# Geometry-Aware Point-to-Voxel Fusion via Gaussian Splatting for Intracranial Aneurysm Segmentation

Tian Qin[1*], Zhongji Zhang[2*], Yuqi Feng[3], Tianyu Huang[3], Zihan Yin[2], and Tao Luo[3†]

[1] The University of Sydney, Sydney, NSW, Australia
tqin0515@uni.sydney.edu.au
[2] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[3] Beijing University of Posts and Telecommunications, Beijing, China
tluo@bupt.edu.cn

**Abstract.** Intracranial aneurysms (IAs) carry high mortality upon rupture, and their accurate segmentation from CTA is essential for clinical intervention. However, many IAs are small and morphologically resemble normal vascular bends, rendering them difficult to distinguish by intensity alone. Voxel-only networks frequently misclassify these subtle protrusions as tortuous healthy vessels, while point-based architectures capture geometry but lack dense semantic context. To bridge this gap, we propose GeoP2VNet, a geometry-aware method that injects explicit surface geometry into a voxel backbone via three modules: Geometric Feature Extraction (GFE) computes rotation-invariant point descriptors; a differentiable Point-to-Voxel (P2V) Gaussian splatting layer projects sparse geometric cues onto aligned voxel features; and Dual-Gated Residual Fusion (DGRF) selectively amplifies geometric priors at anatomically significant regions while suppressing them in normal segments. On 205 CTA cases (266 lesions) with 5-fold cross-validation, GeoP2VNet achieves a Dice of 0.770, outperforming nnU-Net (0.663) and SwinUNETR (0.578), while reducing false positives by 92%, from 0.317 to 0.024 per case. Code is available at `https://github.com/somtiannes/GeoP2VNet`.

**Keywords:** Intracranial Aneurysm · Segmentation · Point-Voxel Fusion · Gaussian Splatting

## 1 Introduction

Intracranial aneurysms (IAs) are pathological, localized deformations of the cerebral arteries. While often asymptomatic, their rupture precipitates subarachnoid hemorrhage (SAH), a catastrophic event with high mortality [2]. Clinical practice relies on manual delineation by expert neuroradiologists—a process that is

[*] Equal contribution. [†] Corresponding author.

time-intensive and difficult to scale [21]. Deep learning solutions are increasingly deployed to alleviate this burden, yet a fundamental challenge persists: many IAs are small (often <5 mm) [19] and morphologically indistinguishable from normal vascular bends on CTA intensity alone.

Regardless of pipeline architecture, a refinement stage is needed to delineate fine boundaries from coarse ROIs. Existing approaches face complementary limitations. Voxel-based methods [3,7,16,17] apply uniform filters that cannot differentiate geometric importance across locations; repeated downsampling further discards boundary details critical for sub-5 mm lesions [5], and heavy reliance on intensity statistics produces false positives where aneurysms exhibit similar intensities to tortuous healthy vessels. Foundation models such as SAM [12] struggle with low-contrast medical targets and produce inter-slice artifacts in 3D vascular reconstruction. Point-based architectures [4,14,15] excel at geometric modeling but yield sparse manifolds decoupled from global CTA context. In summary, intensity-only methods lack location-aware geometric sensitivity, while geometry-only methods lack dense semantic grounding; neither alone suffices.

Most existing multi-modal fusion works combine “image with image,” failing to bridge heterogeneous representations such as point clouds and voxel grids. Yet the pathological hallmark of an IA is inherently geometric: a focal outpouching disrupting the smooth tubular profile of the parent artery. Hemodynamic studies confirm that aneurysm shape is distinguishable via curvature, surface irregularity, and density indices [20], and that geometry-aware priors enhance performance on IA-related tasks [10, 19]. This motivates fusing explicit point-cloud geometry with voxel-grid intensity, an approach effective in other 3D domains [18].

We propose GeoP2VNet, which extracts explicit geometric features from vessel-surface point clouds and projects them into a voxel backbone via differentiable Gaussian splatting. A dual-gated residual fusion (DGRF) mechanism then dynamically modulates the importance of these geometric priors across spatial locations, amplifying them at anatomically critical regions while suppressing noise in normal tubular segments, thereby bridging the “geometric blindness” of voxel networks and the “contextual blindness” of point networks.

The contributions are: (1) a lightweight Geometric Feature Extraction (GFE) module encoding rotation-invariant surface descriptors; (2) a differentiable Point-to-Voxel (P2V) Gaussian splatting layer for geometry-to-voxel alignment; and (3) a dual-gated residual fusion mechanism performing location-adaptive geometric injection.

## 2 Method

### 2.1 Overview of GeoP2VNet

GeoP2VNet is a geometry-aware hybrid framework for fine-grained aneurysm segmentation: our method focuses exclusively on the segmentation stage and fuses explicit point-cloud geometry with CTA intensity (Fig. 1). Given a cropped

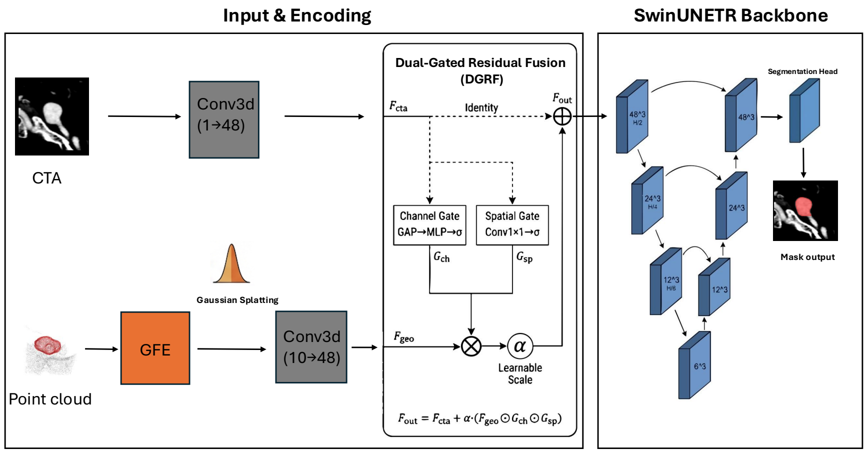



Fig. 1: Overview of GeoP2VNet: GFE extracts 10-dimensional geometric features, P2V projects them to voxel space, and DGRF injects them into the CTA stream before Swin-UNETR segmentation.

CTA volume $\mathbf{X}_{img}$ and a vascular surface point cloud $\mathcal{P}$, the pipeline proceeds in three stages. **(1) GFE** computes a 10-dimensional rotation-invariant descriptor per surface point, encoding curvature, density, and normal consistency. These intensity-independent features directly capture wall morphology and distinguish aneurysmal protrusions from tubular vessels. **(2) P2V** projects these sparse point-level features into the dense voxel grid via a differentiable fixed-kernel operation, ensuring spatial alignment with the CTA backbone while preserving sub-voxel continuity. **(3) DGRF** injects the projected geometric map into the CTA feature stream via channel and spatial gating: the channel gate selects informative descriptors while the spatial gate amplifies priors at critical locations (e.g., bifurcations) and suppresses them in normal segments, acting as a position-aware denoiser. The fused representation is decoded by a Swin-UNETR [3] backbone. This design ensures geometry serves as structured inductive bias rather than unconstrained auxiliary input.

### 2.2 Explicit Geometric Feature Extraction

GFE computes a 10-dimensional signature $\mathbf{f}_i^{geo} \in \mathbb{R}^{10}$ for each point using a multi-scale $k$-NN graph ($k_s$=8, $k_l$=32) and local PCA: NC, $|\mathbf{n}|$, $\sigma_n$, $\kappa$, $\rho$, $d_c$, $r$, $h$, NV, and $\Delta\kappa$. These handcrafted, rotation-invariant descriptors cover local smoothness (NC, NV), curvature ($\kappa$, $\Delta\kappa$), and spatial distribution ($\rho$, $d_c$, $r$, $h$), providing intensity-independent priors that distinguish aneurysmal protrusions from tubular vessels across both shape and scale.

GFE is *train-free* and lightweight: it uses no learnable parameters and relies only on local $k$-NN and PCA, avoiding the need for point-cloud pretraining or extra GPU memory for a point encoder. In contrast, PointNet-based encoders [14, 15] and point-based volumetric backbones [4] learn hierarchical features via shared MLPs and require large annotated 3D datasets to train; they introduce substantial parameter overhead and risk overfitting when geometry is auxiliary to a voxel backbone. By fixing geometric descriptors as handcrafted, rotation-invariant statistics, GFE keeps the pipeline simple and interpretable while still supplying the voxel stream with discriminative surface cues (e.g., curvature and density at bifurcations) that learned point encoders would require additional data and capacity to capture.

### 2.3 Differentiable Gaussian Splatting for Point-to-Voxel Projection

To bridge points and voxels, P2V projects point features into voxel space with a fixed isotropic Gaussian kernel. Each point $p_i$ contributes to voxels in a $(2r+1)^3$ neighborhood ($r$=2, $K$=125), giving $\mathcal{O}(NK)$ complexity. The projected volume is:

$$\mathbf{V}_{geo}(v) = \frac{\sum_{p_i \in \mathcal{N}(v)} w_i(v) \cdot \mathbf{f}_i^{geo}}{\sum_{p_i \in \mathcal{N}(v)} w_i(v) + \epsilon} \tag{1}$$

where $\mathcal{N}(v)$ denotes points in the local $5^3$ neighborhood and $w_i(v)$ is

$$w_i(v) = \exp\left(-\frac{\|v - p_i\|^2}{2\sigma^2}\right) \tag{2}$$

The bandwidth is set to $\sigma$=1.0 voxel with $\epsilon$=$10^{-6}$; when $\mathcal{N}(v) = \emptyset$, Eq. (1) yields $\mathbf{V}_{geo}(v) = \mathbf{0}$. This formulation draws on the Gaussian splatting primitive [8, 9], i.e., weighted accumulation onto a grid, but employs fixed kernels for geometric projection rather than learnable rendering [11], preserving sub-voxel continuity via weight normalization.

### 2.4 Dual-Gated Residual Fusion

Geometric features are more discriminative at lesion and bifurcation sites but may introduce noise in normal segments. DGRF dynamically adjusts each descriptor's importance per spatial location, conditioned on local CTA context [6, 13]. CTA and geometric features are first projected to a shared dimension ($C$=48):

$$\mathbf{F}_{cta} = \psi_{cta}(\mathbf{X}_{img}), \quad \mathbf{F}_{geo} = \psi_{geo}(\mathbf{V}_{geo}), \quad \mathbf{F}_{cta}, \mathbf{F}_{geo} \in \mathbb{R}^{C \times D \times H \times W} \tag{3}$$

where $\psi_{cta}$ is Conv3d(1→48, kernel 3×3×3) and $\psi_{geo}$ is Conv3d(10→48, kernel 1×1×1). Two CTA-guided gates are then computed: a channel gate $\mathbf{G}_{ch}$ (GAP+MLP, $r$=4) and a spatial gate $\mathbf{G}_{sp}$ (1×1×1 convolution):

$$\mathbf{G}_{ch} = \sigma(\mathbf{W}_2 \cdot \mathrm{ReLU}(\mathbf{W}_1 \cdot \mathrm{GAP}(\mathbf{F}_{cta}))) \in \mathbb{R}^{C \times 1 \times 1 \times 1} \tag{4}$$

$$\mathbf{G}_{sp} = \sigma(\mathbf{W}_{sp} * \mathbf{F}_{cta}) \in \mathbb{R}^{1 \times D \times H \times W} \tag{5}$$

The final fusion is

$$\mathbf{F}_{out} = \mathbf{F}_{cta} + \alpha \cdot (\mathbf{F}_{geo} \odot \mathbf{G}_{ch} \odot \mathbf{G}_{sp}) \tag{6}$$

where $\odot$ is broadcasted element-wise multiplication and $\alpha$ is a learnable scalar initialized to 0.1.

$\mathbf{G}_{ch}$ selects informative descriptors while $\mathbf{G}_{sp}$ modulates priors per local anatomy; $\alpha$ is optimized end-to-end, with the residual path preserving CTA semantics.

## 3 Experiments

### 3.1 Dataset

**Cohort.** A total of 205 CTA scans (266 unruptured saccular aneurysms) were collected from Beijing Tiantan Hospital, Capital Medical University, over 2012–2024. Inclusion criteria required confirmed unruptured, saccular aneurysms; dissecting, fusiform, blister-like, and AVM-related aneurysms were excluded. The study was approved by the institutional review board (IRB) of Beijing Tiantan Hospital, Capital Medical University (KY2026-044-02), and oral informed consent was obtained. Scans were acquired on multi-vendor scanners (e.g., Siemens, Philips) with varying acquisition protocols across sites.

**Annotation.** Two physicians independently annotated aneurysm sac and neck; the parent vessel was excluded. Disagreements were resolved by consensus, with inter-observer consistency assessed on a double-annotated subset.

**Evaluation Protocol.** All experiments use one shared patient-level 5-fold split. In each fold, vessel, ROI, and IA models are trained only on training patients and evaluated on held-out patients; held-out cases never enter training, validation, tuning, or model selection.

Table 1: Dataset statistics ($n$ = 266 lesions from 205 cases): lesion size and anatomical distribution.

| **Lesion Size** | $n$ **(%)** | **Location** | $n$ **(%)** |
|---|---|---|---|
| Small (<3 mm) | 63 (23.7) | ICA | 162 (60.9) |
| Medium (3–5 mm) | 124 (46.6) | MCA | 36 (13.5) |
| Large (5–10 mm) | 57 (21.4) | ACA | 26 (9.8) |
| Very Large (≥10 mm) | 22 (8.3) | BA / VA / AComA | 42 (15.8) |

### 3.2 Experimental Setup

We evaluate given-ROI segmentation rather than the full detection-to-segmentation pipeline. Candidate ROIs are generated by nnDetection [1] (84.6% lesion recall);

missed lesions use oracle crops, and all methods receive the same $96^3$ patches. Vessel masks come from a frozen fold-wise nnU-Net [7] (mean Dice 77.35%); held-out masks are predicted from CTA only, converted to whole-brain surface point clouds, and cropped to each ROI.

**Implementation Details.** The voxel backbone is Swin-UNETR [3], trained with a combined Dice and cross-entropy loss using AdamW for 500 epochs on three NVIDIA RTX 4090 GPUs. Inference time per $96^3$ patch is 127.41 ms for GeoP2VNet versus 56.58 ms for SwinUNETR (2.25× overhead), which remains clinically acceptable given the substantial accuracy gains.

### 3.3 Comparison with State-of-the-Art Methods

GeoP2VNet is compared against SwinUNETR (voxel-only baseline) and nnU-Net [7]. As shown in Table 2, the proposed method achieves consistent improvements across all metrics.

Table 2: Quantitative comparison ($N$=205, 5-fold CV, mean ± 95% CI). DSC/IoU/HD95 are voxel-level; Recall is lesion-level; FPs/case = false-positive components per case.

| Method | DSC↑ | IoU↑ | HD95↓ | Recall↑ | FPs↓ |
|---|---|---|---|---|---|
| SwinUNETR | 0.578±.044 | 0.470±.039 | 3.76±.96 | 75.00±5.24 | 0.317±.09 |
| nnU-Net | 0.663±.037 | 0.547±.036 | 3.63±.84 | 96.15±2.42 | 0.122±.05 |
| **GeoP2VNet** | **0.770±.024** | **0.652±.026** | **1.25±.32** | **98.08±1.80** | **0.024±.02** |

GeoP2VNet improves DSC by +0.192 over SwinUNETR and +0.107 over nnU-Net, while reducing FPs by 92% and HD95 from 3.76 to 1.25 mm. These gains indicate that explicit geometric priors resolve boundary ambiguities that neither Transformer-based SwinUNETR nor self-configuring nnU-Net can address through intensity cues alone.

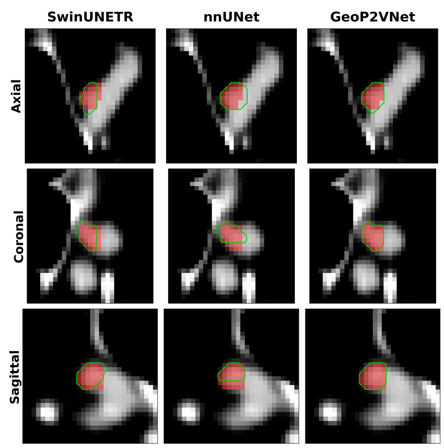
### 3.4 Ablation Study

Table 3 summarizes ablation results comparing baseline Swin-UNETR, a variant with direct geometric integration but no gating, and the full GeoP2VNet with DGRF.

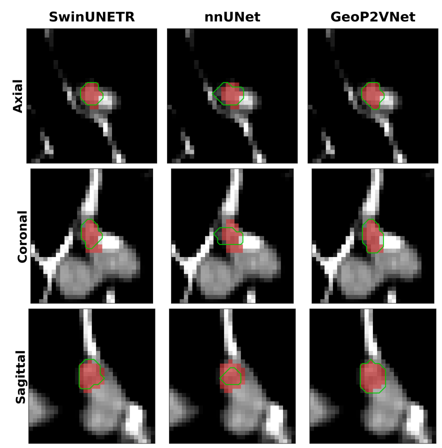
Variant 2, without gating, improves Recall to 84.62% and halves FPs from 0.317 to 0.195, confirming that raw geometric features help distinguish pathological protrusions. However, the DSC gain is only +0.013 because uniform injection also introduces noise from normal vessels. Adding DGRF in Variant 3 raises DSC to 0.770 and drops FPs to 0.024: the spatial gate suppresses responses in normal segments while the channel gate selects discriminative descriptors, refining raw priors into targeted boundary corrections through position-aware denoising.

Table 3: Ablation study ($N$=205 cases). GFE: Geometric Feature Extraction; DGRF: Dual-Gated Residual Fusion.

| **Variant** | **DSC↑** | **IoU↑** | **HD95↓** | **Recall↑** | **FPs↓** |
|---|---|---|---|---|---|
| 1 (Baseline) | 0.578±.044 | 0.470±.039 | 3.76±.96 | 75.00±5.24 | 0.317±.093 |
| 2 (w/o Gating) | 0.591±.040 | 0.475±.038 | 3.42±.76 | 84.62±4.38 | 0.195±.054 |
| **3 (Full)** | **0.770±.024** | **0.652±.026** | **1.25±.32** | **98.08±1.80** | **0.024±.021** |




(a) Case 1




(b) Case 2

Fig. 2: Qualitative results. Rows: Axial, Coronal, Sagittal views. Columns: SwinUNETR, nnUNet, GeoP2VNet. Red: ground truth; Green: prediction.

# 4 Analysis

## 4.1 Qualitative Results

Fig. 2 presents two small aneurysm cases: Case 1, with max. diameter of 3.91 mm and 120 voxels, and Case 2, with max. diameter of 3.42 mm and 122 voxels, where precise boundary delineation is most challenging. In Case 1, SwinUNETR and nnUNet both exhibit boundary offsets in the Coronal and Sagittal planes, particularly at the neck–parent-vessel junction. GeoP2VNet achieves sub-voxel accuracy, with HD95 of 0.057 mm, 0.455 mm, and 0.793 mm and Dice of 0.899, 0.793, and 0.774 for the proposed method, SwinUNETR, and nnUNet, respectively. Case 2 confirms this trend: nnUNet under-segments the dome with Recall of 0.795, while SwinUNETR over-extends in the Sagittal view, achieving Precision of 0.912 but Recall of only 0.762. In contrast, GeoP2VNet maintains tight contour alignment across all planes, with Dice of 0.928, HD95 of 0.110 mm, and Precision/Recall of 0.957/0.902. On these sub-4 mm lesions, geometric priors provide a substantial boundary advantage over intensity-only methods.

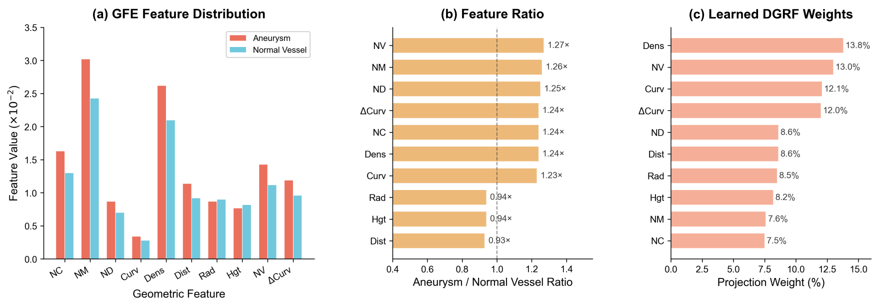



Fig. 3: Feature discriminability analysis. (a) Mean feature values for aneurysm vs. normal vessel regions across ten GFE descriptors. (b) Aneurysm-to-vessel ratio per descriptor; values above the dashed line (= 1.0) indicate higher discriminability. (c) Learned DGRF channel projection weights, showing that the network assigns the largest weights to the most discriminative descriptors (Density, NV, Curvature).

### 4.2 Feature Discriminability Analysis

Fig. 3 visualizes GFE feature distributions. As shown in Panel a, aneurysm regions exhibit systematically higher values than normal vessels. Panel b reveals that NV, Normal Magnitude, and Curvature $\kappa$ show the strongest discriminative power, with aneurysm-to-vessel ratios of 1.27×, 1.26×, and 1.23×, respectively. The learned DGRF projection weights in Panel c assign the highest channel attention to Density at 13.8% and NV at 13.0%, both of which capture local point density and surface irregularity at bifurcations and aneurysm-prone junctions, confirming that DGRF automatically prioritizes anatomically discriminative cues, consistent with its position-aware denoising role.

## 5 Conclusion and Limitation

In this paper, we present GeoP2VNet, a geometry-aware framework that addresses boundary ambiguity in intracranial aneurysm segmentation by injecting explicit surface geometry into a voxel backbone via differentiable Gaussian splatting. Three modules serve complementary roles: GFE encodes intensity-independent surface descriptors, P2V projects sparse geometric cues into dense voxel space, and DGRF performs position-aware gating to amplify geometric priors at critical anatomical sites while suppressing noise in normal segments. On 205 CTA cases, GeoP2VNet achieves a Dice of 0.770 (+0.192 over SwinUNETR) and reduces HD95 from 3.76 to 1.25 mm with 92% fewer false positives. Ablation analysis confirms that gated fusion is essential, as ungated injection yields only +0.013 DSC whereas position-aware gating unlocks the full gain. These findings suggest that structured geometric inductive bias is critical for small-lesion boundary precision and that the proposed fusion paradigm may extend to other vascular pathologies where surface morphology carries diagnostic significance.

**Limitation.** GeoP2VNet evaluates segmentation given candidate ROIs and pre-computed vessel masks, not a full screening pipeline. The study is in-domain; public IA resources differ in access, modality, or mask format, so external CTA validation remains future work. Broader comparisons with vessel-aware pipelines and joint vessel–IA modeling will also be pursued.

**Acknowledgments.** Supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 82402367). The study was approved by the institutional review board (IRB) of Beijing Tiantan Hospital, Capital Medical University (Approval No. KY2026-044-02).

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

## References

1. Baumgartner, M., Jäger, P.F., Isensee, F., Maier-Hein, K.H.: nndetection: a self-configuring method for medical object detection. In: MICCAI. pp. 530–539 (2021)
2. Brown, R.D., Broderick, J.P.: Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening. The Lancet Neurology **13**(4), 393–404 (2014)
3. Hatamizadeh, A., Nath, V., Tang, Y., et al.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
4. Ho, N.V., Nguyen, T., et al.: Point-unet: A context-aware point-based neural network for volumetric segmentation. In: MICCAI. pp. 644–655. Springer (2021)
5. Hu, X., Wang, Y., Fuxin, L., Samaras, D., Chen, C.: Topology-aware segmentation using discrete morse theory. arXiv preprint arXiv:2103.09992 (2021)
6. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. arXiv preprint arXiv:1802.04712 (2018)
7. Isensee, F., Jaeger, P.F., et al.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Kerbl, B., Kopanas, G., et al.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
9. Liu, H., Liu, B., Hu, Q., Du, P., Li, J., Bao, Y., Wang, F.: A review on 3d gaussian splatting for sparse view reconstruction. Artificial Intelligence Review **58**(7), 215 (2025)
10. Luo, X., Wang, J., et al.: Prediction of cerebral aneurysm rupture using a point cloud neural network. Journal of NeuroInterventional Surgery **15**(4), 380–386 (2023)
11. Marzol, K., Kolton, I., Smolak-DyĹL'ewska, W., Kaleta, J., Mazur, M., Spurek, P., et al.: Medgs: Gaussian splatting for multi-modal 3d medical imaging. arXiv preprint arXiv:2509.16806 (2025)
12. Mazurowski, M.A., Dong, H., Gu, H., et al.: Segment anything model for medical image analysis: an experimental study. Medical Image Analysis **89**, 102918 (2023)
13. Oktay, O., Schlemper, J., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
14. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: CVPR. pp. 652–660 (2017)

15. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: NeurIPS. pp. 5099–5108 (2017)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
17. Shi, Z., Miao, C., Schoepf, U.J., et al.: A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. Nature communications **11**(1), 6090 (2020)
18. Tang, H., Liu, Z., Zhao, S., Lin, Y., Lin, J., Wang, H., Han, S.: Searching efficient 3d architectures with sparse point-voxel convolution. In: European conference on computer vision. pp. 685–702. Springer (2020)
19. Wei, J., Song, X., Wei, X., et al.: Knowledge-augmented deep learning for segmenting and detecting cerebral aneurysms with ct angiography: a multicenter study. Radiology **312**(2), e233197 (2024)
20. Xiang, J., Natarajan, S.K., Tremmel, M., Ma, D., Mocco, J., Hopkins, L.N., Siddiqui, A.H., Levy, E.I., Meng, H.: Hemodynamic–morphologic discriminants for intracranial aneurysm rupture. Stroke **42**(1), 144–152 (2011)
21. Zhou, Z., Jin, Y., Ye, H., Zhang, X., Liu, J., Zhang, W.: Classification, detection, and segmentation performance of image-based ai in intracranial aneurysm: a systematic review. BMC Medical Imaging **24**(1), 164 (2024)