Tumor Segmentation in Medical Images: Analysis of 3D Architecture Design Approaches

Anh Tuan Tran and Jared Young
Department of Computer Science, Kent State University, Kent, Ohio, USA
CS64201 Advanced Artificial Intelligence

***Abstract*—Brain tumor segmentation from multimodal magnetic resonance imaging (MRI) remains a central problem in medical image analysis because accurate delineation of tumor tissue supports diagnosis, treatment planning, and downstream clinical assessment. This paper organizes four selected studies into two architecture themes. The foundational section examines V-Net and nnU-Net as strong volumetric segmentation baselines that emphasize 3D encoder–decoder design, Dice-aware optimization, and task-aligned training pipelines. The hybrid architectures section examines TransBTS and DSNet, two models that strengthen volumetric segmentation through Transformer-based global reasoning, dynamic convolution, attention-guided feature reuse, and adversarial refinement. Taken together, the four papers show that progress in brain tumor segmentation does not come from one single design trick. It emerges from a balance between stable volumetric backbones, richer feature modeling, benchmark-aware optimization, and the practical trade-offs of complexity, memory cost, and clinical extensibility.**

***Index Terms*—brain tumor segmentation, MRI, volumetric deep learning, V-Net, nnU-Net, TransBTS, DSNet**

I. Introduction

Brain tumor segmentation is one of the most important applications of deep learning in medical imaging because accurate delineation of tumor subregions supports diagnosis, treatment planning, therapy monitoring, and broader clinical decision making. Multimodal MRI is especially useful in this setting because T1, T1ce, T2, and FLAIR each reveal different aspects of tumor structure. At the same time, the task remains difficult due to class imbalance, ambiguous tumor boundaries, irregular morphology, and the large variation that appears across patients and scanners.

Although earlier deep learning approaches often applied CNNs to 2D MRI slices or to a small number of fixed views, that strategy can lose important information from the full three-dimensional tumor structure. Slice-wise processing is computationally simpler, but it weakens continuity across adjacent slices and can make it harder to capture irregular tumor shape, ambiguous boundaries, and the broader spatial context that helps determine which regions belong to the lesion. This limitation is one of the main reasons later work shifted toward volumetric 3D segmentation models, where the network can learn tumor extent, shape, and surrounding context directly from the full MRI volume rather than inferring them from separated 2D views [2], [3].

Rather than treating the four selected papers as isolated summaries, this report organizes them into two conceptual groups. The first group covers *foundational volumetric segmentation architectures*. These papers establish the core design logic behind strong 3D segmentation systems, including fully convolutional encoder–decoder structures, Dice-aware optimization, and benchmark-aligned training pipelines. The second group covers *hybrid architectures*, meaning models that begin from strong volumetric segmentation backbones and then strengthen them with more advanced modeling components such as Transformer bottlenecks, dynamic convolution, attention-based skip reuse, and adversarial refinement. Under this framing, V-Net and nnU-Net define the foundational section, while TransBTS and DSNet define the hybrid section [1]–[4].

This structure helps the report focus on how deep learning is being applied to brain tumor segmentation at the architectural level. The key comparison is therefore not simply whether one model scores higher than another, but how each paper handles volumetric context, feature transmission, global reasoning, and the computational trade-offs required to improve segmentation quality.

II. Foundational Volumetric Segmentation Architectures

This section introduces the two papers that define the foundational volumetric side of the report. Both V-Net and nnU-Net are centered on strong 3D encoder–decoder segmentation logic, where the full MRI volume is processed as a volumetric input rather than as isolated slices. What makes them foundational is that they establish the core design principles that later models build on: fully convolutional 3D feature extraction, skip-connected encoder–decoder reconstruction, and training objectives aligned with highly imbalanced medical segmentation tasks [1], [3]. Although the two papers differ in emphasis, they both treat volumetric segmentation as the baseline architectural answer to the limitations of slice-wise reasoning.

Taken together, these two papers show that strong performance in brain tumor segmentation does not always depend on adding more exotic modeling components. V-Net represents a more explicit architectural foundation built around residual volumetric processing and Dice-aware optimization, while nnU-Net shows how a comparatively plain 3D U-Net-like backbone can become highly competitive when the surrounding pipeline is carefully aligned to the benchmark and target regions [1], [3].

A. V-Net Architecture

Paper [3] proposed V-Net which is a volumetric, fully convolutional neural network (FCN) designed specifically for 3D medical image segmentation (e.g., prostate MRI volumes). It is inspired by the 2D U-Net but fully adapted to 3D data: it processes entire MRI volumes end-to-end (no slice-by-slice processing), uses only 3D convolutions, and outputs a dense segmentation mask of the same spatial size as the input.

The V-Net architecture follows a classic encoder-decoder (U-shaped) structure. The left side (encoder/compression path) is responsible for extracting hierarchical features while progressively downsampling the spatial resolution and increasing the number of feature channels. Conversely, the right side (decoder/decompression path) upsamples the low-resolution features back to the original input size while fusing high-resolution details. Skip connections link the corresponding stages on both sides, thereby preserving fine-grained spatial information that would otherwise be lost during downsampling. During the workflow, traditional max-pooling layers are completely replaced by convolutional layers using $2 \times 2 \times 2$ filters with stride = 2 for downsampling. The paper adopts this design because this operation becomes fully learnable rather than fixed.

The encoder(left path) of V-Net is divided into five stages that operate at progressively lower resolutions, with the spatial dimensions being halved at each stage according to the input size sequence $128 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 8$. Within each stage (L-Stage 1 to L-Stage 5), the network applies one to three convolutional layers using $5 \times 5 \times 5$ volumetric kernels with stride 1 and appropriate padding, ensuring that the spatial size remains unchanged inside the stage; each convolution is followed by a PReLU activation. To enable residual learning, the input tensor to the entire stage is added element-wise to the output of the last convolutional layer of that stage, allowing the stage to learn a residual function rather than a direct mapping. Between stages, downsampling is performed by a $2 \times 2 \times 2$ convolution with stride 2, which simultaneously halves the spatial resolution in all three dimensions and doubles the number of feature channels (progressing from 16 channels after the first stage to 256 channels in the deepest bottleneck stage L-Stage 5). The paper adopts this strided-convolution approach instead of traditional max-pooling because it is fully learnable and eliminates the need to store pooling switches for back-propagation, resulting in a smaller memory footprint during training. By the end of L-Stage 5, the theoretical receptive field reaches $372 \times 372 \times 372$ voxels, enabling the deepest features to capture the entire input volume and impose global shape constraints.

On the other hand, the decoder (right path) of V-Net consists of four upsampling stages (R-Stage 4 to R-Stage 1) that progressively restore the spatial resolution back to the original input size following the sequence $8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128$. Within each stage, the network applies one to three convolutional layers using $5 \times 5 \times 5$ volumetric kernels with stride 1 and appropriate padding, ensuring that the spatial size remains unchanged inside the stage; the number of feature channels is halved compared to the previous (deeper) stage, and each convolution is followed by a PReLU activation. Residual learning is applied exactly as on the encoder side, whereby the input tensor to the entire stage is added element-wise to the output of the last convolutional layer of that stage. Between stages, upsampling is performed by a $2 \times 2 \times 2$ de-convolution (transposed convolution) with stride 2, which doubles the spatial resolution in all three dimensions by projecting each input voxel onto a larger region; after each de-convolution, the subsequent convolutional layers refine the upsampled features. Additionally, the skip Connections for fine-grained features from the corresponding left-side stages at matching resolutions are forwarded to the right-side stages through skip connections, enabling the decoder to recover precise spatial details such as edges and textures that would otherwise be lost during aggressive downsampling, while also accelerating the overall convergence speed of the network.

Finally, following the last decoder stage (R-Stage 1) that restores the spatial resolution to the original input size of $128 \times 128 \times 64$, a $1 \times 1 \times 1$ convolution is applied to produce exactly two feature maps, one corresponding to the foreground (prostate) class and the other to the background class. These two feature maps maintain identical spatial dimensions to the input volume. A voxel-wise softmax operation is then performed across the two channels to convert the feature maps into probability volumes, representing the probability of each voxel belonging to the prostate or the background with threshold = 0.5(voxel > 0.5 is prostate).

B. nnU-Net Architecture

Isensee et al. position nnU-Net as both a strong baseline and a configurable framework for BraTS-specific optimization rather than as a radically new backbone [1]. Architecturally, the model remains a plain 3D U-Net-like encoder–decoder with skip connections, five downsampling operations, deep supervision, and a cubic input patch of $128 \times 128 \times 128$ voxels. The network uses strided convolutions for downsampling, transposed convolutions for upsampling, leaky ReLU nonlinearities, and instance normalization, while feature map counts mirror the encoder and decoder across resolutions [1]. This design choice is important because nnU-Net demonstrates that state-of-the-art brain tumor segmentation can still emerge from a comparatively simple volumetric backbone when the surrounding training pipeline is aligned tightly to the task.

A central architectural idea in nnU-Net is that the backbone should remain standardized while the pipeline adapts to the benchmark. The paper replaces ordinary class-wise softmax training with region-based training on whole tumor (WT), tumor core (TC), and enhancing tumor (ET) using sigmoid outputs and binary cross-entropy, directly matching the BraTS evaluation targets [1]. The authors further strengthen the model through more aggressive data augmentation, modified batch settings, and a postprocessing rule that removes very small enhancing-tumor predictions when this improves BraTS ranking behavior. This means the core architecture is intentionally

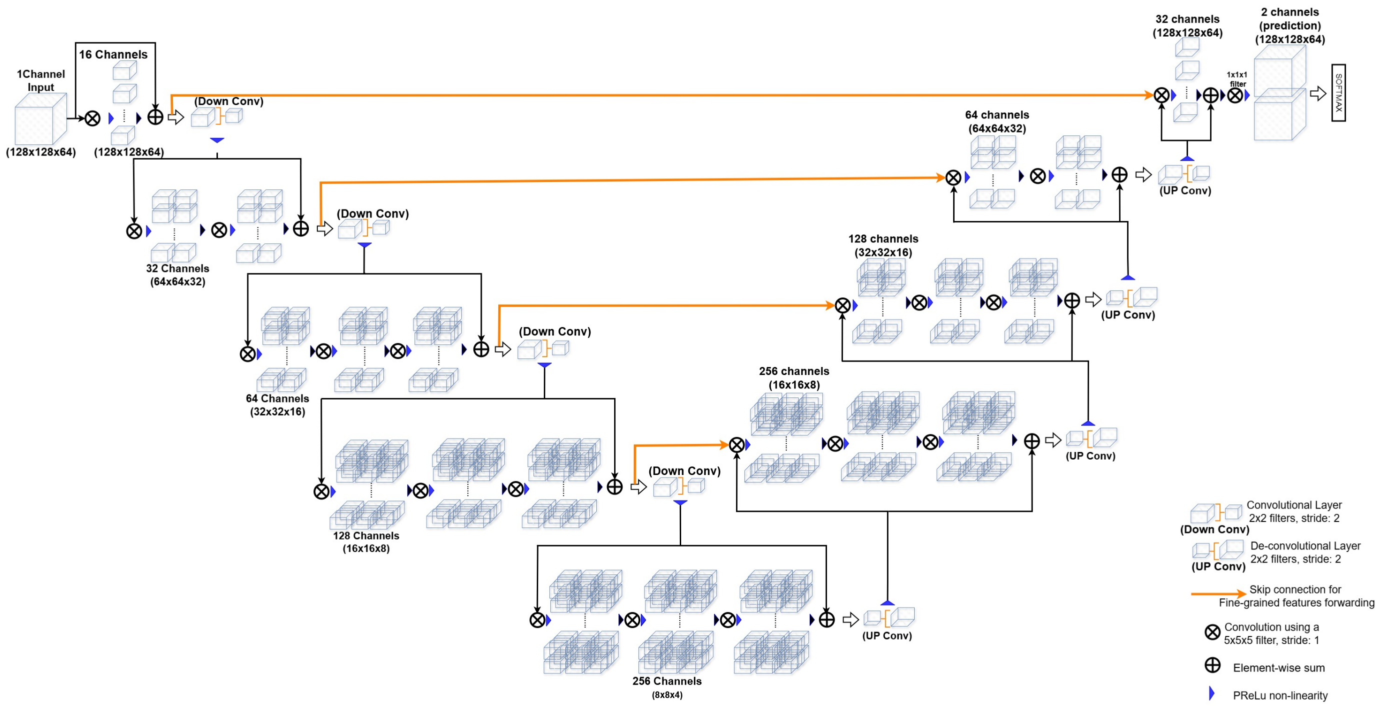

Fig. 1. V-Net Architecture

plain, but its training objective, target representation, and evaluation alignment are highly specialized.

Technically, nnU-Net shows that volumetric segmentation performance depends not only on architectural novelty but also on how training and model selection are formulated. On the BraTS 2020 validation set, its strongest configuration improved the mean Dice score from 84.18 to 85.58 and reduced the mean HD95 from 15.30 to 11.93. On the BraTS 2020 test set, the final submission reported Dice scores of 88.95 for WT, 85.06 for TC, and 82.03 for ET, which the paper identifies as the first-place result in the challenge [1]. For this reason, nnU-Net is best understood here as a foundational volumetric architecture: the backbone is not hybridized with new representational machinery, but the system still reaches state-of-the-art performance by maximizing the effectiveness of a disciplined 3D segmentation pipeline.

III. Hybrid Architectures

This section turns to the two papers that extend volumetric segmentation through more specialized architectural mechanisms. Unlike the foundational models, which rely mainly on strong 3D encoder–decoder design and training discipline, TransBTS and DSNet both attempt to strengthen tumor modeling by adding components that explicitly enhance representation quality. In TransBTS, that refinement comes from a Transformer bottleneck that improves global dependency modeling across the 3D feature space, while in DSNet it comes from dynamic convolution, attention-guided skip reuse, and adversarial refinement for sharper and more adaptive segmentation behavior [2], [4].

Grouped together, these two papers represent the hybrid side of the report because they start from the logic of volumetric segmentation and then push it further with mechanisms intended to capture harder tumor structure, long-range context, or more heterogeneous boundary detail. Their shared idea is that a standard 3D CNN backbone is powerful, but not always sufficient by itself when segmentation quality depends on stronger global reasoning or more adaptive feature selection [2], [4]. This makes them a natural contrast to the foundational section above.

A. *DSNet Architecture*

Paper [4] proposed a volumetric architecture named DSNet, which is inspired by the classic U-Net encoder–decoder structure and integrates an advanced attention mechanism into its skip connections. The attention block in DSNet is implemented as an attention-based skip connection that links each corresponding level of the encoder and decoder. To make the attention more adaptive, the authors use a Dynamic Convolutional Neural Network (DCNN) module. As shown in Figure 2, it first applies 3D adaptive average pooling to compress spatial information, followed by two 3D convolutional layers: the first with ReLU activation and the second with softmax activation. The softmax output produces normalized attention weights that indicate the importance of each kernel for the current input features. These adaptive weights are then used to compute a weighted sum of the multiple kernels, resulting in a dynamic convolution operation. The resulting attention map is multiplied element-wise with the encoder features before they are forwarded to the decoder. This DCNN-powered attention

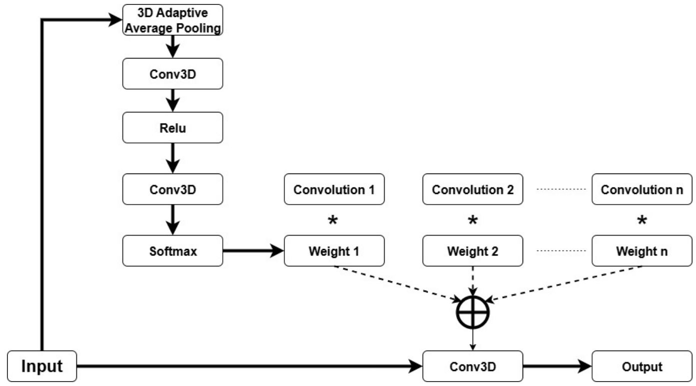

Fig. 2. Dynamic convolution neural network (DCNN) structure.

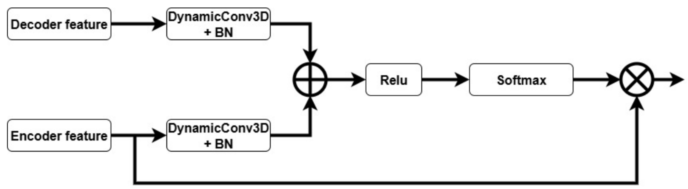

Fig. 3. Attention mechanism structure implemented in DSNet's skip Connections.

design enables the decoder to selectively emphasize the most relevant fine-grained spatial details from the encoder while suppressing irrelevant background information.

The encoder of DSNet follows a 5-level hierarchical structure designed to progressively extract multi-scale features from the 128 $\times$ 128 $\times$ 128 multi-modal MRI input volumes. It consists of repeating Residual Double Convolution (ResDoubleConv) blocks, where each block integrates dynamic convolutional neural network (DCNN) layers, 3D batch normalization, and Leaky ReLU activations. These ResDoubleConv blocks incorporate residual connections to facilitate gradient flow and mitigate the vanishing gradient problem like Resnet. Between each encoder level, downsampling is performed using 3D max-pooling (MaxPool3D) with a kernel size 2 $\times$ 2 $\times$ 2 and stride of 2, which halves the spatial resolution in all three dimensions while reducing computational complexity. The dynamic convolution mechanism within the DCNN layers generates adaptive attention weights over multiple convolution kernels (via a 3D adaptive average pooling layer followed by two Conv3D layers and a softmax), enabling the model to better capture the heterogeneous appearance and complex boundaries of brain tumor subregions such as whole tumor, tumor core.

On the other hand, the decoder of DSNet follows the encoder with five corresponding levels that gradually restore the spatial resolution back to the original 128 $\times$ 128 $\times$ 128 size. Following each ResDoubleConv block in the decoder, an upsampling operation is applied using the Nearest Neighbor interpolation method to double the spatial dimensions. Similar to the encoder, each decoder stage employs Residual Double Convolution blocks containing dynamic convolution layers, batch normalization, and Leaky ReLU activations with residual connections. The encoder and decoder pathways are linked through attention-based skip connections at every level. These attention mechanisms allow the decoder to selectively focus on and incorporate the most relevant fine-grained features from the corresponding encoder stages, significantly improving boundary delineation and localization accuracy for the different tumor subregions.

Finally, after the last decoder stage restores the full spatial resolution, the feature maps are passed through a final 3D convolutional layer followed by a voxel-wise softmax activation to produce probability maps for classes.

B. *TransBTS Architecture*

Wang et al. propose TransBTS as a hybrid architecture that explicitly combines a 3D CNN encoder–decoder with a Transformer bottleneck for multimodal brain tumor segmentation [2]. The network begins with a 3D CNN encoder that extracts local volumetric context and downsamples the input MRI scan into compact feature maps. Instead of decoding those features directly, the model reshapes them into a token sequence, adds positional embeddings, and feeds them into stacked Transformer layers so that long-range dependencies can be modeled across the 3D volume. After global feature modeling, the decoder maps the sequence back into a volumetric representation and progressively upsamples it to produce the final segmentation mask [2].

The architectural motivation is that dense segmentation requires both local and global information. Convolutions capture local 3D structure effectively, but they do not explicitly model distant relationships. Transformer self-attention, by contrast, can encode long-range interactions between arbitrary tokens. TransBTS therefore preserves the strengths of a volumetric

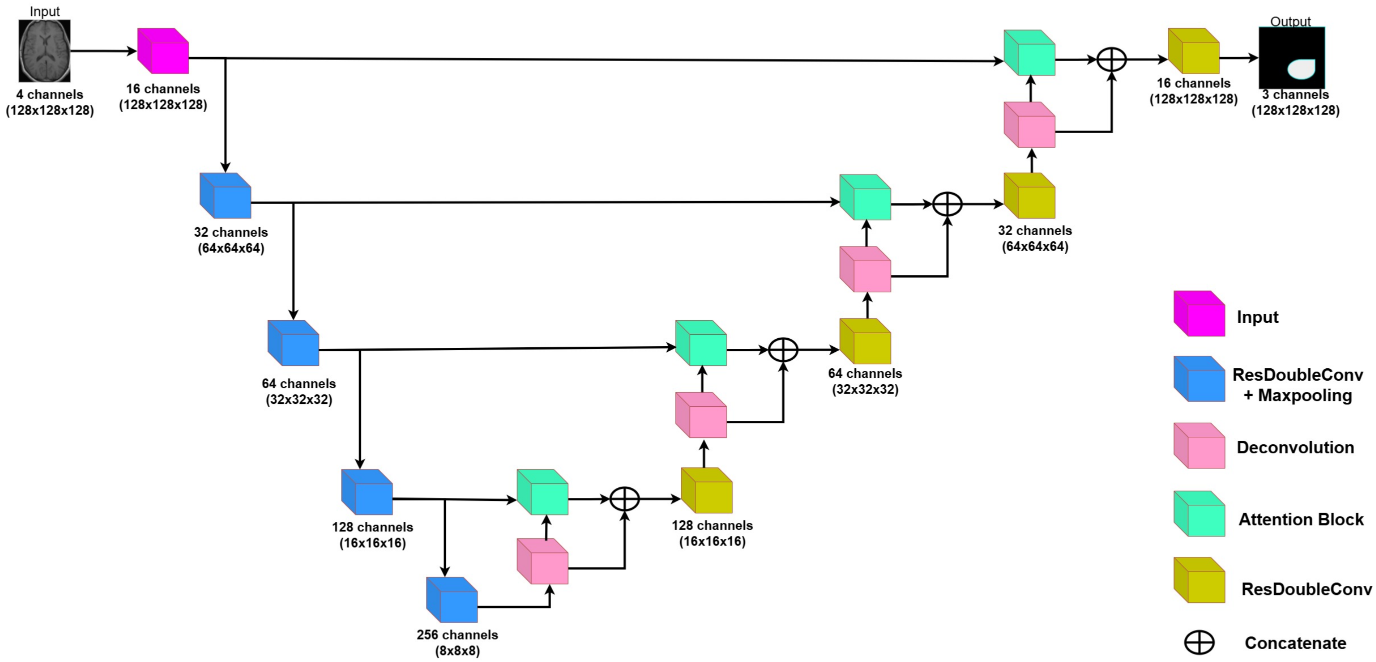

Fig. 4. DSNet segmentation model architecture.

CNN encoder while inserting a Transformer encoder in the latent space to improve global reasoning [2]. The paper uses a 3D CNN encoder with an overall stride of 8 so that the resulting feature map can be converted into 4096 tokens, each embedded to dimension 512 before entering four Transformer layers. The decoder then reshapes the final sequence back into a 3D feature volume and fuses skip-connected CNN features during progressive upsampling [2].

The ablation study clarifies why the architecture works. Increasing the sequence length by reducing the encoder stride from 16 to 8 improves Dice, indicating that token resolution matters for segmentation. The study also shows that CNN skip connections are more effective than Transformer-layer skip connections for recovering fine spatial detail, reinforcing the idea that the hybrid model depends on both branches doing different jobs well [2]. On the BraTS 2019 validation set, TransBTS with test-time augmentation reports Dice scores of 78.93 for ET, 90.00 for WT, and 81.94 for TC. On the BraTS 2020 validation set, it reports 78.73 for ET, 90.09 for WT, and 81.73 for TC, outperforming the 3D U-Net and V-Net baselines listed in the paper [2]. In architectural terms, TransBTS represents a hybrid route to performance: it strengthens volumetric segmentation by injecting explicit global dependency modeling into an otherwise CNN-dominated segmentation pipeline.

IV. Comparative Discussion

The four papers separate naturally into two architecture themes. V-Net and nnU-Net represent foundational volumetric segmentation systems, while TransBTS and DSNet represent hybrid architectures that strengthen a volumetric backbone with additional modeling components. This division is useful because it shows that progress in brain tumor segmentation does not come from one single direction. Some gains come from building a strong and stable 3D encoder–decoder foundation, while other gains come from extending that foundation with more advanced mechanisms for global reasoning, adaptive filtering, or feature refinement.

V-Net and nnU-Net share the core assumption that brain tumor segmentation benefits from processing the MRI as a full 3D volume rather than as disconnected 2D slices. In V-Net, that idea appears through a dedicated volumetric architecture built around residual stages, strided convolutions, deconvolutions, skip connections, and Dice-aware optimization [3]. Its main strength is that it directly models volumetric continuity and spatial context, which helps overcome one of the biggest weaknesses of slice-wise approaches. At the same time, V-Net also exposes several drawbacks of early 3D segmentation design. Its large volumetric kernels and full 3D feature processing increase memory consumption and computational cost, often forcing small batch sizes and making training more expensive. The requirement of fixed-size volumetric inputs also reduces flexibility, and repeated downsampling can still threaten the preservation of very fine boundary detail even when skip connections are used.

nnU-Net keeps the volumetric logic of 3D U-Net but shifts the source of improvement away from architectural novelty and toward task-aligned pipeline design [1]. This is one of its biggest conceptual strengths. The paper shows that a relatively plain 3D encoder–decoder can still reach state-of-the-art performance when training targets, augmentation, batch design, postprocessing, and model selection are carefully aligned to the benchmark. In other words, nnU-Net proves that performance depends not only on the backbone itself but also on how the entire pipeline is formulated. Its drawback, however, is that this strength is partially tied to benchmark alignment. The paper explicitly shows that postprocessing choices can improve BraTS ranking by removing very small enhancing-tumor predictions, even though those predictions may still matter clinically. This means nnU-Net is extremely strong as a benchmark-driven system, but some of its gains come from engineering decisions that may not transfer cleanly to all real clinical settings. It is also operationally heavy, since the strongest results depend on multiple modifications and large ensembles rather than on a single simple deployment model.

TransBTS and DSNet both start from the volumetric baseline and then strengthen it with more specialized representational machinery. TransBTS does this by inserting a Transformer bottleneck into a 3D CNN encoder–decoder pipeline [2]. Its main advantage is that it explicitly introduces long-range dependency modeling into volumetric segmentation. Convolutions are highly effective for local feature extraction, but they do not model distant spatial relationships as directly as self-attention. By reshaping compact 3D feature maps into tokens and processing them with Transformer layers, TransBTS improves global semantic reasoning across the tumor volume. This is especially valuable in brain tumor segmentation, where the overall shape and distributed context of the lesion can matter as much as local edge detail. The drawback is cost. Transformer self-attention scales poorly with token length, so the model must carefully control sequence size through encoder stride and tokenization design. The paper also reports substantial training requirements, including multi-GPU training and large computational overhead, which makes TransBTS more demanding than a conventional volumetric CNN. Its ablation study further shows that the model is sensitive to architecture decisions such as token resolution and skip-connection placement, meaning that its gains are real but not free.

DSNet also represents a hybrid design, but it strengthens volumetric segmentation in a different way [4]. Instead of introducing Transformer-based global reasoning, it improves the segmentation backbone through dynamic convolution, attention-guided skip reuse, residual feature extraction, and adversarial refinement. Architecturally, this gives DSNet a strong ability to adapt to heterogeneous tumor appearance and to refine difficult boundaries. The DCNN module allows the model to generate adaptive convolutional responses, while the attention mechanism helps the decoder emphasize the most relevant encoder features. The adversarial critic further encourages sharper and more realistic segmentation masks. These design choices make DSNet especially strong for challenging tumor subregions and help explain its high reported Dice performance across BraTS datasets [4]. The drawback is that DSNet becomes more complex in several directions at once. Dynamic convolution adds runtime and implementation overhead, attention increases dependency between encoder and decoder paths, and adversarial learning can make training more difficult to stabilize. In addition, DSNet extends beyond segmentation into survival prediction, which makes it clinically ambitious but also introduces a second task whose quality depends on smaller and more limited outcome data than the segmentation module itself.

Taken together, the four papers show that the main architectural trade-off in brain tumor segmentation is not simply "which model scores higher," but rather what kind of improvement is being purchased and at what cost. Foundational volumetric models such as V-Net and nnU-Net provide strong 3D segmentation logic with greater interpretability and clearer training behavior, but they may leave representational capacity on the table when the tumor structure is highly heterogeneous or when long-range context matters. Hybrid models such as TransBTS and DSNet recover some of that missing capacity by adding global attention, adaptive filtering, or adversarial refinement, but they do so at the cost of higher complexity, greater compute demand, and more difficult optimization. The overall lesson across the four papers is that strong brain tumor segmentation depends on how well an architecture balances volumetric context, feature transmission, global reasoning, training stability, and practical deployment constraints.

V. Conclusion

This paper examined four deep learning studies on brain tumor segmentation through two architecture themes: foundational volumetric segmentation and hybrid architectural refinement. Framing the papers this way makes the progression of the field easier to understand. V-Net and nnU-Net represent the foundational side of the literature because they establish the core logic behind strong 3D segmentation systems: volumetric encoder–decoder processing, skip-connected reconstruction, and training objectives aligned with highly imbalanced medical segmentation tasks. TransBTS and DSNet then show how that volumetric foundation can be strengthened with more specialized mechanisms, including Transformer-based global reasoning, dynamic convolution, attention-guided skip reuse, and adversarial refinement.

Taken together, the four papers show that deep learning for brain tumor segmentation has evolved beyond the simple question of whether 2D or 3D processing is better. The more important issue is how a model balances local detail, volumetric continuity, global context, feature reuse, and computational practicality. V-Net demonstrates the importance of fully volumetric segmentation and Dice-aware optimization as an architectural foundation. nnU-Net shows that a relatively plain 3D backbone can still achieve state-of-the-art performance when the pipeline is tightly aligned to benchmark targets, augmentation strategy, and postprocessing logic. TransBTS illustrates how hybridizing a volumetric CNN with a Transformer can improve long-range dependency modeling, while DSNet shows that adaptive filtering, attention, and adversarial refinement can further strengthen tumor boundary delineation and overall segmentation quality.

At the same time, the comparative analysis makes clear that every gain comes with a trade-off. Foundational volumetric systems are typically more stable, interpretable, and easier to analyze, but they may not capture all of the representational complexity needed for highly irregular tumor structure. Hybrid systems recover some of that missing capacity, yet they do so by increasing memory cost, architectural complexity, training difficulty, or engineering overhead. In this sense, the four papers do not point to one universally best design. Instead, they show that strong segmentation performance depends on matching the architecture to the actual bottleneck, whether that bottleneck is volumetric continuity, benchmark alignment, long-range reasoning, or adaptive feature selection.

Overall, the larger conclusion of this report is that progress in brain tumor segmentation comes from building on strong volumetric foundations rather than abandoning them. The most effective newer models do not replace the core encoder–decoder logic; they refine it. This makes the field's direction clearer: future work is likely to continue combining robust 3D segmentation backbones with selective hybrid enhancements that improve global understanding, boundary precision, and ultimately clinical usefulness.

References

[1] 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*, Lecture Notes in Computer Science, vol. 12659, pp. 118–132, 2021, doi: 10.1007/978-3-030-72087-2_11.

[2] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal brain tumor segmentation using transformer," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, vol. 12901, pp. 109–119, 2021, doi: 10.1007/978-3-030-87193-2_11.

[3] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. 4th Int. Conf. 3D Vision (3DV)*, pp. 565–571, 2016, doi: 10.1109/3DV.2016.79.

[4] A. Yeafi, M. Islam, and M. S. U. Yusuf, "A deep learning framework for 3D brain tumor segmentation and survival prediction," *Healthcare Analytics*, vol. 8, p. 100418, 2025, doi: 10.1016/j.health.2025.100418.