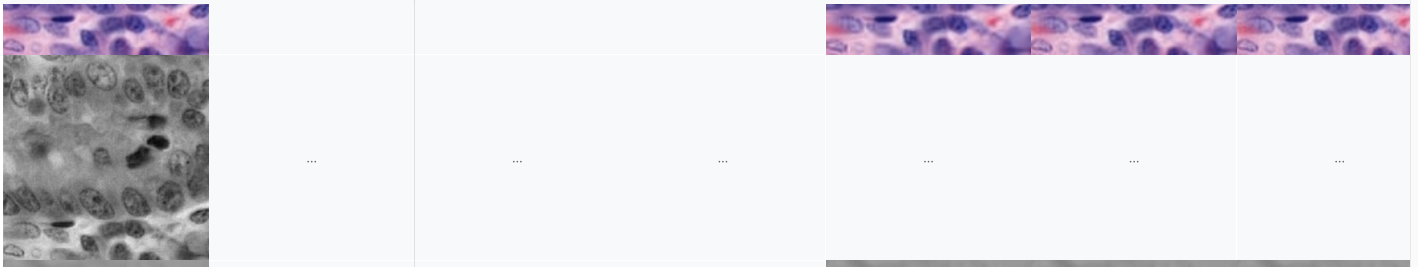
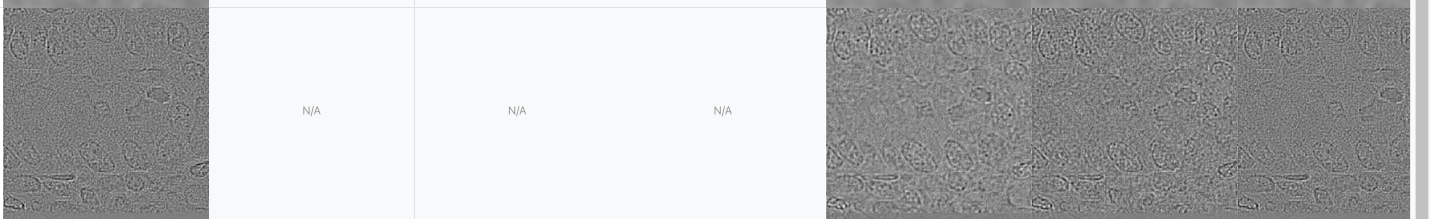
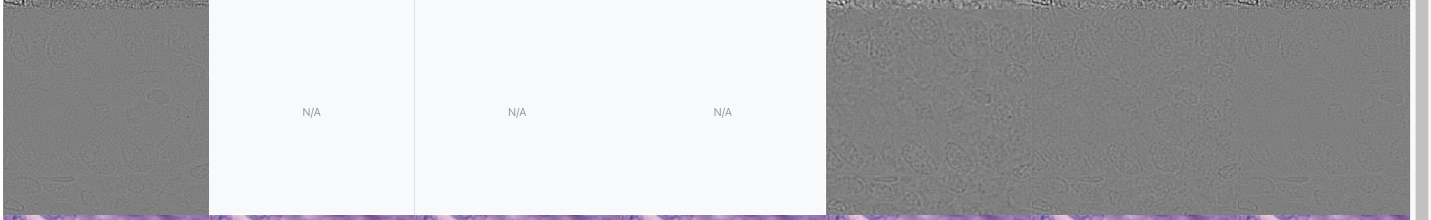
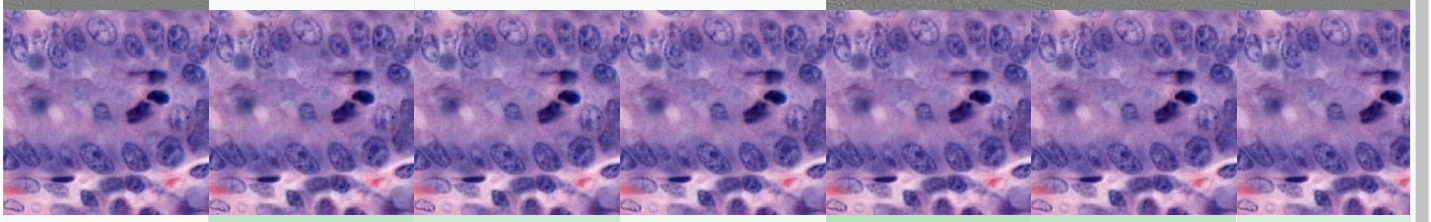
ORIGAMI Multi-Run Comparison Viewer

▶ Pipeline Diagram

Tile: L0_0_2 (L0 at 0,2) Level: L0, Position: (0, 2) Add Column: Select run... Add Load All

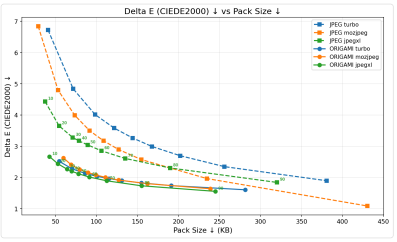
Zoom: Fit Actual (1:1) 2x

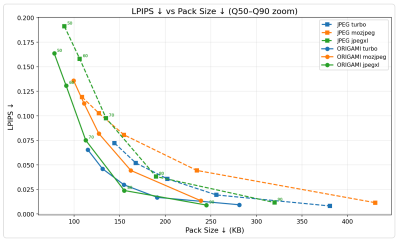
Stage	Original	JPEG turbo 30	JPEG turbo 60	JPEG turbo 90	ORIGAMI turbo 30	ORIGAMI turbo 60	ORIGAMI turbo 90
Y (Luma) Original Luma channel extracted from original RGB after YCbCr conversion		...	...	...	...	...	...
Cb (Blue Chroma) Prediction Blue-difference chroma predicted from L2 parent (reused in reconstruction)		N/A	N/A	N/A			
Cr (Red Chroma) Prediction Red-difference chroma predicted from L2 parent (reused in reconstruction)		N/A	N/A	N/A			
Y Residual Raw Difference between original and predicted luma: R = Y - Y_pred		N/A	N/A	N/A			
Encoded Residual Quantized residual centered to [0,255] and JPEG compressed		N/A	N/A	N/A			
Reconstructed RGB Final reconstructed tile after decompression: Y_recon + predicted chroma -> RGB							
Tile Size ↓ Per-tile encoded size (baseline: tile file, ORIGAMI: residual)	29.5 KB (source ÷ 20)	5.02 KB	8.38 KB	18.48 KB	**2.66** KB (residual)	4.80 KB (residual)	11.69 KB (residual)
Tile Ratio ↑ (source ÷ 20) ÷ tile size	1:1	5.9:1	3.5:1	1.6:1	**11.1:1**	6.2:1	2.5:1
Family Size ↓ L0 (16) + L1 (4) = 20 tiles, LZ4 packed or summed	590.2 KB (1024px source)	98.5 KB	167.7 KB	380.7 KB	**46.2** KB	91.1 KB	241.3 KB
Family Ratio ↑ Original family size ÷ encoded family size	1:1	6.0:1	3.5:1	1.6:1	**12.8:1**	6.5:1	2.4:1
Avg PSNR ↑ Peak Signal-to-Noise Ratio (dB) — averaged across 20 L0+L1 tiles	∞ dB	32.26 dB	34.87 dB	39.38 dB	34.38 dB	36.28 dB	**39.72** dB
Avg SSIM ↑ Structural Similarity Index (0-1) — averaged across 20 L0+L1 tiles	1.0000	0.8803	0.9251	0.9703	0.9146	0.9462	**0.9810**
Avg MSE ↓ Mean Squared Error — averaged across 20 L0+L1 tiles	0.0	40.2	22.3	8.0	24.1	15.6	**7.0**
Avg VIF ↑ Visual Information Fidelity (0-1) — averaged across 20 L0+L1 tiles	1.0000	0.4696	0.5698	0.7329	0.5498	0.6258	**0.7515**
Avg Delta E ↓ CIE Delta E color difference — averaged across 20 L0+L1 tiles	0.00	3.88	2.87	1.80	2.13	1.90	**1.60**
Avg LPIPS ↓ Learned Perceptual Image Patch Similarity (0=identical) — averaged across 20 L0+L1 tiles	0.0000	0.1240	0.0518	**0.0079**	0.1218	0.0461	0.0094

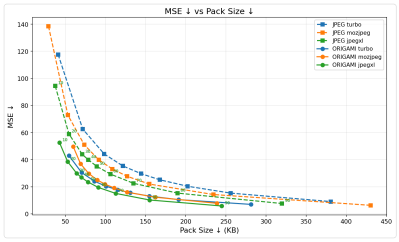
← Previous Tile | Next Tile →

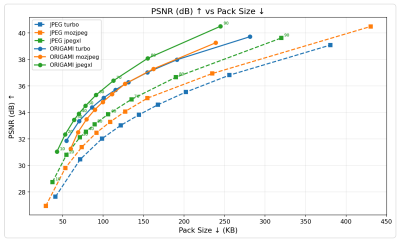
Metric Charts

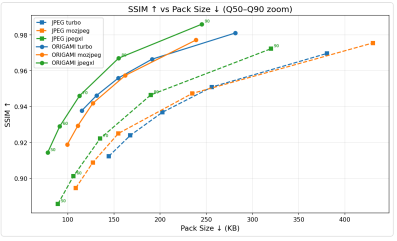
All (18) | Metric vs Quality (8) | Size vs Quality (2) | Metric vs Size (2) | Metric vs Pack Size (6) | PSNR (3) | SSIM (4) | VIF (2) | Delta E (2) | MSE (2) | LPIPS (3)

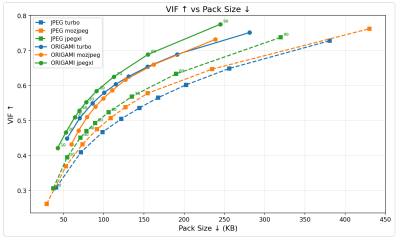












On Image Quality Metrics in Digital Pathology

No validated clinical thresholds exist for PSNR, SSIM, LPIPS, or other full-reference image quality metrics in digital pathology. The literature explicitly warns against treating these metrics as reliable proxies for diagnostic quality. The values shown above are useful for *relative comparison* between encoders and quality levels, but should not be interpreted as pass/fail criteria.

No Reliable Metric Predicts Diagnostic Impact

> *"There is no metric available that can reliably predict human judgments of image quality in compressed images."*
> — Fischer et al. 2024, WSI Compression Baselines (PMC)

Compression Tolerance Is Higher Than Metrics Suggest

The only study to measure actual segmentation and pathologist performance across compression levels found that 85% compression preserved 95% of deep learning performance, and that the practical breakdown point aligned with where pathologists themselves reported difficulty — not where metrics like PSNR or SSIM first begin to degrade.

> *"DP images can be compressed by 85% while still maintaining the performance of the DL algorithms at 95% of what is achievable without any compression."*
> — Chen et al. 2020, JCO Clinical Cancer Informatics

> *"The maximum compression level sustainable by DL algorithms is similar to where pathologists also reported difficulties in providing accurate interpretations."*
> — Chen et al. 2020, JCO Clinical Cancer Informatics

LPIPS Was Not Designed for Medical Imaging

> *"[LPIPS] has not been rigorously tested nor developed for medical images."*
> — Kastryulin et al. 2025, IQA Reassessment

Metrics Systematically Fail in Medical Contexts

> *"Currently and to the best of our knowledge, there is no publicly available database with full-reference ratings for medical images."*
> — Kastryulin et al. 2025, IQA Reassessment

> *"Discrepancies in medical scenarios ... might imply wrong judgement of novel methods for medical images."*
> — Kastryulin et al. 2025, IQA Reassessment

Codec-Specific Distortion Patterns Confound Metrics

Different codecs (JPEG, JPEG XL, WebP, learned methods) produce structurally different distortion patterns. Pixel-level metrics like PSNR/SSIM and even learned metrics like LPIPS may penalize one codec's artifacts more harshly than another's at equivalent perceptual quality. JPEG XL in particular uses adaptive perceptual quantization that can produce lower metric scores while remaining visually indistinguishable at moderate compression ratios.

References

[1] Chen et al. "Assessment of Computational Pathology Deep Learning Algorithms' Robustness to Image Compression." JCO Clinical Cancer Informatics, 2020. PMC • Measures nuclei segmentation F-score and pathologist comfort across JPEG/JPEG2000 compression levels on uncompressed WSIs. Found 85% compression preserved 95% of DL performance. Recommends caution below PSNR 40 dB for new applications.

[2] Fischer et al. "Enhanced Diagnostic Fidelity in Pathology Whole Slide Image Compression via Deep Learning." MLMI/MICCAI 2023. Springer • Tests JPEG, WebP, JPEG-XL, and deep learning codecs on pathology WSIs. Provides rate-distortion curves but establishes no diagnostic quality cutoffs.

[3] Fischer et al. "WSI Compression Baselines." PMC 2024. PMC • Evaluates JPEG, JPEG-XL, WebP, and learned codecs against uncompressed originals. Proposes foundation-model feature similarity as an alternative to pixel-level metrics. No thresholds recommended.

[4] Kastryulin et al. "Image Quality Assessment for Medical Imaging: A Reassessment." 2025. arXiv • Comprehensive audit of PSNR, SSIM, and LPIPS across six medical imaging modalities. Finds systematic failures in all three metrics and recommends caution in their use for medical image evaluation.