VoxShield: Protecting 3D Medical Datasets from Unauthorized Training via Frequency-Aware Inter-Slice Disruption

Xinyao Liu[1,2], Zhipeng Deng[1✉], Wenhan Jiang[1], Haolin Wang[3], Xun Lin[4], Yafei Ou[5], and Yefeng Zheng[1✉]

[1] Westlake University, Hangzhou, China
{dengzhipeng, jiangwenhan, zhengyefeng}@westlake.edu.cn
[2] Dalian University of Technology, Dalian, China
xinyao.liu266@outlook.com
[3] Hokkaido University, Sapporo, Japan
[4] The Chinese University of Hong Kong, Hong Kong SAR, China
[5] RIKEN, Japan
yafei.ou@riken.jp

Abstract. The release of public 3D medical image segmentation (MIS) datasets accelerates clinical research but simultaneously heightens risks of unauthorized AI model training. While **Unlearnable Examples (UE)** offer protection by injecting imperceptible perturbations to prevent effective model learning, existing methods primarily target 2D scenarios. They neglect the *volumetric spatial correlations* and *inter-slice anatomical consistency* inherent in 3D medical volumes, which serve as critical learning priors for 3D segmentation networks. To bridge this gap, we propose **VoxShield**, a UE framework that explicitly targets the volumetric inductive biases of 3D networks. Our core insight is that by systematically dismantling the cross-slice continuity that 3D architectures rely on, we can fundamentally impair their spatial aggregation process. Specifically, we introduce an *Inter-Slice Frequency Consistency Disruption* mechanism that maximizes the spectral divergence between adjacent slices, injecting structural incoherence along the z-axis. Complementing this structural attack, a *Semantic Prediction Disruption* module is incorporated. By maximizing the ℓ_1 divergence between clean and perturbed logits, it forces the injected noise to penetrate the entire network and corrupt the final semantic mapping. Experiments on BraTS19 and FLARE21 demonstrate that VoxShield successfully degrades 3D segmentation performance, reducing the DSC from 80.0% to near 0.0% and from 88.6% to 6.8%, respectively. All protections are achieved with minimal perturbation (ϵ=4/255) to preserve high visual fidelity. The code is available at https://github.com/KK266299/VoxShield.

Keywords: Unlearnable Examples · Data Protection · 3D Medical Image Segmentation

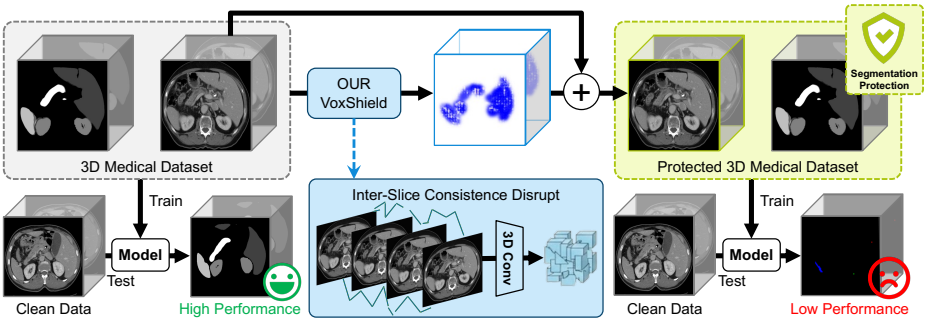

Fig. 1. The defense mechanism of VoxShield. **Left:** A model trained on clean data learns coherent volumetric representations and segments accurately on clean test volumes. **Right:** VoxShield disrupts inter-slice consistency by injecting spectrally diverse perturbations across adjacent slices. Consequently, a model trained on such protected data fails to extract coherent cross-slice features, yielding severely degraded segmentation on clean test inputs.

1 Introduction

3D medical image segmentation (MIS) is fundamental to disease diagnosis, treatment planning, and surgical navigation. The success of modern 3D segmentation models [4, 16, 10, 8, 20] critically depends on large-scale, voxel-annotated datasets. To facilitate clinical collaboration and education, institutions increasingly share these valuable datasets through public repositories like TCIA [1], often with complementary findings published in literature indexed by PubMed [2]. Crucially, such sharing is typically governed by strict data-use agreements and licensing terms (e.g., CC BY-NC) that permit educational and non-commercial research, while restricting unauthorized AI model training. Unfortunately, legal restrictions alone cannot physically prevent data scraping. Once in the public domain, 3D medical datasets are routinely downloaded and repurposed to train unauthorized networks. This misuse violates intellectual property rights and privacy laws, including the GDPR [6, 11], highlighting an urgent need for proactive, technical data protection mechanisms.

To counter such risks, Unlearnable Examples (UE) [9] offer an appealing proactive defense by injecting imperceptible perturbations into training data *before* release. Models trained on the perturbed data learn "shortcut" patterns instead of genuine semantic features, and consequently fail to generalize to clean test inputs. Representative methods include EM [9], which crafts error-minimizing noise via bi-level optimization; TAP [7], which leverages targeted adversarial perturbations; SEP [13], which stabilizes noise through a stable error-minimizing objective; LSP [25], which aligns perturbations with linearly separable features; and CUDA [18], which generates class-wise perturbations via convolution. While UE have been explored for 2D image classification and more recently extended to 2D segmentation [19] and 3D point clouds [23], existing

methods remain ineffective for 3D medical volumetric segmentation due to a critical architectural blind spot: they neglect the volumetric priors unique to 3D data.

Modern 3D segmentation architectures rely on a key inductive bias: *anatomical inter-slice consistency* [10, 5]. Organs and lesions vary smoothly along the z-axis, and 3D convolution kernels explicitly aggregate cross-slice context to exploit this continuity. The importance of this prior is evidenced by a well-established empirical finding: under severe anisotropy, where large inter-slice spacing naturally disrupts volumetric continuity, 2D models consistently outperform their 3D counterparts [10, 5]. Although voxel spacing is an intrinsic acquisition parameter that cannot be altered post hoc, this observation reveals that inter-slice consistency is the key factor governing 3D segmentation performance. This further suggests that deliberately disrupting the cross-slice aggregation process of 3D convolutions via injecting inter-slice incoherent perturbations along the z-axis can directly undermine the very inductive bias these models depend on, thereby providing a principled attack vector against 3D architectures.

As illustrated in Fig. 1, our key insight is that inter-slice consistency is simultaneously the source of 3D segmentation power and its exploitable vulnerability. Adjacent slices of a volume naturally share similar in-plane spatial patterns, which is a property that 3D convolutions heavily rely on to aggregate contextual cues along z. We exploit this by injecting perturbations whose *in-plane spectral characteristics* vary maximally between neighboring slices, so that 3D kernels spanning the z-axis receive conflicting frequency patterns at every spatial position, fundamentally undermining their ability to learn coherent volumetric representations.

Building on this insight, we propose **VoxShield**, a UE framework tailored for 3D medical image segmentation. VoxShield comprises two core modules: (i) an *Inter-Slice Frequency Consistency Disruption* module that maximizes spectral discrepancies between adjacent slices in the frequency domain, injecting "inter-slice flickering" that directly undermines the z-axis continuity assumption of 3D models; and (ii) a *Semantic Prediction Disruption* module that maximizes the ℓ_1 distance between clean and perturbed logits, forcing perturbations to propagate through the entire network and corrupt its semantic understanding.

2 Methodology

2.1 Preliminaries

3D Medical Image Segmentation. Given a clean training set $\mathcal{D}_{\text{clean}} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$, where $\boldsymbol{x}_i \in \mathbb{R}^{C\times D\times H\times W}$ is a 3D medical volume with D slices and $\boldsymbol{y}_i \in \{0,1\}^{D\times H\times W\times K}$ is the corresponding voxel-wise annotation over K classes, 3D medical image segmentation aims to optimize a network $\mathcal{F}(\cdot\,;\theta)$ to establish the volumetric mapping from $\boldsymbol{x}$ to $\boldsymbol{y}$. This optimization can be formulated as:

$$\arg\min_{\theta}\ \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}_{\text{clean}}}\big[\mathcal{L}_{\text{seg}}\big(\mathcal{F}(\boldsymbol{x};\theta),\,\boldsymbol{y}\big)\big], \tag{1}$$

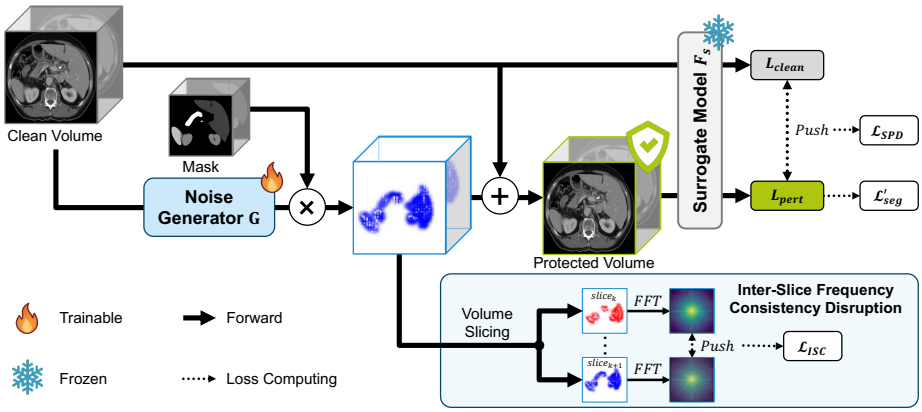

Fig. 2. Overview of the proposed VoxShield framework. A noise generator $\mathcal{G}$ produces perturbations that are added to clean 3D medical images to create protected volumes. A surrogate model $\mathcal{F}_s$ computes $\mathcal{L}_{\text{SPD}}$ and $\mathcal{L}'_{\text{seg}}$ to drive protected images away from the original decision boundaries. An inter-slice frequency consistency disruption module enforces $\mathcal{L}_{\text{ISC}}$ by maximizing spectral discrepancy between adjacent slices, destroying the volumetric continuity that 3D models rely on.

where $\mathcal{L}_{\text{seg}}$ is the segmentation loss (*e.g.*, cross-entropy loss and Dice loss) and θ denotes the trainable parameters of $\mathcal{F}$.

Unlearnable Examples. UE are designed to prevent unauthorized models from extracting useful knowledge by adding imperceptible perturbations to the training images of $\mathcal{D}_{\text{clean}}$. For better understanding, we take the representative error-minimizing (EM) noise [9] as an example. EM jointly optimizes a surrogate model and sample-wise perturbations to generate protective noise:

$$\min_{\theta} \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}_{\text{clean}}}\left[\min_{\boldsymbol{\delta}\in\mathcal{I}} \mathcal{L}_{\text{seg}}\big(\mathcal{F}_s(\boldsymbol{x}+\boldsymbol{\delta};\theta),\,\boldsymbol{y}\big)\right], \tag{2}$$

where $\mathcal{F}_s$ is a surrogate model, $\mathcal{I}=\{\boldsymbol{\delta} : \|\boldsymbol{\delta}\|_\infty \le \epsilon\}$ is the feasible region that ensures imperceptibility, and $\boldsymbol{x}+\boldsymbol{\delta}$ denotes the protected image (also called the unlearnable example). In this process, θ and $\boldsymbol{\delta}$ are alternately updated to minimize $\mathcal{L}_{\text{seg}}$, yielding the protected dataset $\mathcal{D}_{\text{protect}}=\{(\boldsymbol{x}_i+\boldsymbol{\delta}_i,\,\boldsymbol{y}_i)\}_{i=1}^{N}$.

Task Objective. The *data protector* has access to $\mathcal{D}_{\text{clean}}$ and publishes $\mathcal{D}_{\text{protect}}$. The *data exploiter* trains a 3D segmentation architecture on $\mathcal{D}_{\text{protect}}$. We evaluate the protection effectiveness by testing the model trained by the exploiter on held-out clean data: if the test-time performance degrades significantly, the protection is considered successful. The protector aims to minimize such test-time performance while keeping perturbations imperceptible.

2.2 VoxShield Framework

Overview. Existing UE fail in 3D medical segmentation because they ignore *inter-slice anatomical consistency*, the key inductive bias exploited by 3D models.

As illustrated in Fig. 2, VoxShield introduces two dedicated modules: (i) *Inter-Slice Frequency Consistency Disruption* maximizes the spectral discrepancy of perturbations between adjacent slices in the frequency domain, breaking the inter-slice continuity prior that 3D models rely on; (ii) *Semantic Prediction Disruption* maximizes the ℓ_1 distance between clean and perturbed logits to corrupt semantic predictions. To focus perturbations on anatomically meaningful regions, the raw generator output is masked by a binary region-of-interest mask $\boldsymbol{M} \in \{0,1\}^{D\times H\times W}$ derived from $\boldsymbol{y}$:

$$\boldsymbol{\delta} = \mathrm{clamp}\big(\mathcal{G}(\boldsymbol{x}) \odot \boldsymbol{M},\ -\epsilon,\ \epsilon\big), \tag{3}$$

where $\odot$ denotes element-wise multiplication. The protected volume is then $\boldsymbol{x}_p = \boldsymbol{x} + \boldsymbol{\delta}$.

2.3 Inter-Slice Frequency Consistency Disruption

3D segmentation networks fundamentally assume *volumetric continuity*, i.e., anatomical structures change smoothly along z. 3D convolution kernels with receptive fields spanning k_z slices aggregate cross-slice features under this assumption. Our strategy is to violate this assumption by forcing the perturbation's spectral characteristics to vary *maximally* between consecutive slices.

For each slice z, we compute the 2D amplitude spectrum of the perturbation: $\boldsymbol{S}_z = |\mathcal{F}_{\mathrm{2D}}(\boldsymbol{\delta}_z)| \in \mathbb{R}^{H\times W}$, where $\boldsymbol{\delta}_z \in \mathbb{R}^{H\times W}$ is the z-th slice of the generated noise. The ISC loss is:

$$\mathcal{L}_{\mathrm{ISC}} = -\frac{1}{D-1}\sum_{z=1}^{D-1} \big\|\boldsymbol{S}_{z+1} - \boldsymbol{S}_z\big\|_2. \tag{4}$$

Minimizing $\mathcal{L}_{\mathrm{ISC}}$ (i.e., maximizing inter-slice spectral divergence) forces drastic spectral shifts between adjacent slices, injecting high-frequency "spectral jitter" along z that fundamentally contradicts the smooth anatomical continuity 3D convolutions rely on. We measure divergence in the *frequency domain* rather than the spatial domain because magnitude spectra are translation-invariant [24, 22], providing a more stable and texture-focused measure of inter-slice dissimilarity.

2.4 Semantic Prediction Disruption

Effective data protection requires that perturbations corrupt not only low-level pixel statistics but also the high-level semantic mapping learned by the surrogate. If the surrogate can still produce similar predictions on clean and perturbed inputs, the learned representations remain intact and protection fails. We therefore explicitly maximize the discrepancy between the surrogate's outputs on clean and perturbed volumes, forcing perturbations to propagate through the entire network and disrupt its semantic understanding.

Let $\boldsymbol{L}_{\mathrm{clean}} = \mathcal{F}_s(\boldsymbol{x};\theta)$ and $\boldsymbol{L}_{\mathrm{pert}} = \mathcal{F}_s(\boldsymbol{x}_p;\theta)$ denote the clean and perturbed logit volumes ($\in \mathbb{R}^{D\times H\times W\times K}$), respectively. We maximize the ℓ_1 distance between them:

$$\mathcal{L}_{\mathrm{SPD}} = -\|\,\boldsymbol{L}_{\mathrm{pert}} - \boldsymbol{L}_{\mathrm{clean}}\|_1. \tag{5}$$

Algorithm 1: VoxShield Training Procedure

Input: Clean dataset $\mathcal{D}_{\text{clean}}$, surrogate $\mathcal{F}_s$, generator $\mathcal{G}$, budget ϵ, weights $\lambda_{\text{SPD}}, \lambda_{\text{ISC}}$, learning rates α_g
Output: Protected dataset $\mathcal{D}_{\text{protect}}$

for epoch $= 1$ **to** n **do**
  for each $(\boldsymbol{x}, \boldsymbol{y})$ in $\mathcal{D}_{\text{clean}}$ **do**
    $\boldsymbol{\delta} \leftarrow \text{Clamp}_{[-\epsilon,\epsilon]}[\mathcal{G}(\boldsymbol{x}) \odot \boldsymbol{M}]$; $\boldsymbol{x}_p \leftarrow \text{Clamp}_{[0,1]}[\boldsymbol{x} + \boldsymbol{\delta}]$
    Compute $\mathcal{L}_{\text{total}}$ // *Eq.* (6)
    $\theta_{\mathcal{G}} \leftarrow \theta_{\mathcal{G}} - \alpha_g \nabla_{\theta_{\mathcal{G}}} \mathcal{L}_{\text{total}}$
  end for
end for
for each $(\boldsymbol{x}, \boldsymbol{y})$ in $\mathcal{D}_{\text{clean}}$ **do**
  $\mathcal{D}_{\text{protect}} \leftarrow \mathcal{D}_{\text{protect}} \cup \{(\text{Clamp}_{[0,1]}[\boldsymbol{x} + \mathcal{G}(\boldsymbol{x})],\ \boldsymbol{y})\}$
end for
return $\mathcal{D}_{\text{protect}}$

Compared with the ℓ_2 norm, the ℓ_1 norm is less dominated by a few large-deviation voxels, thereby encouraging the generator to degrade predictions broadly across the volume rather than at isolated locations.

2.5 Optimization

The total generator objective combines segmentation, logits divergence, and inter-slice consistency disruption:

$$\mathcal{L}_{\text{total}} = \mathcal{L}'_{\text{seg}}\big(\mathcal{F}_s(\boldsymbol{x}_p; \theta),\ \boldsymbol{y}\big) + \lambda_{\text{SPD}}\, \mathcal{L}_{\text{SPD}} + \lambda_{\text{ISC}}\, \mathcal{L}_{\text{ISC}}, \tag{6}$$

where $\mathcal{L}'_{\text{seg}}$ (the combination of Dice and cross entropy loss) anchors training stability by ensuring the surrogate continues to learn on perturbed data (thus providing meaningful gradients), while $\mathcal{L}_{\text{SPD}}$ and $\mathcal{L}_{\text{ISC}}$ inject semantic and structural corruption, respectively. λ_{SPD} and λ_{ISC} balance the three terms.

The optimization of VoxShield follows a bi-level alternating strategy between the surrogate model $\mathcal{F}_s$ and the noise generator $\mathcal{G}$. During the generator's update phase, $\mathcal{F}_s$ is frozen to provide consistent adversarial gradients for perturbation refinement. This iterative cycle ensures that $\mathcal{G}$ learns to synthesize noise patterns that are specifically destructive to 3D volumetric priors. The detailed optimization steps for the generator are summarized in Algorithm 1. Upon convergence, the finalized generator is deployed as a fixed protective operator to transform the clean dataset $\mathcal{D}_{\text{clean}}$ into its unlearnable counterpart, $\mathcal{D}_{\text{protect}}$.

3 Experiments and Results

3.1 Experimental Setup

Datasets. We evaluated the protection effectiveness of VoxShield on two 3D MIS datasets: **BraTS19** [15] (brain tumor MRI) and **FLARE21** [14] (abdominal

Table 1. Protection effectiveness on victim 3D UNet. DSC (%) ↓ and HD95 (mm) ↑ indicate stronger protection; PSNR (dB) ↑ and SSIM ↑ measure perturbation imperceptibility. Best UE results in **bold**.

Method	BraTS19				FLARE21			
	DSC↓	HD95↑	PSNR↑	SSIM↑	DSC↓	HD95↑	PSNR↑	SSIM↑
Clean	80.0	16.0	–	–	88.6	10.2	–	–
EM [9]	69.3	30.2	39.7	0.9219	57.1	90.1	30.0	0.9312
LSP [25]	80.5	15.0	39.6	0.9916	88.5	17.5	43.0	0.9970
SEP [13]	72.3	16.7	33.9	0.7544	88.4	15.7	40.8	0.9511
TAP [7]	68.4	22.7	38.0	0.8792	45.7	136.7	37.7	0.9084
PUE [21]	60.0	26.5	40.2	0.9279	36.1	150.2	37.9	0.9155
UMed [12]	40.2	34.7	54.9	0.9990	26.3	107.3	48.0	0.9973
VoxShield	$< 1e^{-3}$	**217.4**	**56.3**	**0.9994**	**6.8**	**172.4**	**49.0**	**0.9980**

organ CT), covering different anatomical regions and imaging modalities. All volumes were resized to $160 \times 160 \times 160$, and intensities were normalized to the range of $[0, 1]$.

Models. We used a 3D UNet [4] implemented in MONAI for both the surrogate model and the generator for UE generation. To evaluate cross-architecture transferability, we trained four victim models: *3D-UNet* [4], *Attention UNet* [17], *UNet++* [26], and *TransUNet* [3].

Baselines. We compared against: (1) *Clean*; (2) *EM* [9]; (3) *LSP* [25]; (4) *SEP* [13]; (5) *TAP* [7]; (6) *PUE* [21]; (7) *UMed* [12]. All baselines were adapted to 3D volumetric inputs.

Metrics and Implementation. We reported the Dice Similarity Coefficient (DSC, %) and 95th-percentile Hausdorff Distance (HD95, mm). Lower DSC and higher HD95 on victim models reflect more effective data protection. The perturbation budget was $\epsilon = 4/255$. We trained VoxShield for 100 epochs with Adam optimizer (lr=1e-4) on a single NVIDIA RTX 5090 GPU. The loss weights were set to $\lambda_{\text{ISC}} = 0.2$ and $\lambda_{\text{SPD}} = 0.05$.

3.2 Main Results

Table 1 reported the protection performance on both datasets. VoxShield achieved a dramatic reduction in victim model DSC, substantially outperforming all baselines on both datasets. 2D-UE (EM, SEP, LSP, PUE, UMed) showed limited effectiveness on 3D data, because their slice-independent noise patterns were partially smoothed by 3D convolution kernels, confirming our hypothesis that disrupting inter-slice consistency is essential for effective 3D protection.

Cross-Architecture Transferability. Table 2 demonstrated that VoxShield perturbations transferred effectively across architectures. Since the perturbations targeted the fundamental volumetric continuity prior shared by *all* 3D architectures (not architecture-specific features), the protection generalized from

Table 2. Cross-architecture transferability on FLARE21 (Mean DSC %). Noise generated with UNet surrogate, evaluated on different victim architectures.

Method	3D-UNet	Att-UNet	UNet++	TransUNet
Clean	88.6	83.2	83.4	80.7
EM	57.1	48.6	68.6	74.2
VoxShield	**6.8**	**9.0**	**10.8**	**3.1**

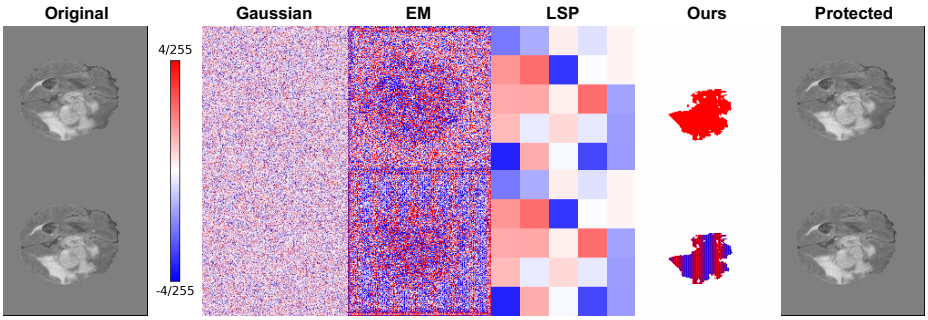

Fig. 3. Visualization of BraTS19 noise patterns across consecutive z-slices.

Table 3. Ablation study on FLARE21. Best results in **bold**.

Configuration	DSC (%) ↓	HD95 (mm) ↑
VoxShield (full)	**6.8**	**172.4**
w/o $\mathcal{L}_{\text{SPD}}$ (no logits divergence)	24.8	113.5
w/o $\mathcal{L}_{\text{ISC}}$ (no ISC disruption)	32.9	168.6

the UNet surrogate to Attention UNet, UNet++, and even Transformer-based TransUNet.

Ablation Study. Table 3 validated the contribution of each component on FLARE21. Removing $\mathcal{L}_{\text{ISC}}$ caused the largest performance drop, confirming that inter-slice spectral diversity is the most critical attack vector against 3D segmentation. Removing $\mathcal{L}_{\text{SPD}}$ also degraded protection, as the generator lost its ability to corrupt deep semantic representations.

Inter-Slice Frequency Consistency Disruption. We visualized the noise patterns across consecutive z-slices for different methods. VoxShield produced spectrally diverse perturbations that shifted drastically between adjacent slices, while 2D methods (EM, SEP, LSP, PUE, UMed) generated temporally coherent noise that 3D kernels could readily filter. Fig. 3 illustrates these differences.

Robustness to Preprocessing and Defenses. We evaluated simple mean and median filtering as potential preprocessing defenses, and VoxShield maintained strong protection under these local smoothing operations. This is because such filtering cannot reconstruct the inter-slice spectral coherence deliberately

disrupted by VoxShield. A broader evaluation against defense-aware frequency augmentation and purification-based defenses, especially generative purification, remains an important direction for future work.

4 Conclusion

To the best of our knowledge, VoxShield is the first UE framework that explicitly targets inter-slice continuity priors in 3D medical segmentation. By targeting inter-slice anatomical consistency and effectively degrading segmentation performance, VoxShield creates the noise to disrupt structural continuity and semantic predictions. VoxShield achieves the strongest protection among the evaluated baselines, demonstrating that securing 3D data requires attacking the volumetric priors fundamental to 3D models.

Acknowledgments. This work was supported by Zhejiang Leading Innovative and Entrepreneur Team Introduction Program (2024R01007) and the "Pioneer" and "Leading Goose" Research and Development Program of Zhejiang (2025C02077).

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

References

1. The cancer imaging archive (tcia) (2025), https://www.cancerimagingarchive.net/, accessed: 2025-01-15
2. Pubmed (2025), https://pubmed.ncbi.nlm.nih.gov/, accessed: 2025-01-15
3. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 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. Dong, Z., He, Y., Qi, X., Chen, Y., Shu, H., Coatrieux, J.L., Yang, G., Li, S.: Mnet: rethinking 2d/3d networks for anisotropic medical image segmentation. arXiv preprint arXiv:2205.04846 (2022)
6. European Parliament and Council of the European Union: Regulation (eu) 2016/679 (general data protection regulation). Official Journal of the European Union, L 119, 1–88 (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj
7. Fowl, L., Goldblum, M., Chiang, P.y., Geiping, J., Czaja, W., Goldstein, T.: Adversarial examples make strong poisons. Advances in Neural Information Processing Systems **34**, 30339–30351 (2021)
8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)

9. Huang, H., Ma, X., Erfani, S.M., Bailey, J., Wang, Y.: Unlearnable examples: Making personal data unexploitable. arXiv preprint arXiv:2101.04898 (2021)
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
11. Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F.: Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence **2**(6), 305–311 (2020)
12. Lin, X., Yu, Y., Xia, S., Jiang, J., Wang, H., Yu, Z., Liu, Y., Fu, Y., Wang, S., Tang, W., et al.: Safeguarding medical image segmentation datasets against unauthorized training via contour-and texture-aware perturbations. arXiv preprint arXiv:2403.14250 (2024)
13. Liu, Y., Xu, K., Chen, X., Sun, L.: Stable unlearnable example: Enhancing the robustness of unlearnable examples via stable error-minimizing noise. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 3783–3791 (2024)
14. Ma, J., Zhang, Y., Gu, S., An, X., Wang, Z., Ge, C., Wang, C., Zhang, F., Wang, Y., Xu, Y., et al.: Fast and low-gpu-memory abdomen ct organ segmentation: the flare challenge. Medical Image Analysis **82**, 102616 (2022)
15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
17. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
18. Sadasivan, V.S., Soltanolkotabi, M., Feizi, S.: Cuda: Convolution-based unlearnable datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3862–3871 (2023)
19. Sun, Y., Zhang, H., Zhang, T., Ma, X., Jiang, Y.G.: Unseg: One universal unlearnable example generator is enough against all image segmentation. Advances in Neural Information Processing Systems **37**, 79168–79193 (2024)
20. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20730–20740 (2022)
21. Wang, D., Xue, M., Li, B., Camtepe, S., Zhu, L.: Provably unlearnable data examples. arXiv preprint arXiv:2405.03316 (2024)
22. Wang, H., Wu, X., Huang, Z., Xing, E.P.: High-frequency component helps explain the generalization of convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8684–8694 (2020)
23. Wang, X., Li, M., Liu, W., Zhang, H., Hu, S., Zhang, Y., Zhou, Z., Jin, H.: Unlearnable 3d point clouds: Class-wise transformation is all you need. Advances in Neural Information Processing Systems **37**, 99404–99432 (2024)
24. Yin, D., Gontijo Lopes, R., Shlens, J., Cubuk, E.D., Gilmer, J.: A fourier perspective on model robustness in computer vision. Advances in Neural Information Processing Systems **32** (2019)

25. Yu, D., Zhang, H., Chen, W., Yin, J., Liu, T.Y.: Availability attacks create shortcuts. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 2367–2376 (2022)
26. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: International workshop on deep learning in medical image analysis. pp. 3–11. Springer (2018)