Learned Image Compression with Dictionary-based Entropy Model

Supplementary Material

A. More detailed network architecture

The overall framework of the proposed network. The encoder-decoder and hyperprior module are composed of Transformer blocks, downsample modules, upsample modules, and convolutions. We set the number of Transformer layers to 1 in the hyperprior module and we use the Factorized Entropy Model to estimate the distribution of side information z.

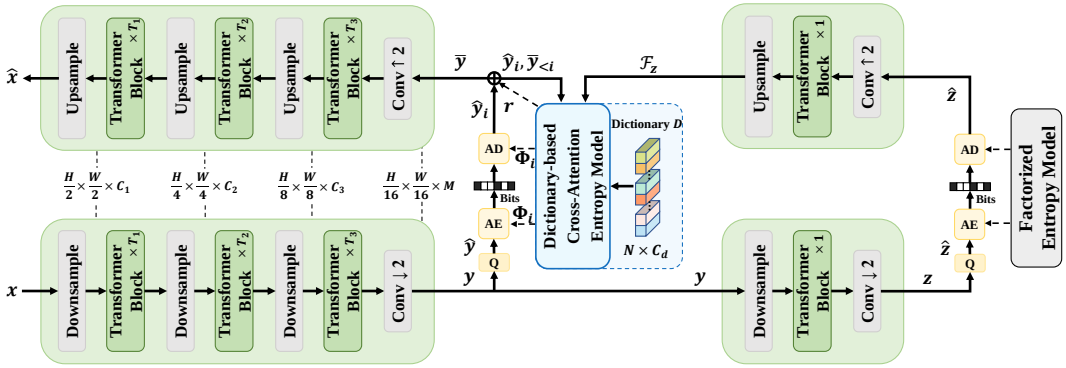

Figure 1. The overall framework of the proposed network.

B. More ablation studies

To further demonstrate the effectiveness of our entropy model, we replace the entropy models in three state-of-the-art methods, ELIC (CVPR'22), TCM (CVPR'23), and FTIC (ICLR'24), with ours. All studies are trained with an initial learning rate of $1e-4$ for 20 epochs, followed by 5 epochs at $1e-4$, using a batch size of 8. As shown in Tab. 1, ELIC's BD-rate improves by 3.75% (-2.62% to -6.37%), TCM by 2.41% (-7.80% to -10.21%), and FTIC by 3.19% (-6.52% to -9.71%). In addition, our entropy model can enhance the existing methods in an plug-and-play manner. Furthermore, combining our model with ELIC's further improves BD-rate from -2.62% to -9.66% (ELIC uses a spatial-channel autoregressive model, while ours is channel-wise only).

Table 1. Comparative evaluation by replacing the entropy model.

Autoencoder	ELIC			TCM		FTIC	
Entropy Model	ELIC	ours	ELIC+ours	TCM	ours	FTIC	ours
BD-rate	-2.62%	-6.37%	-9.66%	-7.80%	-10.21%	-6.52%	-9.71%

C. Multi-Scale visualizations

To study how multi-scale context influences dictionary query, we visualize attention maps at different feature levels from the EConv layers. Specifically, we use three EConv layers for feature extraction and progressively set each layer's output to zero, starting from the last, creating four models: "scale4", "scale3", "scale2", and "scale1". As shown in Fig. 2, with a growing number of EConv layers, the extracted features facilitate progressively more accurate dictionary queries.

D. More visual examples

Fig. 3 and Fig. 4 shows the reconstruction results between our method and State-of-the-Art Methods. Our model achieves better texture detail restoration at similar bpp levels. For example, in Fig. 3, our model more effectively restores the striped

textures on the sails of the boat (kodim09) and the stitching details on the dress (kodim18).

E. More rate-distortion results

For completeness, we present additional methods for comparison (Fig. 5, Fig. 6, and Fig. 7). In addition, we also provide MS-SSIM optimized models (Fig. 8) to compare with other methods. Our model achieves state-of-the-art results on all datasets.

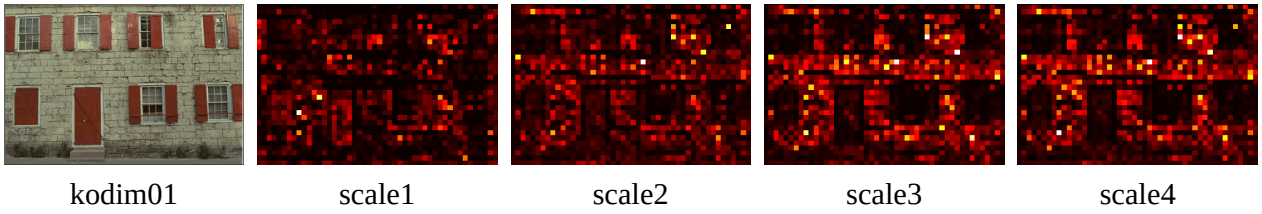

Figure 2. Multi-Scale visualizations

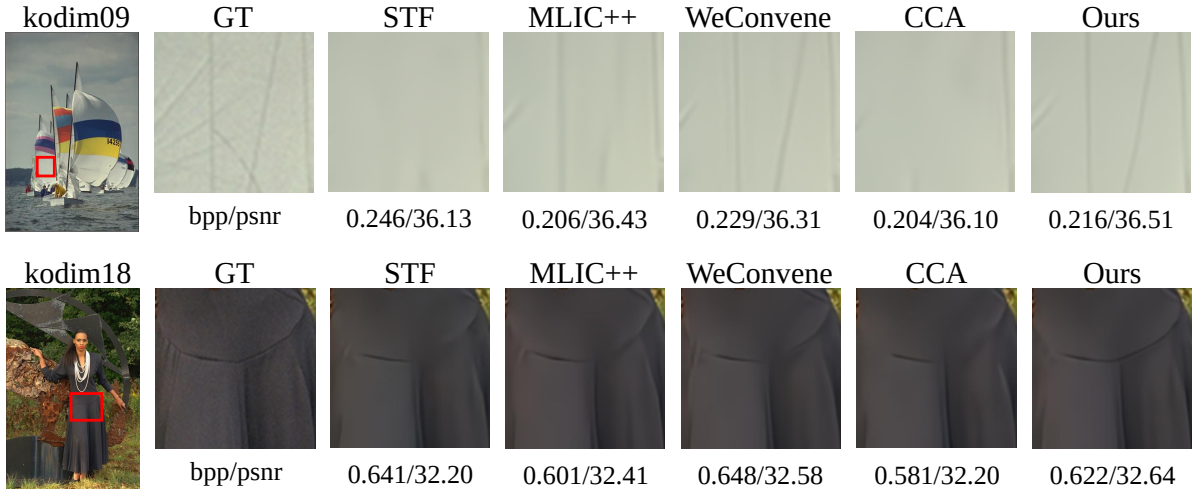

Figure 3. Reconstructed images from the Kodak dataset.

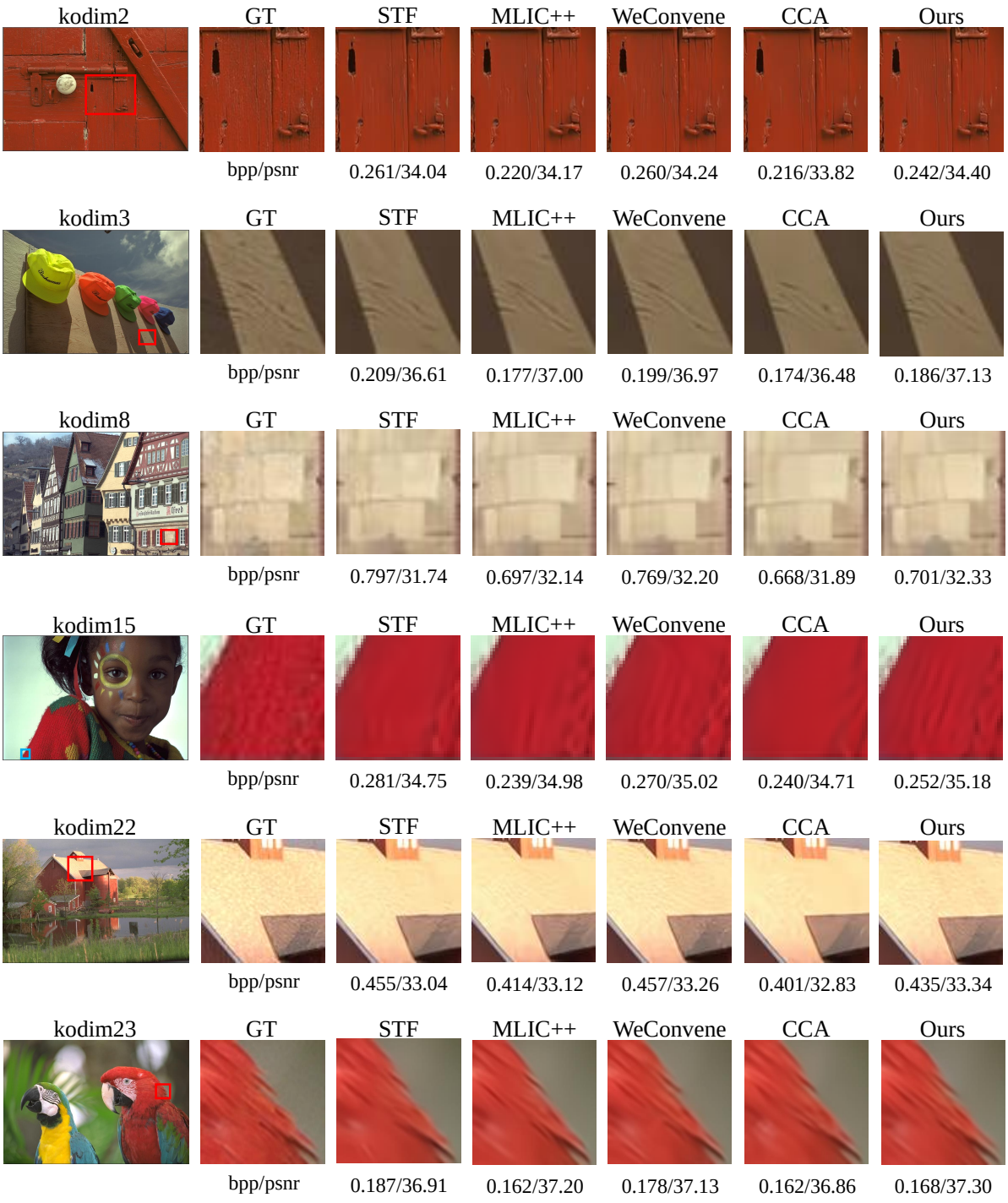

Figure 4. Reconstructed images from the Kodak dataset.

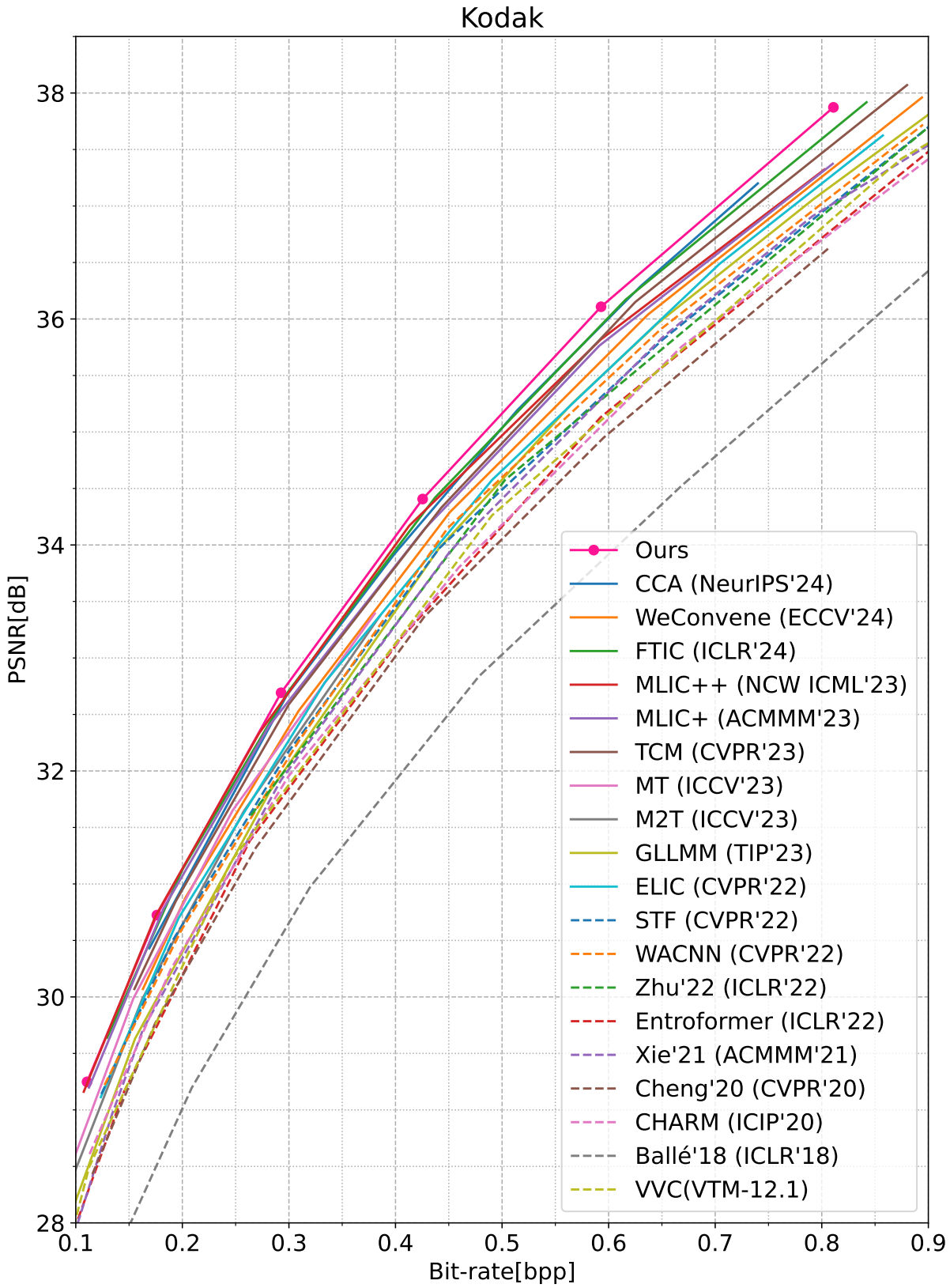

Figure 5. Performance evaluation (PSNR) on the Kodak dataset.

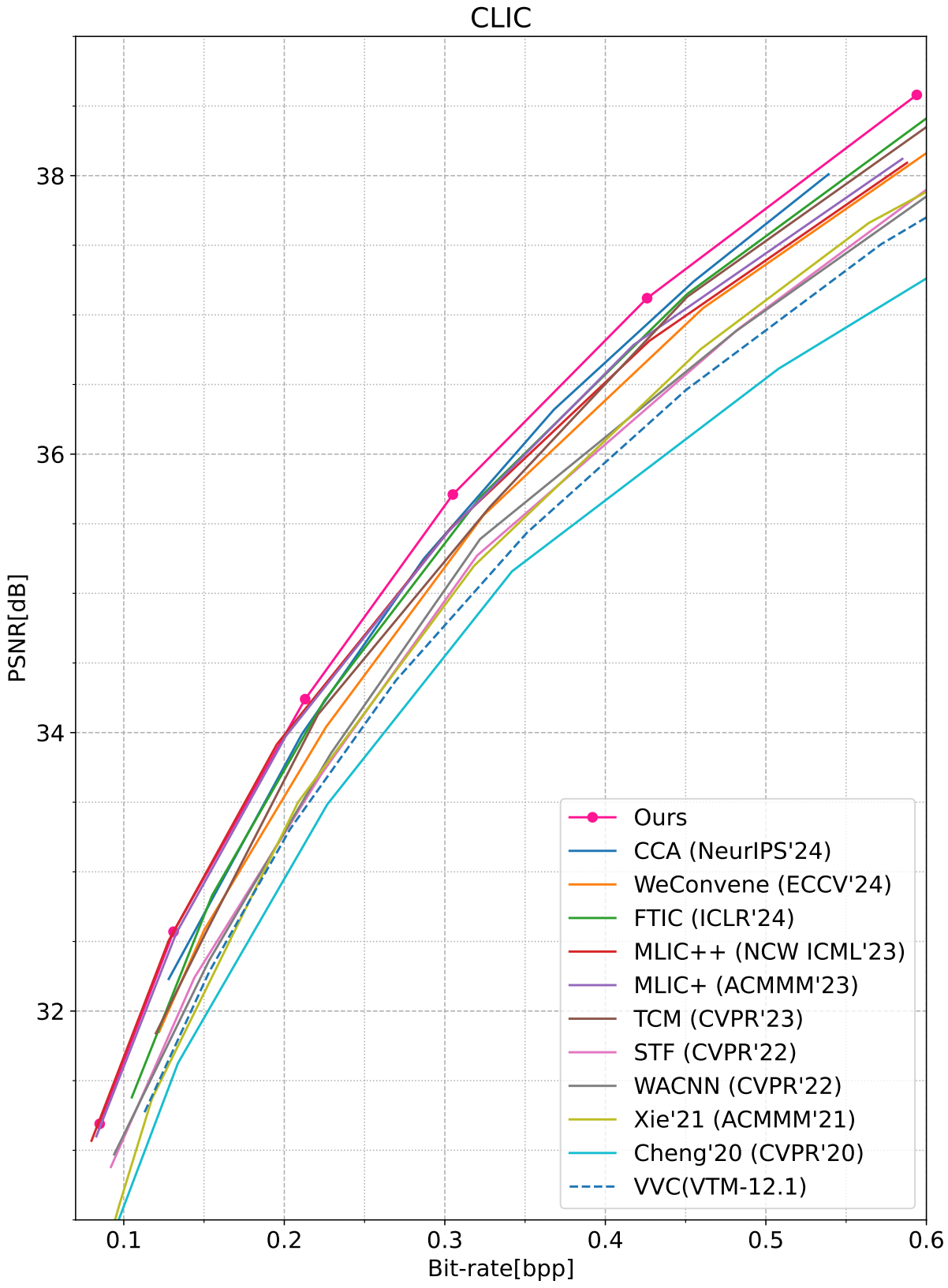

Figure 6. Performance evaluation (PSNR) on the CLIC dataset.

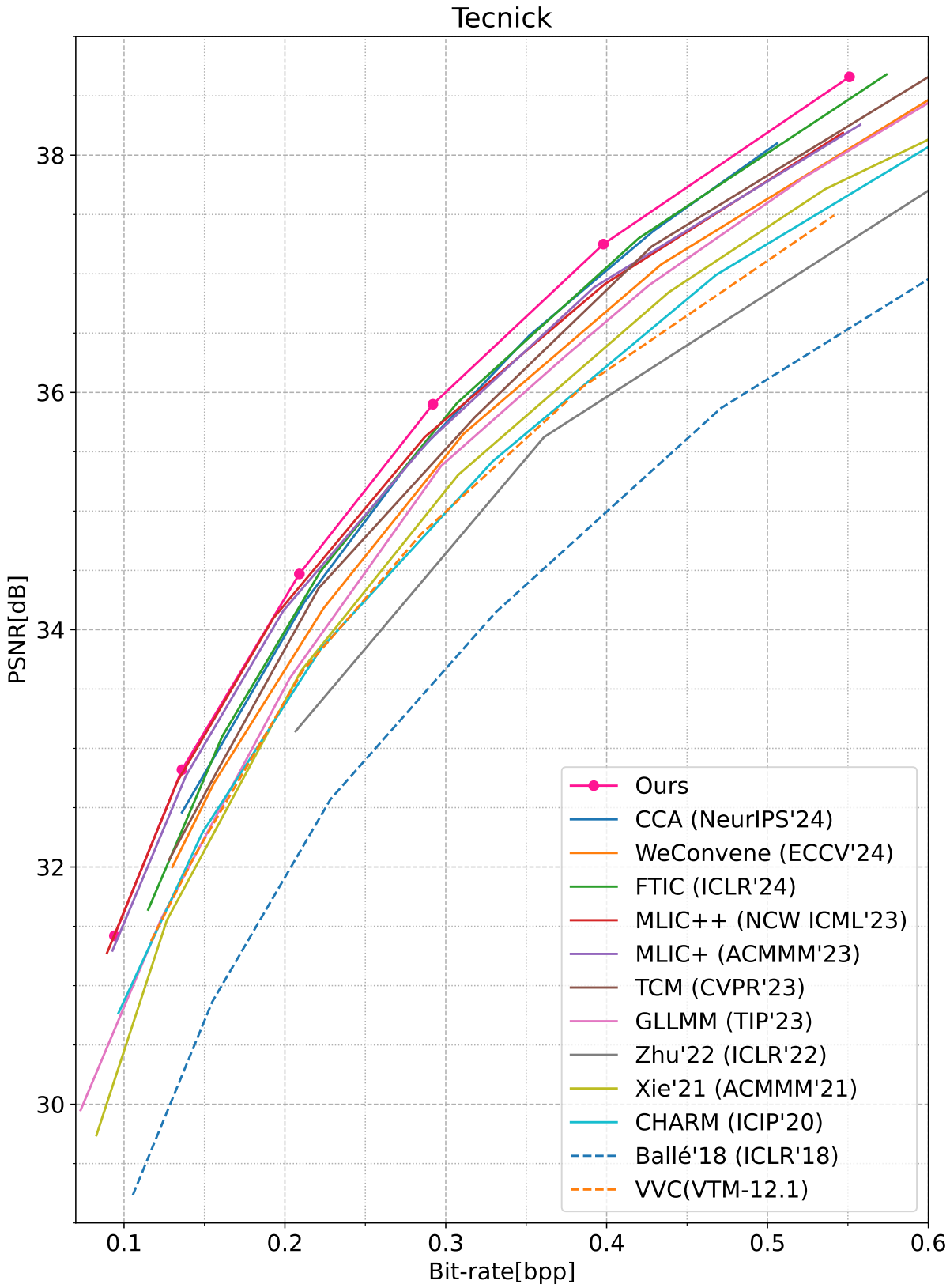

Figure 7. Performance evaluation (PSNR) on the Tecnick dataset.

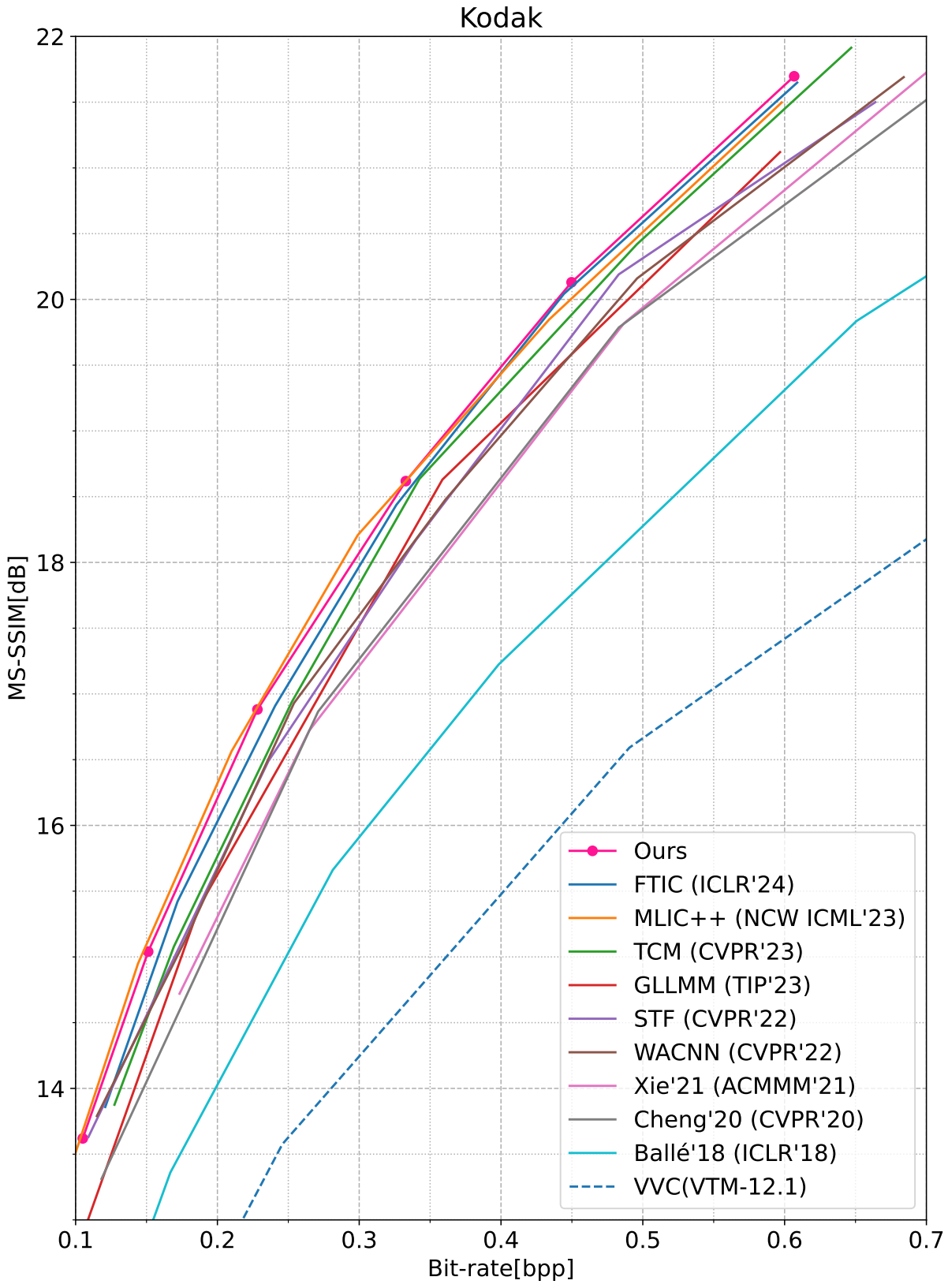

Figure 8. Performance evaluation (MS-SSIM) on the Kodak dataset.