White Paper: A Practical Guide to Selecting Compression Levels for 3D Light Sheet Fluorescent Microscopy Data

Author: Nathan J. O'Connor PhD MS

Date: July 23, 2025

1. Executive Summary

Light sheet fluorescence microscopy (LSFM) is a powerful technique for high-resolution, three-dimensional imaging of biological specimens over time. However, LSFM generates massive datasets, often on the scale of terabytes per experiment, posing significant challenges for data storage, transfer, and analysis. While lossless compression offers a partial solution, certain lossy compression methods, such as wavelet-based methods, can provide much greater data reduction. This white paper provides a framework for selecting an appropriate lossy wavelet-based JPEG 2000 compression level for 3D LSFM data, balancing the need for data reduction with the preservation of scientific integrity. We present a methodology for evaluating the impact of different compression ratios on image quality and downstream analysis, enabling researchers to make informed decisions for their specific research questions and data types. We used this methodology to analyze the images from the SLICE™ light sheet microscopy system from MBF Bioscience. The methodology presented here may also be used to analyze other microscopes. However, the results presented in this paper are for SLICE.

Key Findings

For SLICE microscopy data, we identified an optimal compression range of **15:1 to 20:1** using JPEG 2000 wavelet-based compression. This "sweet spot" represents the highest compression ratios that maintain scientific integrity while delivering substantial storage savings.

This recommendation is supported by both human visual assessment and quantitative analysis. Expert observers confirmed that compression artifacts remain imperceptible within this range, while automated quantitative measurements showed no significant impact on downstream analysis results.

Practical Impact: A conservative 15:1 compression ratio reduces file sizes by **93%**—transforming a 500 GB dataset into just 33 GB while preserving scientific validity. This dramatic reduction addresses the storage and transfer challenges that currently limit light sheet microscopy workflows, making high-throughput imaging more practical and cost-effective.

2. The Data Deluge in Light sheet Microscopy

The advantages of LSFM—high speed, low phototoxicity, and deep tissue penetration—have led to its widespread adoption in developmental biology, neuroscience, and other fields. A single LSFM experiment can generate multiple 3D image stacks, resulting in datasets that quickly overwhelm local storage and challenge institutional data management infrastructure.

Key Data Challenges:

- **Storage Costs:** The sheer volume of raw data incurs significant and ongoing storage costs.
- **Data Transfer:** Moving terabyte-scale datasets between acquisition systems, processing workstations, and long-term archives is slow and resource-intensive.
- **Analysis Bottlenecks:** Large file sizes can hinder the performance of image analysis software, slowing down quantitative measurements and discovery.

3. JPEG 2000: A Primer for Scientific Imaging

JPEG 2000 is a wavelet-based image compression standard[1] that offers several advantages over the traditional JPEG format, making it particularly well-suited for scientific data:

- **Superior Compression Performance:** At high compression ratios, JPEG 2000 typically produces fewer visual artifacts than JPEG.
- **Lossless and Lossy Compression:** The standard supports both lossless and lossy compression, providing flexibility for different use cases.
- **Scalability:** JPEG 2000 allows for the decoding of images at different resolutions and quality levels from a single compressed file.
- **Region of Interest (ROI) Coding:** Specific regions of an image can be compressed at a higher quality than the background, preserving important features.

4. Methodology for Evaluating Compression Levels

[1] The described compression ratio values refer to settings which are available in MBF Bioscience applications, e.g., BrightSLICE. These ratios correspond to a bitrate setting which is the ratio between the total number of compressed bits and the product of the largest horizontal and vertical image component dimensions. The optimal rate setting is determined for the image being compressed, along with numerous other settings such as progression ordering, wavelet decomposition levels, quality layers, etc.

The central question for any researcher considering lossy compression is: "How much compression can I apply before it negatively impacts my scientific conclusions?" The answer to this question is application dependent. We propose a systematic approach to determine the optimal compression level for your data.

4.1. Image Quality Metrics

A combination of qualitative and quantitative metrics should be used to assess the impact of compression:

- **Qualitative Assessment (Visual Inspection):**
  - **Side-by-Side Comparisons:** Visually compare the original, uncompressed data with data compressed at various ratios (see **Appendix C: Sample compressed images from SLICE stacks** for example images).
  - **Expert Evaluation:** Have researchers familiar with the biological structures in the images assess the impact of compression on their ability to identify key features.
- **Quantitative Assessment (Objective Metrics):**
  - **Peak Signal-to-Noise Ratio (PSNR):** A classic metric that measures the ratio between the maximum possible power of a signal and the power of corrupting noise. Higher PSNR generally indicates better quality.
  - **Structural Similarity Index (SSIM):** A perceptually based metric that considers changes in structural information, luminance, and contrast. A value closer to 1 indicates higher similarity (Wang et al., 2004).
  - **Visual Information Fidelity (VIF):** An advanced metric based on information theory. It quantifies the amount of visual information lost between the original and compressed image from the perspective of the human visual system. A value closer to 1 indicates less information has been lost (Sheikh & Bovik, 2006).
  - **Learned Perceptual Image Patch Similarity (LPIPS):** A state-of-the-art metric that uses a deep neural network to compare images in a way that aligns with human perception. Lower scores indicate greater similarity (Zhang et al., 2018). For scientific images, the choice of the underlying network is important. While the default AlexNet is fast, the deeper and more uniform VGG network is often more sensitive to the subtle textural changes and fine details that are critical in microscopy data.

- **Application-Specific Metrics:** The most critical evaluation is to measure the impact of compression on the specific quantitative analysis being performed (e.g., cell counting, segmentation, intensity measurements).

4.2. Recommended Experimental Protocol

1. **Select Representative Datasets:** Choose a few representative datasets that encompass the range of image quality and biological structures present in your typical experiments.
2. **Define a Range of Compression Ratios:** Start with a range of compression ratios (e.g., 5:1, 10:1, 20:1, 50:1, 100:1).
3. **Compress the Data:** Use a reliable JPEG 2000 implementation to compress the data at each chosen ratio.
4. **Perform Qualitative and Quantitative Analysis:** Apply the metrics described in section 4.1 to the compressed datasets.
5. **Analyze the Results:** Plot the quantitative metrics as a function of the compression ratio to identify the point of diminishing returns.
6. **Determine the Optimal Compression Level:** Based on the analysis, select a compression level that provides a significant reduction in file size without unacceptably compromising the scientific integrity of the data.

5. Results: Evaluating Compression Impact on SLICE Images of Mouse Brains

This section presents the findings from applying the evaluation methodology to a representative LSFM dataset of cleared mouse brain regions images by the SLICE microscope from MBF Bioscience. The impact of compression was evaluated across four conditions: using both the AlexNet and VGG deep learning backbones on images with either 2x2 or 3x3 pixel binning.

5.1. Qualitative Observations

Visual inspection by trained researchers revealed that compression artifacts became noticeable at ratios above 20:1, manifesting as a subtle softening of fine cellular processes. However, despite visible compression artifacts that began to occur at a 25:1 compression ratio, observers noted that the visualization of biological structures (here cells) was not adversely impacted. This qualitative assessment agrees well with the statistically similar results of automated cell detection across compress levels described in section 5.2.2 Application-Specific Analysis.

5.2. Quantitative Analysis

5.2.1 Image Fidelity Analysis

The impact of compression on image fidelity was assessed using two categories of metrics. Figure 1 shows the results for classical, error-based metrics (PSNR and MSE), while Figure 2 shows the results for modern, perception-based metrics (VIF and LPIPS).

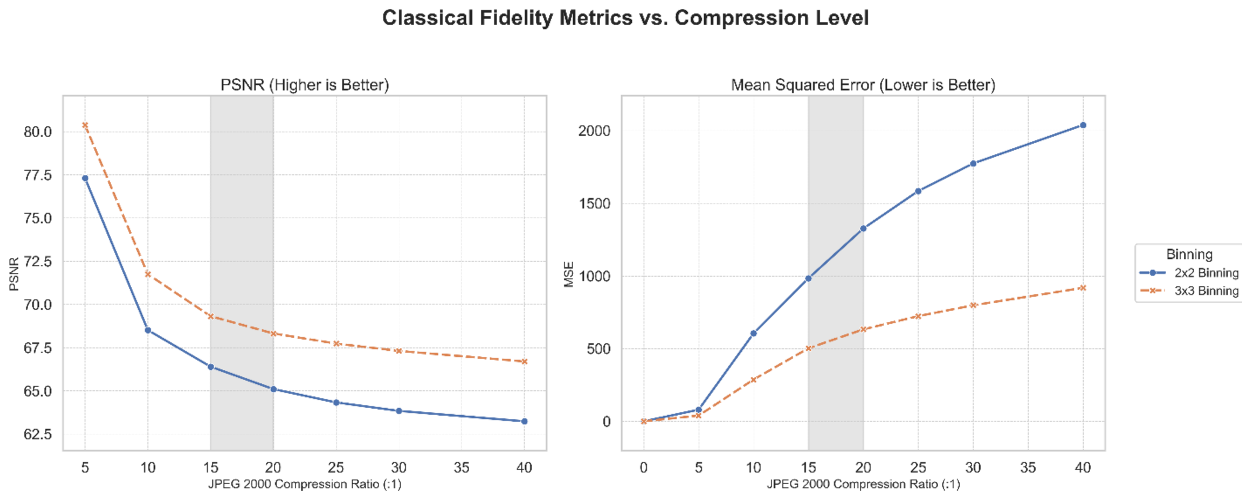


Figure 1: Classical Fidelity Metrics. Plots of PSNR and MSE versus compression ratio for 2x2 and 3x3 binned images. Both metrics show a clear, expected degradation in quality as compression increases.

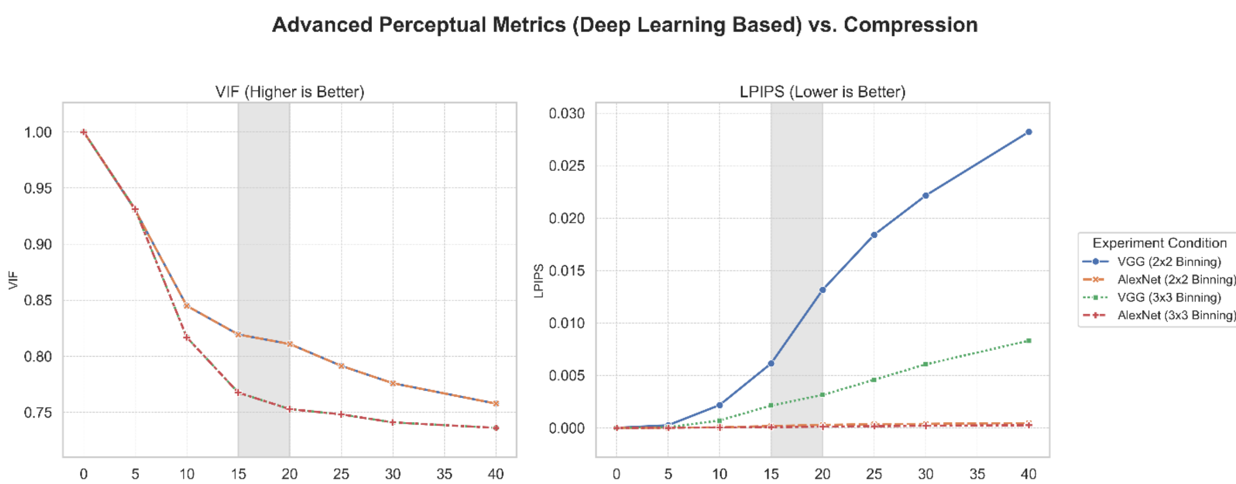


Figure 2: Advanced Perceptual Metrics. Plots of VIF and LPIPS versus compression ratio across all four experimental conditions. These metrics reveal more subtle differences between the LPIPS network backbones and binning levels.

In Figures 1 and 2 data points for uncompressed and lossless compressed image data are identical. This is because the image data from lossless jpeg2000 compression is truly lossless. This was validated with a direct comparison of every pixel pair between uncompressed and lossless compressed image stacks. Correspondingly, the cell detection results described in 5.2.2. Application-Specific Analysis are identical for uncompressed and lossless compressed data. Additionally, lossless JPEG 2000 compression consistently achieved a compression ratio of approximately 2:1 for all 2x2 and 3x3 camera binned images from SLICE.

5.2.2 Application-Specific Analysis

To directly measure the impact of compression on scientific outcomes, an automated cell counting algorithm was run on all SLICE image stacks. For this validation, a traditional Laplacian of Gaussian (LoG) based method, and a more modern neural network (NN) based approach were chosen. The LoG method, being a classic blob detector, is highly sensitive to changes in pixel intensity and edge sharpness. This makes it a more rigorous "critic" for detecting subtle compression artifacts. A NN-based detector while potentially "too smart" (i.e, it is trained to be robust to noise and minor variations and could mask the true point at which data integrity begins to degrade) was nonetheless utilized for comparison to the LoG method. Figure 3 plots the resulting cell counts, normalized to the count from the original uncompressed image, against the compression ratio.

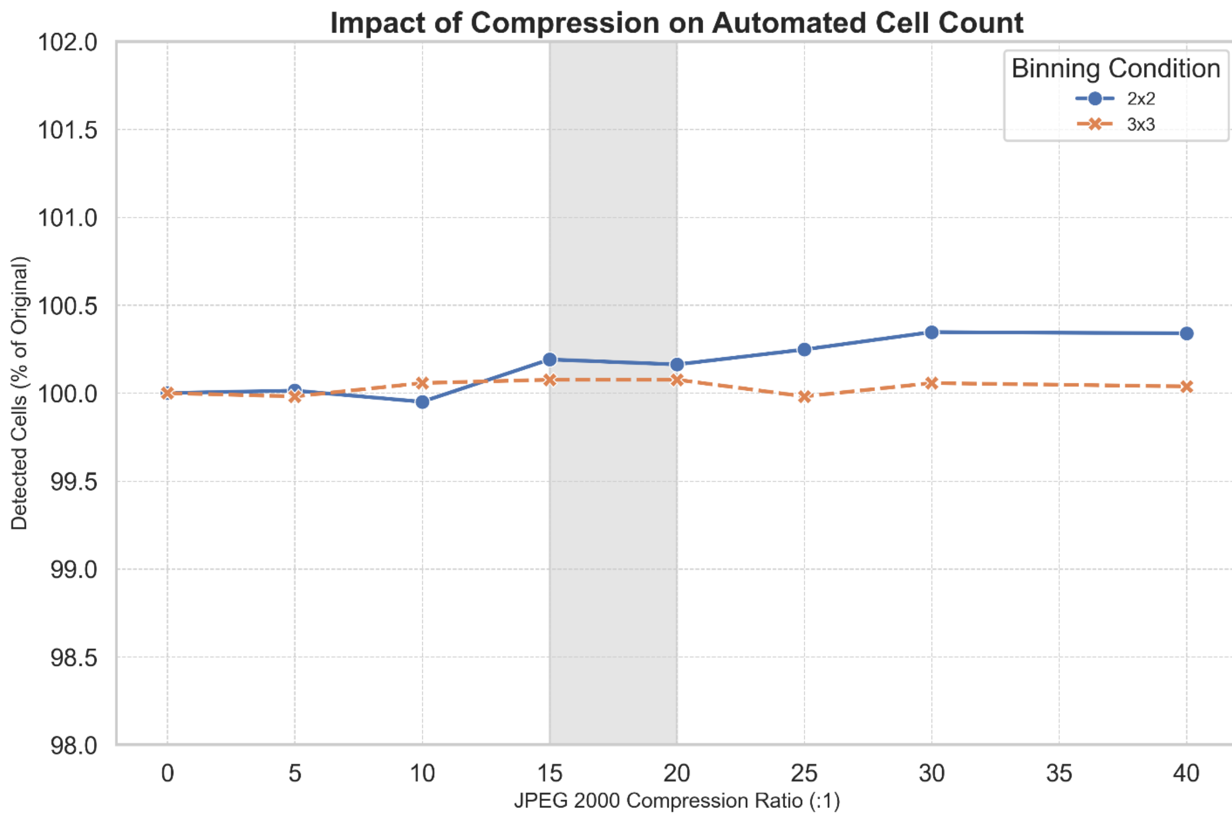


Figure 3: Impact of Compression on Cell Count. The plot shows the percentage of cells detected in compressed images relative to the original. This metric directly assesses whether compression artifacts are significant enough to alter quantitative biological results.

5.3. Conclusion of Findings

The quantitative analysis reveals several key insights. As shown in Figure 1, classical metrics like PSNR and MSE effectively demonstrate that image fidelity decreases with higher compression. However, the advanced perceptual metrics in Figure 2 provide a more nuanced understanding. The LPIPS plot shows that the deeper VGG network consistently registers a larger perceptual error than AlexNet across all compression levels. This suggests that **VGG is a more sensitive detector of the subtle textural artifacts** introduced by compression, making it a more rigorous metric for scientific images.

Ultimately, the impact on the experimental endpoint should be a guiding force in the selection of a compression level. The application-specific analysis in Figure 3, supported by the statistical analysis in Appendix B, provides the ultimate ground truth. It shows that for these data, automated cell detection numbers were largely unimpacted and statistically consistent even at high compression levels. This suggests that for workflows that rely solely on this specific automated analysis, and not visual presentation (e.g., in a figure), higher compression ratios could be tolerated, leading to significant savings in data storage and budget.

Based on these findings, we identified a "sweet spot" for compression between **15:1 and 20:1**. This range, highlighted in the figures, was determined by human observers to be the highest acceptable range for these specific datasets before noticeable artifacts appeared. This recommendation balances the robustness of the automated analysis with the preservation of visual fidelity for human review and discovery.

6. Recommendations and Best Practices

- **Start Conservatively:** If you are new to lossy compression, begin with lower compression ratios (e.g., 5:1 to 15:1).
- **Archive Raw Data (If Desired):** For critical experiments, archive the original raw data with lossless compression if resources permit.
- **Document Your Workflow:** Record the compression software, version, and parameters used for each dataset. This is crucial for reproducibility.
- **Educate Your Team:** Ensure that all members of your research group understand the implications of working with lossy compressed data.
- **Re-evaluate for New Experiments:** The optimal compression level may vary for different sample types, imaging conditions, and analysis goals.

7. Conclusion

Lossy wavelet-based compression with JPEG 2000 offers an excellent solution to the data storage and management challenges posed by modern light sheet microscopy. By adopting a systematic evaluation process that includes advanced perceptual metrics and, most importantly, application-specific validation, researchers can confidently choose a compression level that significantly reduces data volumes without sacrificing the scientific validity of their results. The final choice must be guided by the experimental endpoint; if reproducible quantitative results are achievable at high compression levels, substantial savings in storage costs and budget can be realized, making research more sustainable.

8. References

1. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612.

2. Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on image processing*, 15(2), 430-444.

3. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 586-595).

Appendix A: Supplemental Fidelity Data Tables

Table A1: Classical Fidelity Metrics

Binning	**Compression Level (N:1)**	**PSNR**	**SSIM**
2x2	**Uncompressed**	inf	1.0
	Lossless	inf	1.0
	5	80.38	0.999988
	10	68.51	0.999839
	15	66.40	0.999738
	20	65.10	0.999647
	25	64.33	0.999579
	30	67.31	0.999787
3x3	**Uncompressed**	inf	1.0
	Lossless	inf	1.0
	5	80.38	0.999988

	10	71.75	0.999923
	15	68.32	0.999831
	20	67.73	0.999806
	25	67.31	0.999787
	30	67.31	0.999787

Table A2: Perceptual Fidelity Metrics

Compression Level (N:1)	**VIF (2x2)**	**LPIPS (AlexNet, 2x2)**	**LPIPS (VGG, 2x2)**	**VIF (3x3)**	**LPIPS (AlexNet, 3x3)**	**LPIPS (VGG, 3x3)**
Uncompressed	1.0	0.0	0.0	1.0	0.0	0.0
Lossless	1.0	0.0	0.0	1.0	0.0	0.0
5	0.8879	0.000021	0.0002	0.8879	0.000021	0.0002
10	0.8449	0.000046	0.0022	0.8169	0.000024	0.0007
15	0.8193	0.000170	0.0053	0.7528	0.000103	0.0031
20	0.8108	0.000277	0.0132	0.7481	0.000152	0.0046

25	0.7914	0.000388	0.0184	0.7410	0.000191	0.0069
30	0.7410	0.000191	0.0069	0.7410	0.000191	0.0069

Appendix B: Supplemental Cell Count Analysis

LoG Detector Parameters

The following parameters were used for the Laplacian of Gaussian (LoG) based cell detection across all datasets.

Parameter	**2x2 Binning Value**	**3x3 Binning Value**
LoG Strength Threshold	150	150
Minimum Object Size	40	10
Maximum Object Size	10	40

Statistical Analysis of Cell Counts

To verify the stability of the cell counts across compression levels, Z-score and Interquartile Range (IQR) outlier analyses were performed on the LoG cell counts for each binning condition.

Table B1: Statistical Analysis for 2x2 Binned Data

- **Mean:** 14143.7
- **Standard Deviation:** 21.3
- **IQR Lower Fence:** 14061.75
- **IQR Upper Fence:** 14227.75

Compression Level (N:1)	Cell Count	Z-score	Outlier?
Uncompressed	14124	-0.92	No
Lossless	14124	-0.92	No
5	14126	-0.83	No
10	14117	-1.25	No
15	14151	0.34	No
20	14147	0.15	No
25	14159	0.72	No
30	14173	1.37	No
35	14172	1.33	No

Table B2: Statistical Analysis for 3x3 Binned Data

- **Mean:** 5270.56
- **Standard Deviation:** 2.16
- **IQR Lower Fence:** 5262.5
- **IQR Upper Fence:** 5278.5

Compression Level (N:1)	Cell Count	Z-score	Outlier?
Uncompressed	5269	-0.72	No

Lossless	5269	-0.72	No
5	5268	-1.18	No
10	5272	0.67	No
15	5273	1.13	No
20	5273	1.13	No
25	5268	-1.18	No
30	5272	0.67	No
35	5271	0.20	No

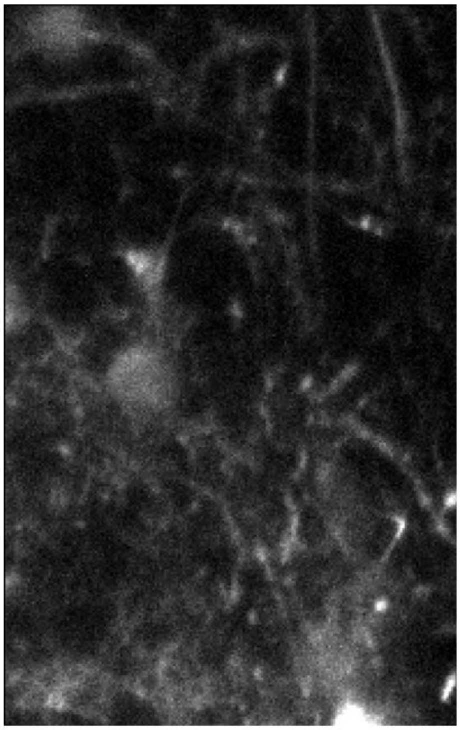
Conclusion: As shown in the tables, no cell counts at any compression level were identified as statistical outliers by either the Z-score or IQR methods for both binning conditions. This provides strong quantitative evidence that the LoG cell count is robust to compression artifacts within the tested range.

Appendix C: Sample compressed images from SLICE stacks

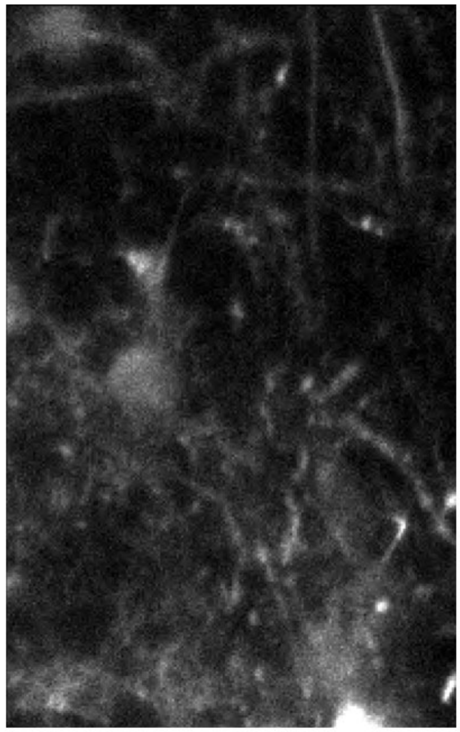
Sample XY optical sections from SLICE image stacks at 2x2 and 3x3 camera binning. For each binning level, both uncompress and compressed (JPEG 2000 15:1) images are displayed.

2x2 Camera binning

1:1

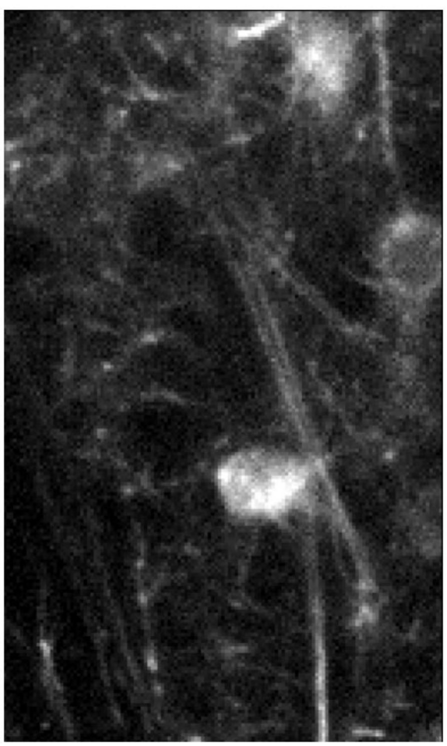


15:1



3x3 Camera binning

1:1



15:1