VolMicro: Experimental Results

1 Experiments

1.1 Dataset and evaluation protocol

We evaluate on a single 3D neuron microscopy volume of size $100 \times 647 \times 813$ (D×H×W), totaling 52,601,100 voxels and occupying 105.2 MB in `uint16` (Table 4). To exploit extreme background sparsity, we apply our sparse occupancy gating (TOPS-Gate) with threshold $\tau = 0.5$ to select 8,452,906 voxels (16.1%) for sparse optimization. Importantly, decoding and quality metrics are always computed on the full 52.6M-voxel grid.

1.2 Baselines and metrics

We compare VolMicro—a sparse mixture of anisotropic 3D Gaussian primitives inspired by Gaussian splatting representations [2]—against:

- **Traditional codecs:** JPEG2000 [6], HEVC/x265 (H.265/HEVC) [5], and ZFP [3].
- **Neural implicit baseline:** COIN [1] with a SIREN decoder [4], which decodes by evaluating a coordinate MLP at every voxel location.

We report PSNR↑, SSIM↑ [7], LPIPS↓ [8], and bitrate in bits-per-voxel (bpp)↓. For efficiency we measure full-volume decode time (ms)↓ and throughput (MVox/s)↑ on an NVIDIA RTX 3080.

1.3 Matched-bitrate comparison (core result)

Table 1 provides a strict, storage-matched comparison at 0.073 bpp. COIN achieves higher PSNR (39.23 vs. 36.58 dB), but VolMicro improves structural and perceptual quality (SSIM 0.932 vs. 0.923; LPIPS 0.278 vs. 0.316). The key advantage is decoding speed: VolMicro reconstructs the entire 52.6M-voxel volume in 1,032 ms versus 19,424 ms for COIN, a **19**× speed-up (51.0 vs. 2.7 MVox/s). This reflects a fundamental computational difference: coordinate-field decoders (COIN/SIREN) require a network forward pass per voxel [1, 4], whereas VolMicro decodes by evaluating a sparse list of primitives (Gaussians) whose contributions can be vectorized and parallelized efficiently [2].

1.4 Rate–distortion context

Table 2 places VolMicro in the broader compression landscape. At moderate bitrates (≈0.316–0.333 bpp), JPEG2000 and HEVC achieve ∼41 dB PSNR with high SSIM, highlighting that classical codecs remain strong when a few megabytes are available [6, 5]. ZFP operates at much higher bitrate (10.23 bpp), yielding near-lossless reconstruction but only 3.3× compression [3]. In the extreme-compression regime (0.073 bpp), neural representations provide sub-megabyte models: VolMicro trades PSNR for structural/perceptual fidelity, yielding higher SSIM [7] and lower LPIPS [8] than COIN at the same bitrate [1].

1.5 Model size accounting and storage efficiency

VolMicro stores 22 bytes per Gaussian in float16 (Table 3). With 20,693 Gaussians, the compressed payload is 0.46 MB, giving 0.073 bpp and a 231× compression ratio (Table 5). This explicit accounting makes storage costs transparent: the footprint is directly determined by the primitive list. (Unlike INR compression, where the stored object is a quantized neural decoder [1].)

1.6 Training dynamics

As shown in Fig. 1, the optimization follows a typical *coarse-to-fine* trajectory for Gaussian-mixture fitting with *dynamic model capacity*, consistent with density-control schedules used in Gaussian splatting pipelines [2]. Early in training, the objective decreases rapidly as the model captures dominant intensity structure. During this phase, the curves exhibit a repeated *spike–recovery* behaviour: the loss briefly increases and PSNR temporarily drops, then both recover and improve. This pattern is expected when the representation undergoes *structural updates* (e.g., densification/splitting/cloning of primitives and pruning of redundant ones), which momentarily perturbs the reconstruction before re-optimization converges to a better local optimum. Once the Gaussian count stabilizes (mid-to-late training), the loss and PSNR curves become smooth, indicating a transition from discrete structural changes to continuous fine-tuning.

Table 1: **Fair comparison at matched bitrate (0.073 bpp).** Both methods use an identical storage budget. Decoding benchmark on NVIDIA RTX 3080 for full volume reconstruction (52.6M voxels). VolMicro achieves **19× faster** decoding by avoiding per-voxel MLP evaluation required by INR decoders [1, 4].

Method	**PSNR** ↑ (dB)	**SSIM** ↑ [7]	**LPIPS** ↓ [8]	**Decode** (ms) ↓	**Throughput** (MVox/s)
COIN (SIREN) [1, 4]	39.23	0.923	0.316	19,424	2.7
VolMicro (Ours)	36.58	0.932	0.278	1,032	51.0
Δ: −2.65 dB PSNR, +0.009 SSIM, −0.038 LPIPS, **19**× faster decode					

Table 2: Compression performance on neuron microscopy volume ($100 \times 647 \times 813$ voxels, 16-bit, 105.2 MB). Traditional baselines: JPEG2000 [6], HEVC [5], ZFP [3]. Neural baseline: COIN [1] with SIREN [4]. Metrics: SSIM [7], LPIPS [8].

Method	**PSNR** ↑ (dB)	**SSIM** ↑	**LPIPS** ↓	**bpp** ↓	**Ratio**	**Size** (MB)
Traditional Codecs						
JPEG2000-3D [6]	41.09	0.935	0.204	0.333	101×	2.09
HEVC (x265) [5]	41.62	0.943	0.015	0.316	106×	1.98
ZFP [3]	87.45	1.000	0.000	10.23	3.3×	64.12
Neural Implicit Representation						
COIN (SIREN) [1, 4]	39.23	0.923	0.316	0.073	231×	0.46
3D Gaussian Mixture (Ours)						
VolMicro (Gaussian primitives) [2]	36.58	0.932	0.278	0.073	231×	0.46

Table 3: Per-Gaussian storage breakdown (float16). Total: 22 bytes/Gaussian.

Parameter	**Format**	**Size (bytes)**
Position $\boldsymbol{\mu} \in \mathbb{R}^3$	3× float16	6
Scale $\boldsymbol{\sigma} \in \mathbb{R}^3$	3× float16	6
Rotation $\mathbf{q} \in \mathbb{H}$	4× float16	8
Intensity $\alpha \in \mathbb{R}$	1× float16	2
Total per Gaussian		**22**

Table 4: VolMicro model configuration and training details.

Parameter	Value
Dataset	
Volume dimensions	$100 \times 647 \times 813$ (D $\times$ H $\times$ W)
Total voxels	52,601,100
Original size	105.2 MB (uint16)
TOPS-Gate threshold	$\tau = 0.5$
Gated voxels	8,452,906 (16.1%)
Model	
Number of Gaussians	20,693
Parameters per Gaussian	11 (position: 3, scale: 3, rotation: 4, intensity: 1)
Total parameters	227,623
Compressed size	0.46 MB (float16)
Training	
Epochs	10,000
Learning rate	0.01
KNN neighbors	$k = 32$
Edge boost factor	3.0
Sparsity weight	$\lambda_s = 0.001$
Results	
PSNR	**36.58 dB**
SSIM [7]	0.932
LPIPS [8]	0.278
Compression ratio	**231**$\times$
Bits per voxel	0.073 bpp

Table 5: Compression ratio calculation.

Component	Size	Calculation
Original volume	105.20 MB	$100 \times 647 \times 813 \times 2$ bytes
Compressed	0.46 MB	$20{,}693 \times 22$ bytes
Compression ratio	**231**$\times$	$105.20 \div 0.46$
Bits per voxel	0.073 bpp	$(0.46 \times 8) \div 52.6$M

training_metrics.pdf

Figure 1: **VolMicro training dynamics over 10,000 epochs (coarse-to-fine capacity scheduling). (a) Total objective (log scale).** The overall loss decreases by orders of magnitude and exhibits a sawtooth pattern in early training: short spikes followed by rapid recovery. These transients coincide with discrete *capacity update* events (densification/splitting/cloning and pruning), analogous to density-control schedules in Gaussian primitive optimization [2]. **(b) Reconstruction quality (PSNR).** PSNR improves rapidly while the model captures dominant structure, then saturates; temporary drops align with spikes in (a), reflecting the same structural updates. The dashed line marks the best observed value (36.61 dB). **(c) Reconstruction error (MSE, log scale).** MSE mirrors (a), confirming that most improvement is driven by reduced reconstruction error and stabilizes once structural updates subside. **(d) Model capacity (number of Gaussians).** The primitive count increases during densification (peaking at $\sim$28.5k) and then decreases through pruning to a compact final representation of 20,693 Gaussians (dashed line). Once capacity stabilizes, optimization transitions to continuous fine-tuning and the curves in (a–c) become smooth.

References

[1] Emilien Dupont, Adam Goliński, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. Coin: Compression with implicit neural representations. *arXiv preprint arXiv:2103.03123*, 2021. URL `https://arxiv.org/abs/2103.03123`.

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. doi: 10.1145/3592433.

[3] Peter Lindstrom. Fixed-rate compressed floating-point arrays. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2674–2683, 2014. doi: 10.1109/TVCG.2014.2346458.

[4] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL `https://arxiv.org/abs/2006.09661`.

[5] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12): 1649–1668, 2012. doi: 10.1109/TCSVT.2012.2221191.

[6] David S. Taubman and Michael W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2001.

[7] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

[8] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *arXiv preprint arXiv:1801.03924*, 2018. URL `https://arxiv.org/abs/1801.03924`.