SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

Tissue-Aware Distributional Matching for JPEG 2000: 1.7× Storage Reduction in Whole-Slide Histopathology at JPEG-80-Equivalent Quality

Ritika Agarwal, MSE, Perelman School of Medicine, University of Pennsylvania; Mathew Francis, PhD; Mark D. Zarella, PhD

Introduction/Background
JPEG 2000 achieves higher compression ratios than JPEG on whole-slide cytology images at equivalent quality measured by mean SSIM. However, this framework has not been applied to histopathology, where tissue morphology and staining characteristics differ from cytology and most equipment defaults to JPEG quality 80. Further, relying on mean SSIM to match codecs may obscure differences in compression behavior when background tiles dominate a slide.

Methods/Intervention
We processed 14 uncompressed H&E histopathology slides (0.23 µm/pixel), tiled into 512×512 patches (166,579 tiles). A pre-trained UNet++ segmented tissue at 10 µm/pixel to give each tile a tissue-area fraction. Tiles were compressed with JPEG (quality 80) and JP2 (qualities 35–40), and SSIM was computed per tile against the uncompressed original. We matched codecs two ways: by mean SSIM and by Wasserstein distance between per-tile SSIM distributions. All analyses were repeated on tissue tiles (fraction ≥ 0.5).

Results/Outcome
Across all tiles, both metrics selected JP2 quality 38 as the closest match to JPEG quality 80 (mean-SSIM difference 0.001; Wasserstein 0.006 vs ≥0.021 at neighboring settings), giving 85× compression for JP2 versus 34× for JPEG. But the all-tile SSIM distribution was strongly right-skewed (skewness 2.3): background accounted for 84% of the total and dominated the mean. Restricting to tissue tiles moved the matched setting to JP2 quality 36, where JP2 achieved 28× compression compared to 17× for JPEG, a 1.7× improvement at equivalent quality. The tissue-only Wasserstein curve varied smoothly with quality, unlike the sharp drop on all tiles.

Conclusion
Matching codecs on tissue-tile distributions, rather than on mean SSIM over whole slides, gives a more reliable comparison, because background tiles otherwise dominate the mean and shift the chosen operating point. At the tissue-matched setting, replacing JPEG quality 80 with JP2 improves compression 1.7× at equivalent SSIM, a substantial storage saving for histopathology archives.

Statement of Impact
Adopting JP2 at the tissue-matched setting could cut histopathology archive storage ~1.7× at image quality equivalent, by SSIM, to the JPEG setting already in routine use, with no change to scanner acquisition.

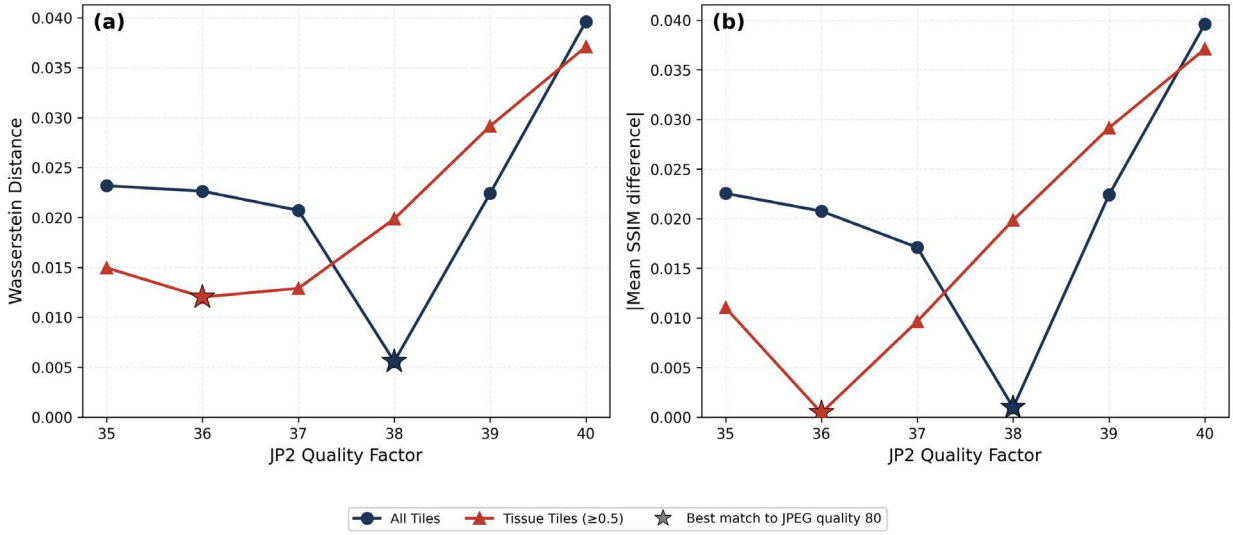

Figure 1. (a) Wasserstein distance and (b) absolute mean SSIM difference between JPEG quality 80 and JPEG 2000 at each quality factor. We computed each metric on all 166,579 tiles (navy circles) and on tissue-only tiles with tissue fraction at least 0.5 (red triangles). Lower values indicate a closer match to JPEG quality 80. Stars mark the best-matching JP2 quality for each curve. On all tiles, both metrics agree on JP2 quality 38; on tissue tiles, both shift to JP2 quality 36.

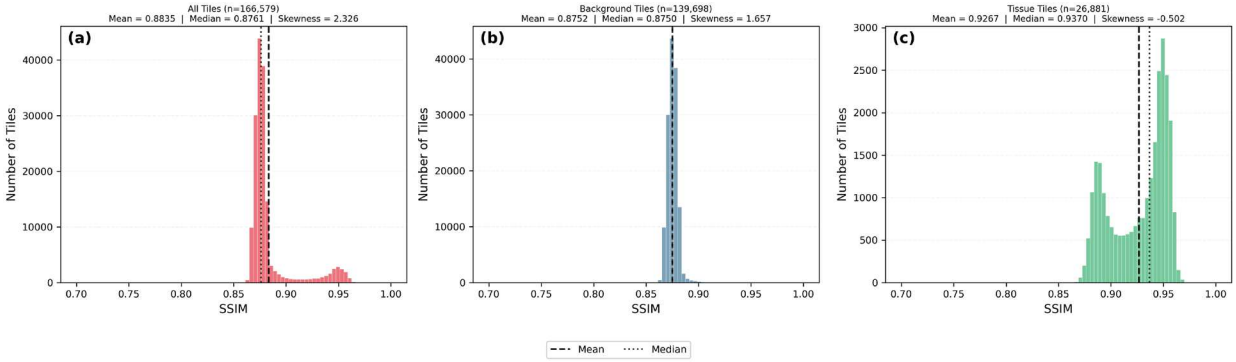

Figure 2. SSIM distribution of JPEG quality 80 compressed tiles for (a) all 166,579 tiles, (b) background tiles with tissue fraction below 0.5 (n = 139,698, 84% of total), and (c) tissue tiles with tissue fraction at least 0.5 (n = 26,881). Dashed and dotted lines mark the mean and median respectively. The all-tile distribution is right-skewed (skewness 2.3) because the narrow background spike at SSIM 0.875 dominates the overall distribution, while tissue tiles form a broader, left-skewed distribution centered near 0.94.

QRFBVSUS-2480909-3-ANY.pdf

Figure 3. Representative 512×512 tissue tiles at the 25th percentile, median, and 75th percentile of the JPEG quality 80 SSIM distribution, showing the original uncompressed tile alongside JPEG quality 80, JP2 quality 36 (tissue-matched setting), and JP2 quality 38 (all-tile-matched setting). All tiles have tissue fraction of 1.0. SSIM and compression ratio are shown for each compressed version.

Keywords

histopathology; whole-slide imaging; image compression; JPEG 2000; SSIM; Wasserstein distance