Coarse-to-Fine Hyper-Prior Modeling for Learned Image Compression

Yueyu Hu Wenhan Yang Jiaying Liu

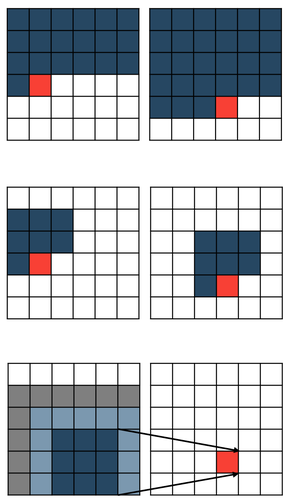
Wangxuan Institute of Computer Technology, Peking University, Beijing, China

Contributions

- Multi-Layer hyper-priors reduce spatial redundancy for improved learned image compression.
- Signal Preserving Hyper Transform facilitates coarse-to-fine modeling for latent representations.
- Information aggregation network to utilize multi-layer hyper-priors for the reconstruction of the image.

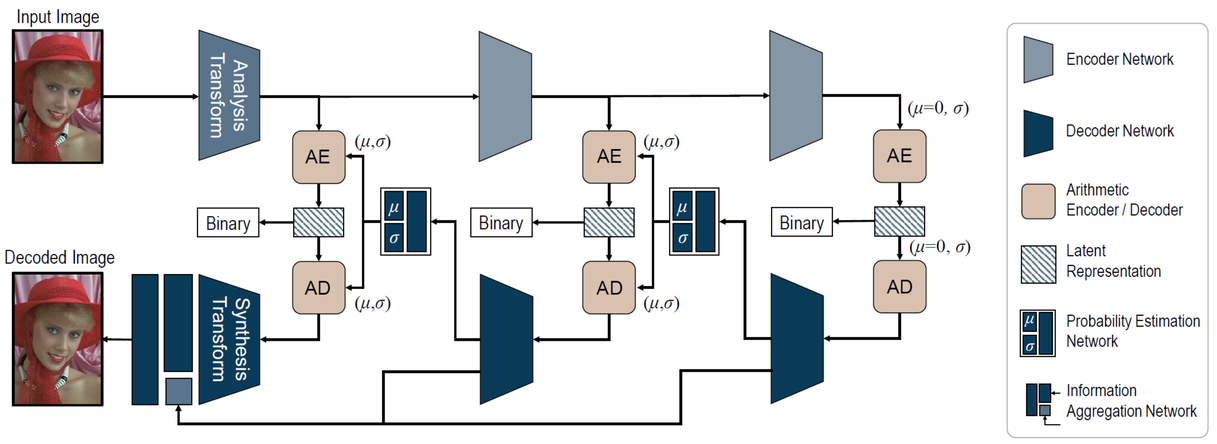
Formulation

- **Joint Probability Estimation**
  $P(\mathbf{X}) = P(X_1)P(X_2 \mid X_1)\cdots P(X_i \mid X_{i-1},...,X_1)$
  Large parameter space → hard to model
- **Locality Assumption**
  $P(\mathbf{X}) \doteq \prod_i P(X_i \mid X_{i-1}, X_{i-2},..., X_{i-m})$
  Hard to maintain accuracy & keep efficiency
- **Coarser-to-Fine Hyper-Prior**
  $P(\mathbf{X}) = P(\mathbf{X},\mathbf{Y}) = P(\mathbf{Y})P(\mathbf{X} \mid \mathbf{Y}), \quad \mathbf{Y} = f(\mathbf{X})$
  Divide and Conquer
  Model $\mathbf{Y}$ and estimate $\mathbf{X} \mid \mathbf{Y}$ conditionally



Coarse-to-Fine Modeling

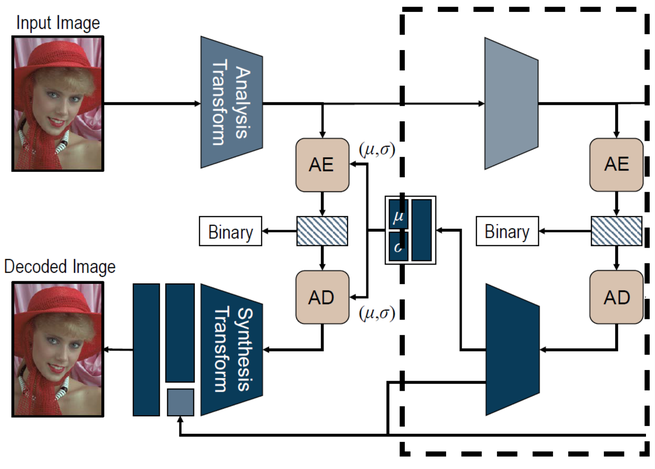
- Extract hyper representation $\mathbf{Y}$ with auto-encoder.
- Assume $\mathbf{X} \mid \mathbf{Y}$ to be Gaussian, μ and σ calculated from $\mathbf{Y}$.
- All layers can be executed in full parallel.



Hyper Transform

- Hyper representations are less correlated
- Large kernels with ReLU result in information loss
- **Expand dimension for non-linearity**
- **Large kernels → *Space-to-Depth* + small kernels**
- **Provide rich information for multi-layer analysis**

Signal Preserving Hyper Transform

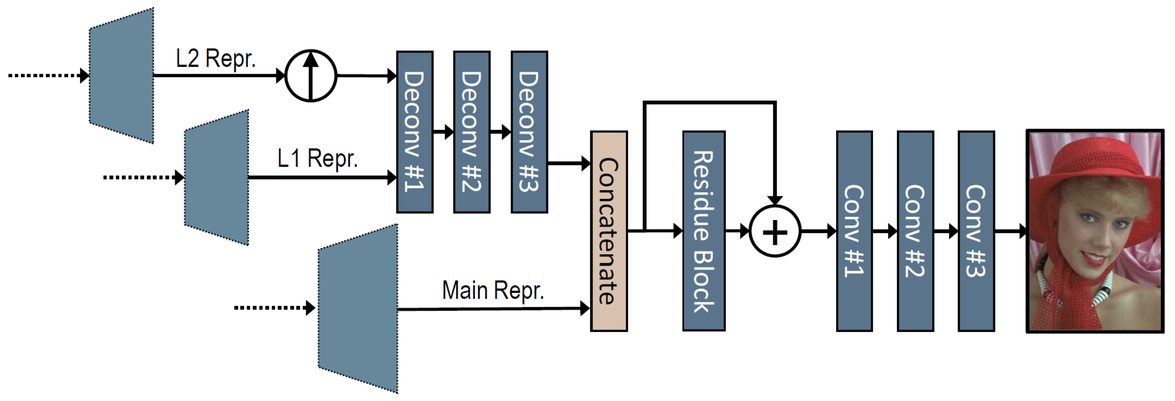


Information Aggregation

- Hyper Representations → Coarse Image Feature
- Facilitate reconstruction of basic components in images

Information Aggregation Reconstruction sub-network

- Aggregate hyper representations of different scales
- Fully convolutional → parallel accelerated

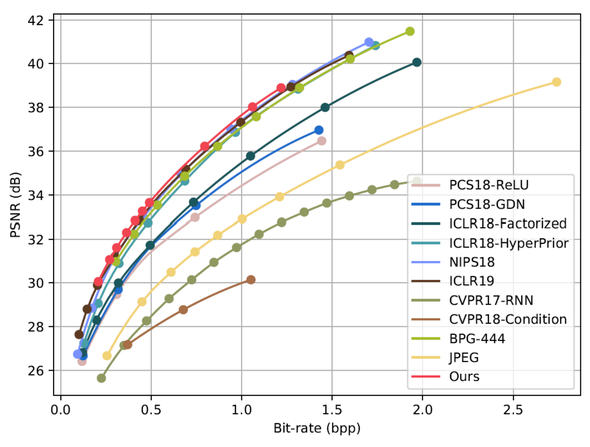


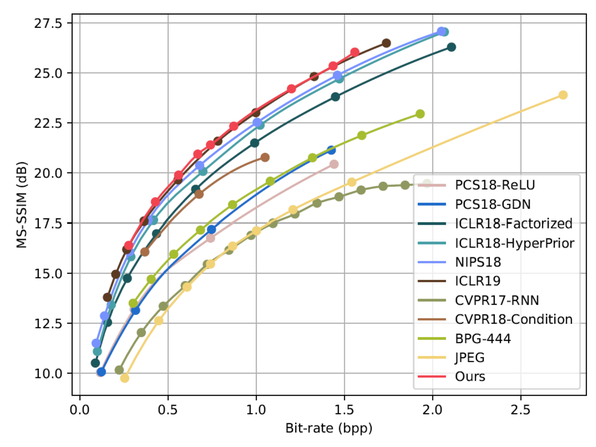
Experimental Results

BD-Rate on Kodak, Tecnick, and CLIC 19 Dataset

Methods	Kodak	Tecnick	CLIC 19
CVPR17-RNN	212.81%	244.26%	N/A
PCS18-ReLU	54.88%	55.17%	56.85%
PCS18-GDN	41.63%	38.18%	53.93%
ICLR18-Factorized	32.45%	32.17%	52.11%
ICLR18-HyperPrior	3.43%	-5.44%	10.15%
NIPS18	-4.80%	**-16.95%**	-1.06%
ICLR19	-4.94%	26.84%	18.48%
Ours	**-9.38%**	-16.50%	**-13.15%**
BPG-4:4:4	0	0	0
JPEG	115.05%	217.84%	120.47%

RD-Curve on Kodak Dataset (PSNR and MS-SSIM)





Visual Quality

Ours 31.3 dB / 0.610 bpp BPG 30.9 dB / 0.610 bpp

For further details

STRUCT Project

STRUCT @ PKU