

Architectural Ablation Study of Conditional GANs for MRI Brain Tumor Segmentation: Resolving Checkerboard Artifacts and the PatchGAN Paradox

1st Dhruv Gupta
B.Tech CSE (AI Specialization)
Bennett University
Greater Noida, India
guptadhruv1402@gmail.com

2nd Devansh Gupta
B.Tech CSE (AI Specialization)
Bennett University
Greater Noida, India
S24CSEU0825@bennett.edu.in

Abstract—Automated brain tumor segmentation using Magnetic Resonance Imaging (MRI) is a critical task in neuro-oncology. While traditional Convolutional Neural Networks (CNNs) like the U-Net excel at spatial mapping, they often produce overly smoothed boundaries. This study explores Conditional Generative Adversarial Networks (cGANs) to enforce high-frequency anatomical realism. To determine the optimal discriminator architecture, we conducted a four-stage ablation study comparing a baseline U-Net against three PatchGAN variations. We mathematically demonstrate that a 3×3 kernel with a stride of 2 introduces destructive checkerboard artifacts. Furthermore, while correcting this flaw using a 3×3 stride-1 kernel with MaxPooling achieves an elite local segmentation loss (0.035), it empirically triggers the “PatchGAN Paradox”—where hyper-optimization for local texture actively degrades the global Structural Similarity Index Measure (SSIM) to 0.934. Ultimately, we establish that a 4×4 PatchGAN with a stride of 2 provides the mathematically optimal balance, achieving the highest global SSIM (0.961) while maintaining sharp, artifact-free tumor boundaries. Future directions propose advanced pre-training image enhancement techniques to further optimize these results.

Index Terms—Conditional GAN, Brain Tumor Segmentation, U-Net, PatchGAN, Checkerboard Artifacts, Image Enhancement

I. INTRODUCTION

Precise segmentation of brain tumors from MRI scans is essential for clinical diagnosis, surgical planning, and post-operative monitoring. Historically, the U-Net architecture has served as the baseline for this task due to its robust spatial mapping capabilities. However, standard U-Net models trained exclusively on pixel-wise loss functions (e.g., Binary Cross-Entropy) tend to predict overly safe, smoothed boundaries, failing to capture the complex, micro-anatomical textures of actual tumors.

To address this, Conditional Generative Adversarial Networks (cGANs) have been introduced to penalize the U-Net for generating unrealistic structures. This project investigates the architectural constraints required to stabilize a cGAN for medical segmentation. Specifically, at the request of the advisory faculty, we conduct a four-stage ablation study on the PatchGAN Discriminator to analyze the mathematical

impact of kernel sizes and spatial reduction strategies on the generation of high-fidelity tumor masks.

II. METHODOLOGY AND ARCHITECTURE

A. The Generator (U-Net Baseline)

The foundation of this study is a standard U-Net, trained exclusively on pixel-wise loss, designed to map 240×240 multi-channel MRI inputs to predicted tumor masks. The architecture features a symmetrical encoder-decoder structure. The contracting path captures context using repeated 3×3 convolutions, each followed by a Rectified Linear Unit (ReLU) activation and a 2×2 max pooling operation for spatial downsampling. The expansive path utilizes 2×2 transposed convolutions for upsampling, concatenated with the corresponding cropped feature maps from the contracting path via skip connections to recover spatial resolution. The final layer utilizes a 1×1 convolution mapped to a 4-channel output representing the specific tumor classes. To prevent the Generator from diverging during adversarial training, the objective function was heavily weighted with an L_1 Segmentation penalty ($\lambda_{seg} = 100$).

B. The Discriminator Ablation Study

To force the Generator to produce anatomically realistic textures, we introduced a PatchGAN Discriminator. We evaluated four distinct architectural configurations:

1) *Model A: Baseline U-Net (No GAN)*: Trained without an adversary to establish a baseline for global structural similarity and peak signal-to-noise ratio.

2) *Model B: cGAN with 3×3 Kernel (Stride 2)*: This model tested a 3×3 convolutional kernel with a stride of 2. As documented by Odena et al., an asymmetrical kernel-to-stride ratio causes uneven receptive overlap. Mathematically, this forces high-frequency noise and “checkerboard artifacts” into the generated masks.

3) *Model C: cGAN with 3×3 Kernel (Stride 1 + Pooling)*: To resolve the checkerboard artifacts, we applied a spatial “Separation of Concerns.” Convolutions were locked to a stride

of 1 for symmetrical feature extraction, and downsampling was handled entirely by 2×2 MaxPooling layers.

4) *Model D: Champion cGAN (4×4 Kernel, Stride 2)*: The final, optimized architecture utilized a standard Pix2Pix-style PatchGAN. Instead of outputting a single scalar probability, the network outputs a 2D matrix where each element evaluates a 70×70 overlapping patch of the input image for realism. The network processes a concatenated input of the MRI slice and the generated mask through a series of 4×4 Convolutional layers utilizing a stride of 2 for simultaneous feature extraction and downsampling. Each layer is followed by Batch Normalization and a LeakyReLU activation (slope = 0.2). A 4×4 kernel perfectly divides by a stride of 2, allowing for symmetrical overlap without structural degradation.

III. RESULTS AND DISCUSSION

A. Performance Metrics

The models were evaluated on an unseen validation set of 4,524 slices containing tumors. We tracked both local Segmentation Loss (to measure microscopic boundary accuracy) and the global Structural Similarity Index Measure (SSIM) (to measure macroscopic anatomical coherence).

TABLE I
COMPARATIVE PERFORMANCE OF SEGMENTATION ARCHITECTURES

Architecture Configuration	Evaluation Metrics		
	Seg Loss	PSNR (dB)	SSIM
Model A: Baseline U-Net	0.140	25.38	0.957
Model B: 3×3 , Stride 2	0.158	24.63	0.951
Model C: 3×3 , Stride 1	0.035	21.61	0.934
Model D: 4×4 , Stride 2	0.080	26.35	0.961

Note: Model D provides the optimal mathematical balance.

B. The PatchGAN Paradox and Final Selection

As demonstrated in Table I, the ablation study revealed a critical trade-off in generative medical segmentation:

- **The Artifact Flaw:** Model B performed the worst structurally due to the stride-induced checkerboard artifacting predicted by deep learning theory.
- **The PatchGAN Paradox:** Model C achieved an elite local Segmentation Loss of 0.035. However, its SSIM dropped significantly to 0.934. The Discriminator became so adept at penalizing local 70×70 patches that the Generator injected high-frequency micro-textures to fool it, actively destroying the smooth, global coherence required for a continuous clinical mask.
- **The Goldilocks Zone:** Model D (4×4 , Stride 2) emerged as the champion architecture. It successfully prevented checkerboard artifacts mathematically, resulting in the highest global SSIM (0.961) and PSNR (26.35 dB), while simultaneously improving the local boundary sharpness over the baseline U-Net.

C. Clinical Implications

In a clinical setting, radiomics and surgical planning rely on continuous, highly confident tumor boundaries. The 0.961 SSIM achieved by Model D indicates that the generated masks maintain the macroscopic integrity required by neurosurgeons, while the adversarial training successfully combats the boundary-blurring typically seen in standard CNNs.

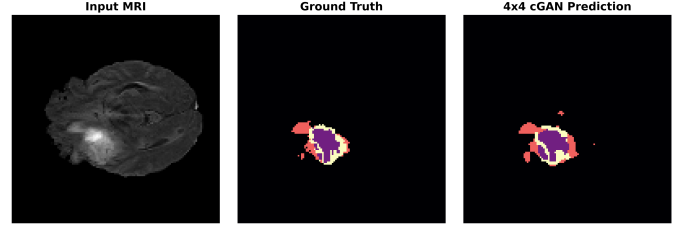


Fig. 1. Visual comparison of the final 4×4 generated tumor mask vs. Ground Truth. The model successfully balances sharp local boundaries with global structural coherence.

IV. CONCLUSION AND FUTURE WORK

This study demonstrates that while cGANs are highly effective at sharpening medical image segmentations, architecture selection is critical. We empirically proved that 3×3 kernels with even strides introduce destructive checkerboard artifacts. Furthermore, while correcting this flaw with stride-1 pooling yields exceptional local texture accuracy, it falls victim to the PatchGAN Paradox, prioritizing local realism over global structural integrity. The 4×4 PatchGAN with a stride of 2 provides the mathematically optimal balance, successfully sharpening tumor boundaries without sacrificing the continuous, global structure of the MRI scans.

Future Work: While the current 4×4 architecture effectively optimizes the generative mapping, the quality of the raw input data remains a bounding factor. Moving forward, we plan to implement a rigorous pre-processing pipeline to enhance the MRI images and ground-truth masks prior to training. Specifically, we intend to apply Contrast Limited Adaptive Histogram Equalization (CLAHE) to the raw MRI slices to amplify the visibility of low-contrast tumor boundaries [4]. Additionally, morphological enhancement techniques will be applied to the training masks to eliminate human-annotation noise. By maximizing the structural clarity of the data before it enters the adversarial loop, we hypothesize the cGAN will achieve even greater clinical precision.

REFERENCES

- [1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” *CVPR*, 2017.
- [2] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” *ICLR*, 2016.
- [3] A. Odena, V. Dumoulin, and C. Olah, “Deconvolution and Checkerboard Artifacts,” *Distill*, 2016.
- [4] S. M. Pizer et al., “Adaptive histogram equalization and its variations,” *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355-368, 1987.