*, vol. 112, pp. 256–262, 2018.
- [10] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, “nnu-net for brain tumor segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (A. Crimi and S. Bakas, eds.), (Cham), pp. 118–132, Springer International Publishing, 2021.
- [11] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, “dresu-net: 3d deep residual u-net based brain tumor segmentation from multimodal mri,” *Biomedical Signal Processing and Control*, vol. 79, p. 103861, 2023.
- [12] Y. Cao, W. Zhou, M. Zang, D. An, Y. Feng, and B. Yu, “Mbanet: A 3d convolutional neural network with multi-branch attention for brain tumor segmentation from mri images,” *Biomedical Signal Processing and Control*, vol. 80, p. 104296, 2023.
- [13] Y. Zhu and A. C. Cangellaris, *Multigrid finite element methods for electromagnetic field modeling*, vol. 28. John Wiley & Sons, 2006.
- [14] B. Rivière, *Discontinuous Galerkin methods for solving elliptic and parabolic equations: theory and implementation*. SIAM, 2008.
- [15] T. M. Shah, “Analysis of the multigrid method,” *NASA Sti/Recon Technical Report N*, vol. 91, p. 23418, 1989.
- [16] J. H. Bramble, *Multigrid methods*. Routledge, 2018.
- [17] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.
- [18] J. He and J. Xu, “MgNet: A unified framework of multigrid and convolutional neural network,” *Science China Mathematics*, vol. 62, no. 7, pp. 1331–1354, 2019.
- [19] A. Celaya and et al., “Pocketnet: A smaller neural network for medical image analysis,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 4, pp. 1172–1184, 2023.
- [20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.

- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, p. 84–90, may 2017.
- [23] B. Menze and et al., “The multimodal brain tumor image segmentation benchmark (brats),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
- [24] S. Bakas and et al., “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge,” *arXiv preprint arXiv:1811.02629*, 2018.
- [25] S. Bakas and et al., “Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features,” *Scientific Data*, vol. 4, no. 1, p. 170117, 2017.
- [26] B. Lowekamp, D. Chen, L. Ibanez, and D. Blezek, “The design of simpleitk,” *Frontiers in Neuroinformatics*, vol. 7, p. 45, 2013.
- [27] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, “Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research,” *Journal of Digital Imaging*, vol. 31, pp. 290–303, 6 2018.
- [28] R. Beare, B. Lowekamp, and Z. Yaniv, “Image segmentation, registration and characterization in r with simpleitk,” *Journal of Statistical Software, Articles*, vol. 86, no. 8, pp. 1–35, 2018.
- [29] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [30] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [31] M. Abadi and et. al., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [32] F. Chollet and et al., “Keras,” 2015. Software available from keras.org.
- [33] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” in *CVPR*, 2019.
- [34] B. Jena, S. Jain, G. K. Nayak, and S. Saxena, “Analysis of depth variation of u-net architecture for brain tumor segmentation,” *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 10723–10743, 2023.
- [35] P. Bilic and et al., “The liver tumor segmentation benchmark (LiTS),” *arXiv preprint arXiv:1901.04056*, 2019.