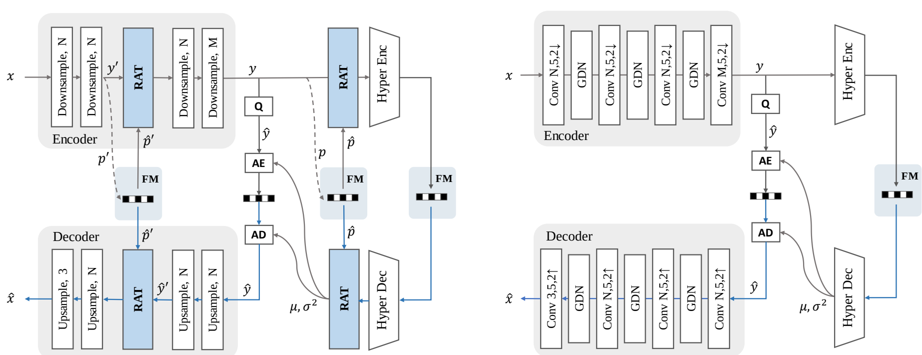
Appendix

6 Network Design of Ours-S and BL-S

The detailed designs of two simplified models Ours-S and BL-S are illustrated in Fig. 10. The Hyper Decoder consists of two modules, Hyper Decoder Scales and Hyper Decoder Means, which have identical network structures and each generates the estimated σ and μ for $\hat{y} \sim \mathcal{N}(\mu, \sigma^2)$. The RAT after the Hyper Decoder also consists of both the scale and mean portions, which is not fully depicted in Fig. 2 (a) and Fig. 10 for brevity. The network architectures of Hyper Encoder and Decoder are shown in Tab. 4.



(a) Ours-S (b) BL-S

Fig. 10: The overall framework of Ours-S and BL-S. We construct these two simplified models without self-attention and auto-regressive mechanisms. FM is a Factorized Model [5]. The details of Hyper Encoder and Decoder are shown in Tab. 4.

7 Rate-Distortion Performance on MS-SSIM

We also compare the MS-SSIM results with other LIC methods and VTM-17.0. Here we additionally introduce some LIC methods, *e.g.* Factorize Model [5] (denoted as Balle2017), Hyperprior Entropy Model [6] (denoted as Balle2018), Joint Autoregressive and Hierarchical Priors Model [33] (denoted as Minnen2018). Fig. 11 shows that our SegPIC outperforms these methods.

Table 4: The network architecture of Hyper Encoder, Hyper Doecoder Scales, and Hyper Decoder Means. "Conv 256,3,2↓" means the convolution operation with out-channel 256, kernel size 3 and downsample step 2. "TConv" means transposed convolution.

Hyper Encoder	Hyper Decoder Scales	Hyper Decoder Means
Conv 320,3,1	Conv 192,3,1	Conv 192,3,1
GELU	GELU	GELU
Conv 288,3,1	TConv 224,3,2↑	TConv 224,3,2↑
GELU	GELU	GELU
Conv 256,3,2↓	Conv 256,3,1	Conv 256,3,1
GELU	GELU	GELU
Conv 224,3,1	TConv 288,3,2↑	TConv 288,3,2↑
GELU	GELU	GELU
Conv 192,3,2↓	Conv 320,3,1	Conv 320,3,1

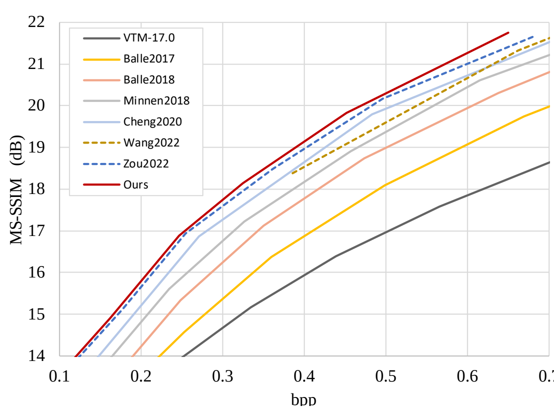


Fig. 11: The rate-distortion performance on Kodak. The metrics are bpp↓ and MS-SSIM↑. All the LIC methods are optimized by loss function Eq. 3, where the distortions can be formulated as $D = (1 - \text{MS-SSIM}(x, \hat{x}))$.