Good, Cheap, and Fast: Overfitted Image Compression with Wasserstein Distortion

Supplementary Material

6. All reconstructions and study data

The original images and all reconstructions in PNG format are available for download at the following locations.

method	file size	URL
original images	129 MB	https://storage.googleapis.com/wasserstein_c3/original.zip
C3/MSE [22]	216 MB	https://storage.googleapis.com/wasserstein_c3/c3-mse.zip
C3/wMSE	209 MB	https://storage.googleapis.com/wasserstein_c3/c3-wmse.zip
C3/MS-SSIM	236 MB	https://storage.googleapis.com/wasserstein_c3/c3-ms-ssim.zip
C3/LPIPS	403 MB	https://storage.googleapis.com/wasserstein_c3/c3-lpips.zip
C3/WD8	439 MB	https://storage.googleapis.com/wasserstein_c3/c3-wd8.zip
C3/WDs	437 MB	https://storage.googleapis.com/wasserstein_c3/c3-wds.zip
VVC [1]	379 MB	https://storage.googleapis.com/wasserstein_c3/vvc.zip
MLIC+ [21]	231 MB	https://storage.googleapis.com/wasserstein_c3/mlicplus.zip
CDC [41]	403 MB	https://storage.googleapis.com/wasserstein_c3/cdc.zip
HiFiC [31]	383 MB	<pre>https://storage.googleapis.com/wasserstein_c3/hific.zip</pre>

The evaluation data (rater responses, metric evaluations) is available at https://storage.googleapis.com/
wasserstein_c3/eval.zip.

7. Additional examples

Please see the next pages for further selected examples.



Figure 8. Horizontal crop of a image 6 from the CLIC2020 professional dataset (best viewed on screen). Left: C3 optimized for MSE, compressed to 0.321 bits/pixel. Right: C3 optimized for WD with $\sigma = 8$, compressed to 0.270 bits/pixel. While optimization for MSE leads to flattened texture and staircasing artifacts, as seen in the reproduction of the vegetation, texture is vastly improved in the WD-optimized version, while using 15% fewer bits.



Figure 9. Horizontal crop of image 19 from the CLIC2020 professional dataset (best viewed on screen). Far left: Original image. Center left: C3 optimized for MSE, compressed to 0.076 bits/pixel. Canter right: C3 optimized for MS-SSIM, compressed to 0.088 bits/pixel. Far right: C3 optimized for WD with $\sigma = 8$, compressed to 0.076 bits/pixel. Under traditional perceptual metrics such as MSE and MS-SSIM, the visual quality of textures such as the starry night sky suffers, even though MS-SSIM is designed to account for visual contrast masking. Neither of them benefit from common randomness provided by the method. WD accounts much better for perception of visual texture and makes use of CR to achieve a more accurate visual impression.



Figure 10. Horizontal crop of image 16 from the CLIC2020 professional dataset (best viewed on screen). Far left: Original image. Center left: C3 optimized for MSE, compressed to 0.088 bits/pixel. Center right: C3 optimized for LPIPS, compressed to 0.076 bits/pixel. Far right: C3 optimized for WD with $\sigma = 8$, compressed to 0.077 bits/pixel. The background texture of the waterfall is largely flattened in the MSE version. The LPIPS version slightly improves on this, albeit at the cost of losing significant detail on the person standing in the foreground. Remarkably, even without employing a saliency model, WD preserves detail in the foreground better than LPIPS, while also reproducing the waterfall texture faithfully.



Figure 11. Horizontal crop of image 11 from the CLIC2020 professional dataset (best viewed on screen). Left: C3 optimized for MSE, compressed to 0.133 bits/pixel. Center: C3 optimized for WD with $\sigma = 8$, compressed to 0.134 bits/pixel. Right: C3 optimized for WD with σ derived from a saliency map, compressed to 0.130 bits/pixel. Again, optimization for MSE leads to flattened texture, as seen in the reproduction of the street and chimneys, while the signage is reproduced well due to its high contrast (difference between dark and light), which MSE is sensitive to. While texture is vastly improved in the WD-optimized version, the flat σ version struggles to reproduce the signage well. On the right, the saliency model assigns high saliency to that image region, leading to a better reconstruction and higher legibility of the text.



Figure 12. Horizontal crop of image 13 from the CLIC2020 professional dataset (best viewed on screen). Far left: Original image. Center left: C3 optimized for LPIPS, compressed to 0.093 bits/pixel. Center right: C3 optimized for WD with $\sigma = 8$, compressed to 0.093 bits/pixel. Far right: C3 optimized for WD with σ derived from a saliency map, compressed to 0.091 bits/pixel. An example of text in the background of the image reproduced as a visual texture. Note that text legibility doesn't improve when comparing the WD version with and without using a saliency model, as in this case, the text is not predicted to be a salient image region. However, the visual quality in both cases is still much preferable to the one achieved by optimizing for LPIPS.