Neural Image Compression summary

G. Drakos, C. Pantazopoulos, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.1

Artificial Intelligence &
Information Analysis Lab

Outline

- **Introduction**
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Introduction

Introduction

- Image compression reduces the amount of data required to represent an image.
- The number of images compressed and decompressed daily is innumerable.
- Image compression plays an important role in multimedia storage and Internet communication.
- It is the most useful and commercially successful technologies in the field of Digital Image Processing.

Outline

- Introduction
- **Image Compression Types**
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Image Compression Types

Lossless image compression

- The original image can be readily retrieved.
- Example: legal and medical multimedia documents.
- It exploits only statistical pixel redundancy.
- Compression ratio is not very big.
- Example image type: .png

Image Compression Types

Lossy image compression.

- Information is lost and the original image cannot be fully retrieved.
- It exploits both statistical and spatial image redundancy.
- It takes into account human visual perception properties.
- Example image type: .jpg

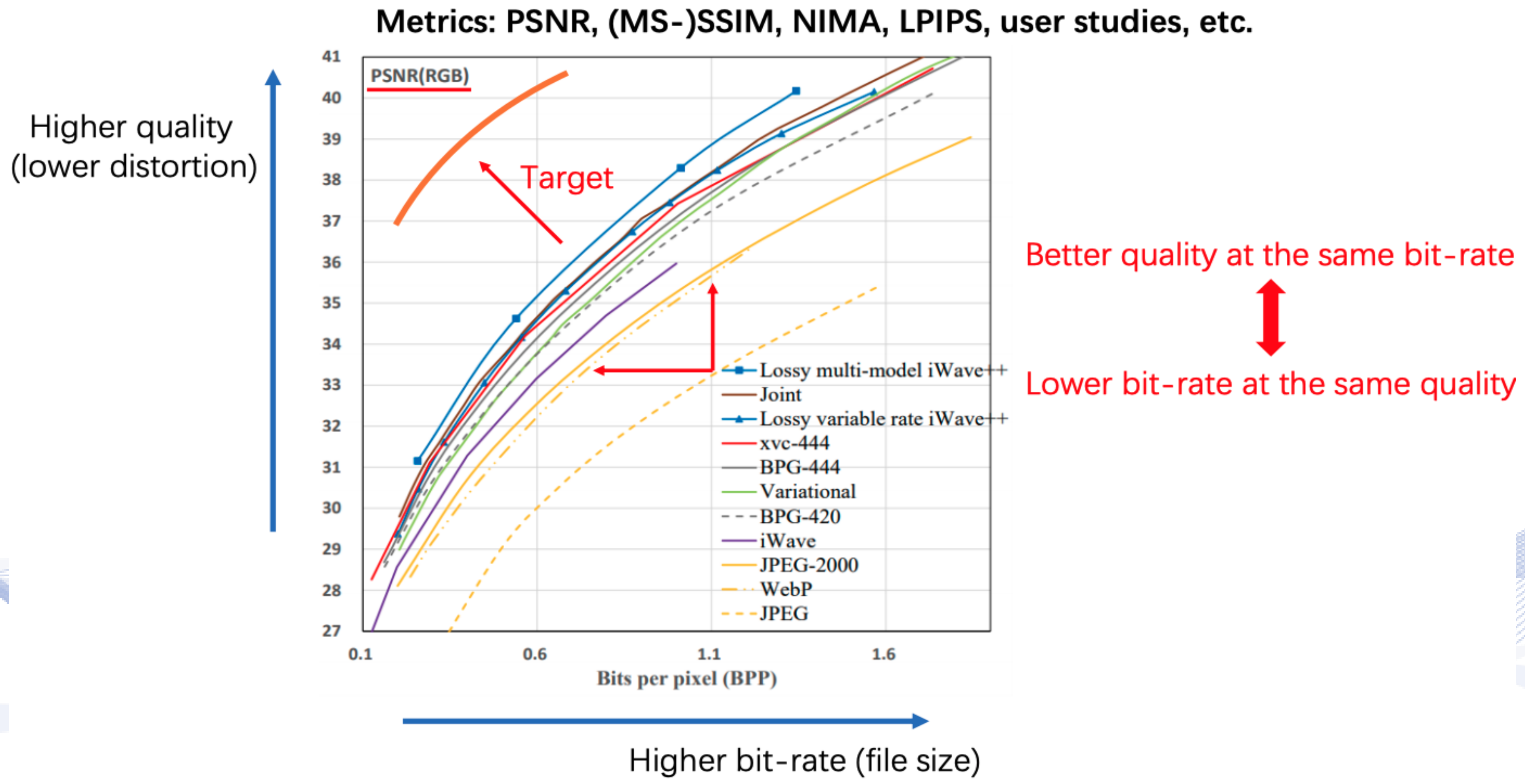
Outline

- Introduction
- Image Compression Types
- **Image Compression Evaluation**
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Image Compression Evaluation

- ***Peak Signal to Noise Ratio*** (PSNR): Ratio between the maximum possible power of an image and the power of the compression noise that affects image quality.
  - It is expressed in the logarithmic decibel scale.
- ***PSNR-HVS***: Peak signal-to-noise ratio extension that takes into account HVS properties (e.g., contrast perception).
- ***Multi-Scale Structural Similarity Measure*** (MS-SSIM): It considers image degradation as *perceived change in structural image content.*
  - It incorporates HVS perceptual characteristics.

Image Compression Evaluation

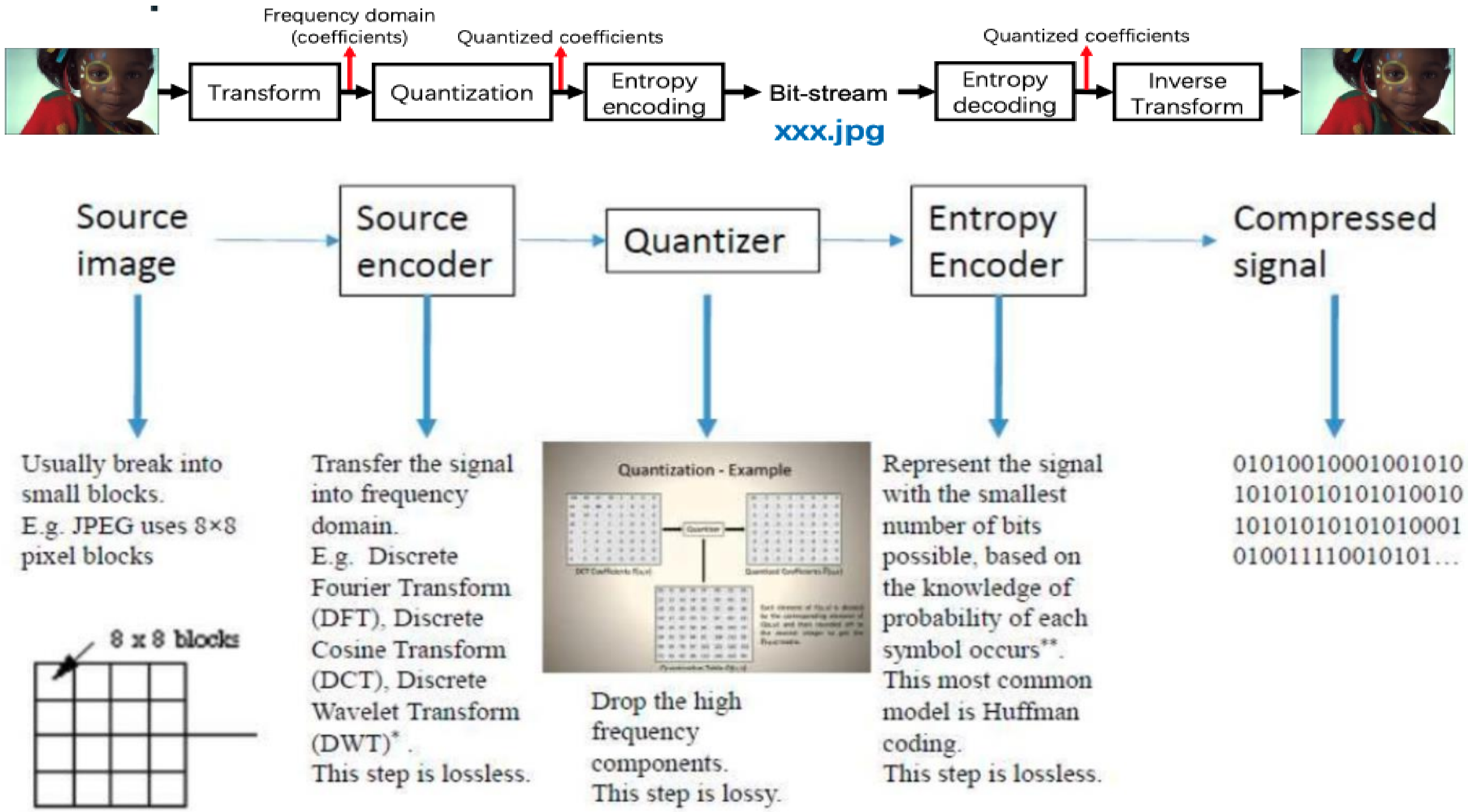


Rate-distortion trade-off [BHA2000].

Outline

- Introduction
- Image Compression Types
- Image Compression Evaluation
- **Transform Image Compression**
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Transform Image Compression



Transform compression pipeline [BHA2000].

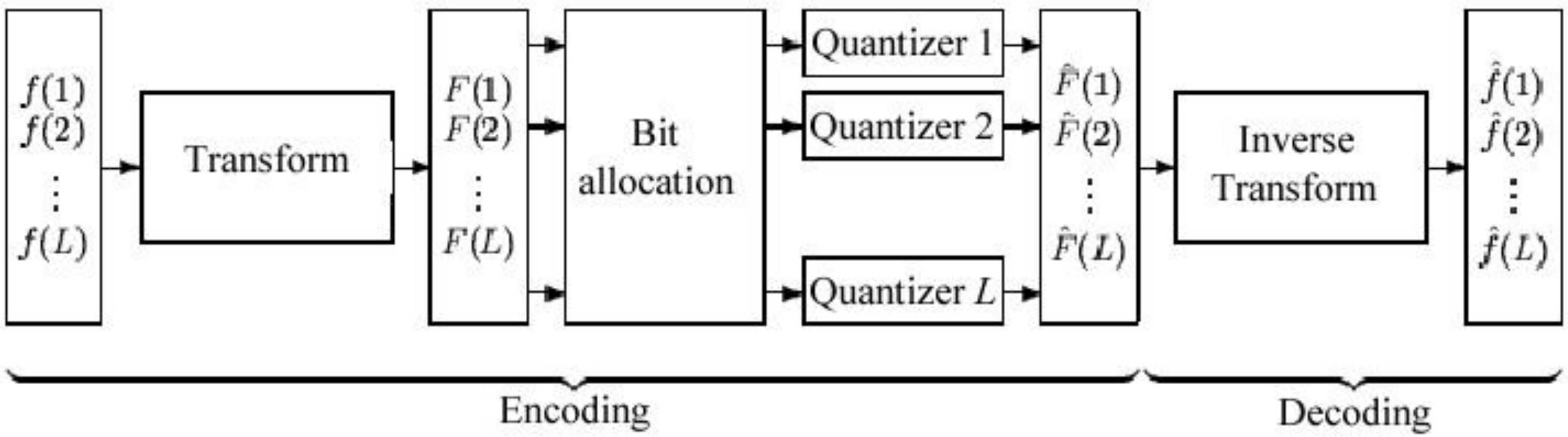
Transform Image Compression

- Linearly transformation of the image data.
- DCT, DWT transforms.
- Many transform coefficients:
  - have small values quantized to 0 without causing noticeable artifacts;
  - are uncorrelated and hence can be coded independently without losing efficiency.

Transform Image Compression

Digital image transforms concentrate image energy in a few transform coefficients.

- Heavy quantization or deletion of most transform coefficients leads to big lossy compression.



Transform encoding/decoding.

Outline

- Introduction
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- **Neural Predictive Image Coding**
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Neural Predictive Image Coding

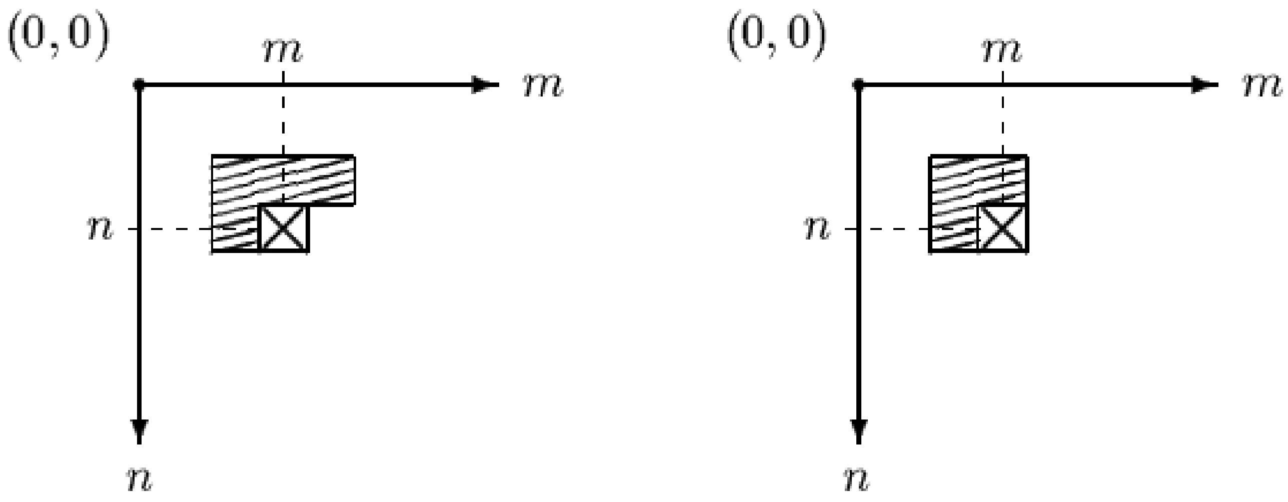
- One way to describe information redundancy in digital images is to use ***local image neighborhood predictability*** .
- Pixel intensity $f(n,m)$ can be predicted from the pixel intensities in its ***local neighborhood*** $\mathcal{A}$:

$$\hat{f}(n,m) = L\big[f(n-i, m-j), \quad (i,j) \in \mathcal{A}, \quad (i,j) \neq (0,0)\big].$$

- ***Causal prediction*** is used, which is based on already reconstructed past pixel values:

$$\hat{f}(n,m) = L\big[f_r(n-i, m-j), \quad (i,j) \in \mathcal{A}\big].$$

Neural Predictive Image Coding



Causal windows used in image prediction.

Neural Predictive Image Coding

- It is sufficient to code the ***prediction error***:

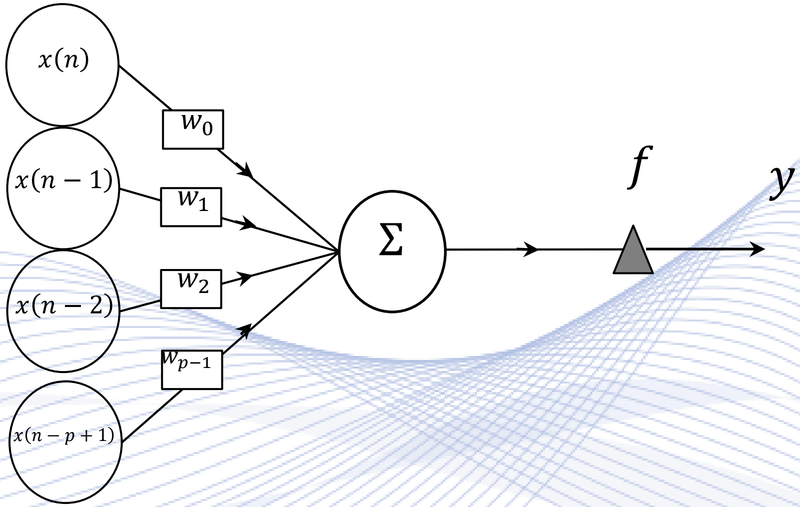
$$e(n,m) = f(n,m) - \hat{f}(n,m).$$

- If the prediction is good, the error term has a small dynamic range and a substantial compression can be achieved.
- For pixel $f_r(n,m)$ reconstruction, the transmission of the prediction coefficients and of the coded error is needed.
- If $e_q(n,m)$ is the quantized and decoded error value, the pixel value can be reconstructed as follows:

$$f_r(n,m) = L\left[f_r(n-i, m-j), \quad (i,j) \in A\right] + e_q(n,m).$$

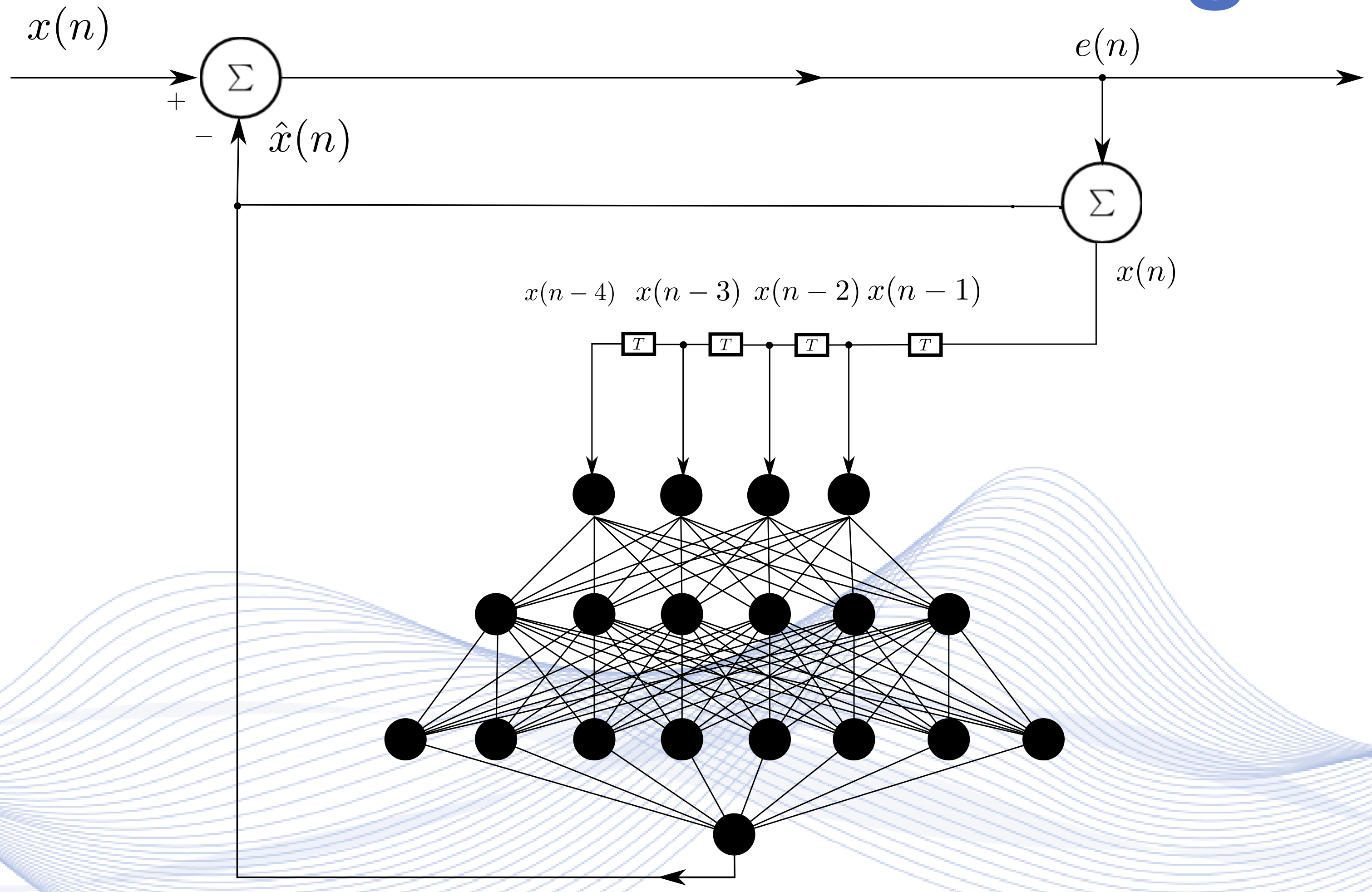
Neural Predictive Coding

Perceptron performs ***non-recurrent*** signal processing.



Perceptron signal processing using sliding window.

Neural Predictive Coding



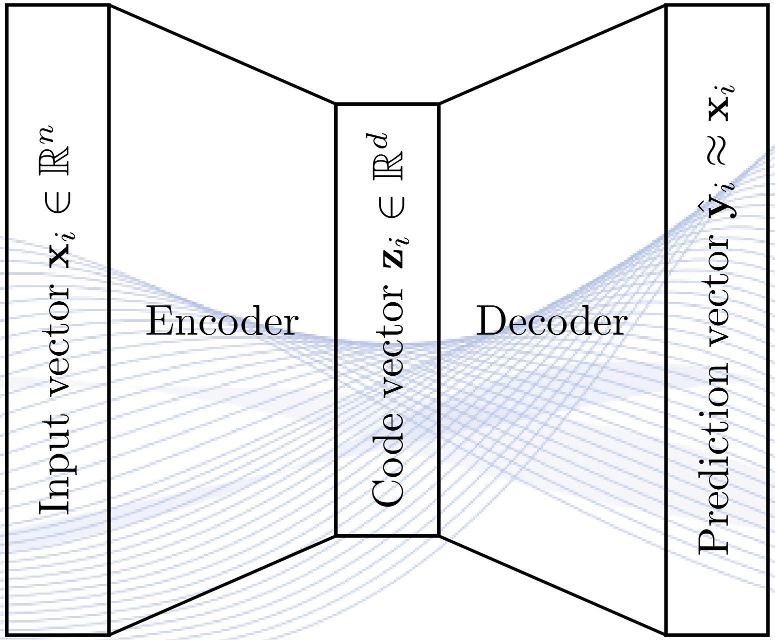
DPCM with a multilayer perceptron.

Outline

- Introduction
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- **Neural Image Autoencoding**
- CNN-Transformer Image Compression
- RNN Image Compression
- Variable Rate RNN Image Compression

Neural Image Autoencoding

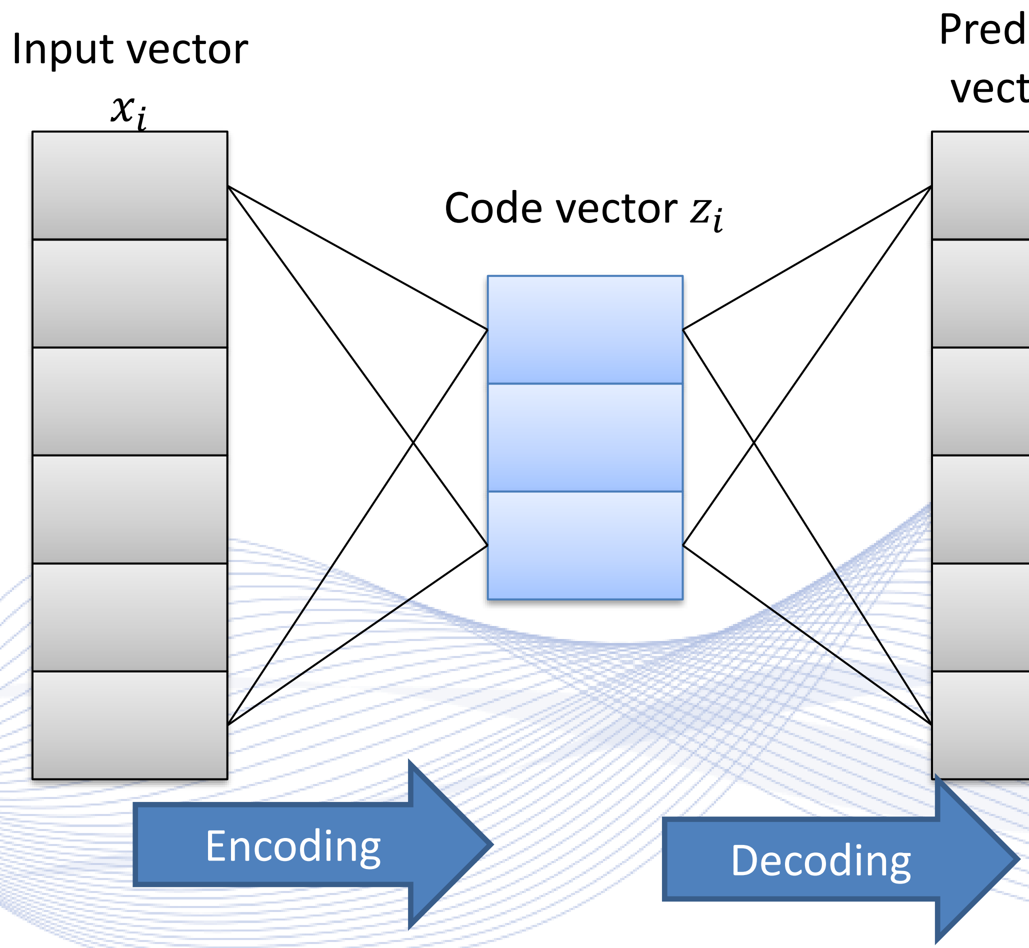
- Autoencoders are typically used for dimensionality reduction and feature extraction.
- The encoder output could be used for image coding.



Autoencoder architecture.

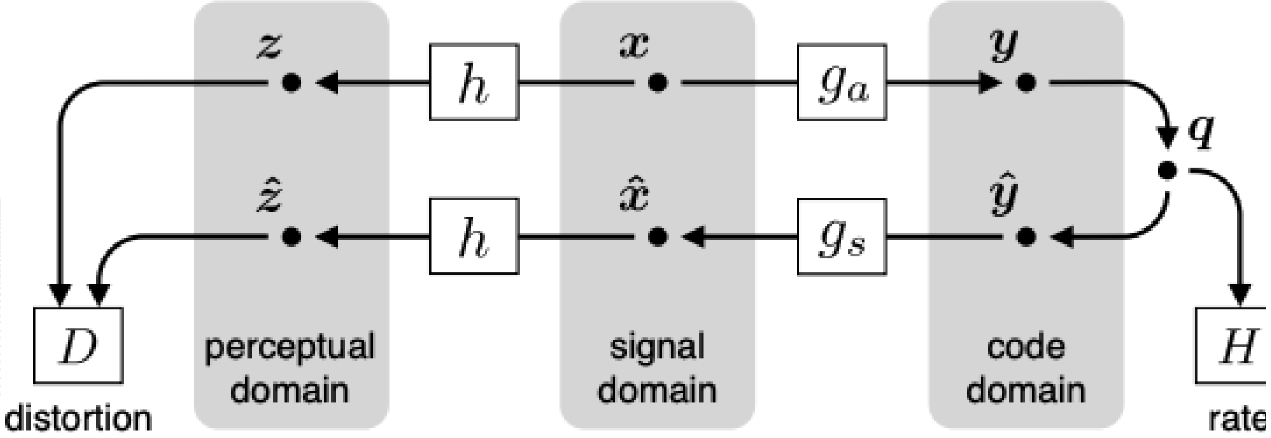
Artificial Intelligence & Information Analysis Lab

Neural Image Autoencoding



Autoencoder architecture.

Neural Image Autoencoding



$$L[g_a, g_s] = H[P_q] + \lambda \mathbb{E}\,\|z - \hat{z}\|.$$

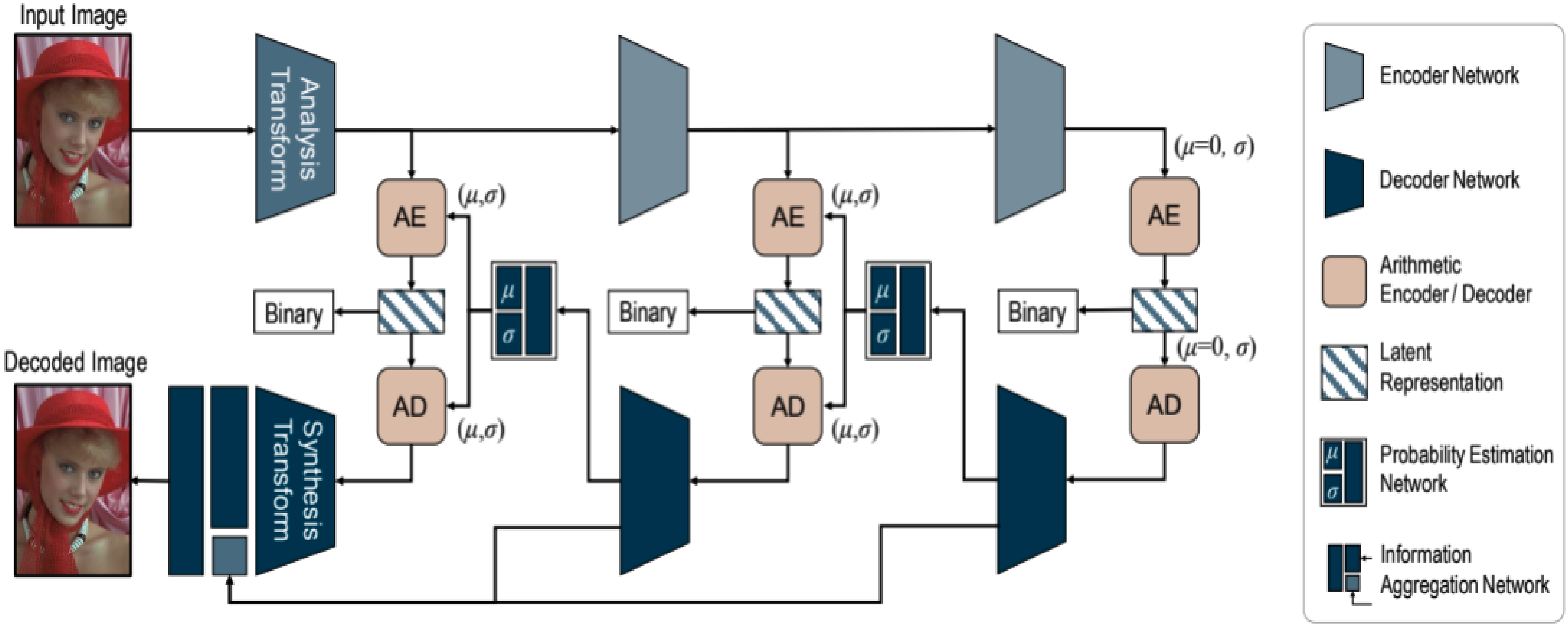
$$\boldsymbol{y} = g_a(\boldsymbol{x}; \boldsymbol{\phi})$$
$$\hat{\boldsymbol{y}} = Q(\boldsymbol{y})$$
$$\hat{\boldsymbol{x}} = g_s(\hat{\boldsymbol{y}}; \boldsymbol{\theta})$$

Joint optimization of the reconstruction quality and the compressed image size [BAL2017].

Outline

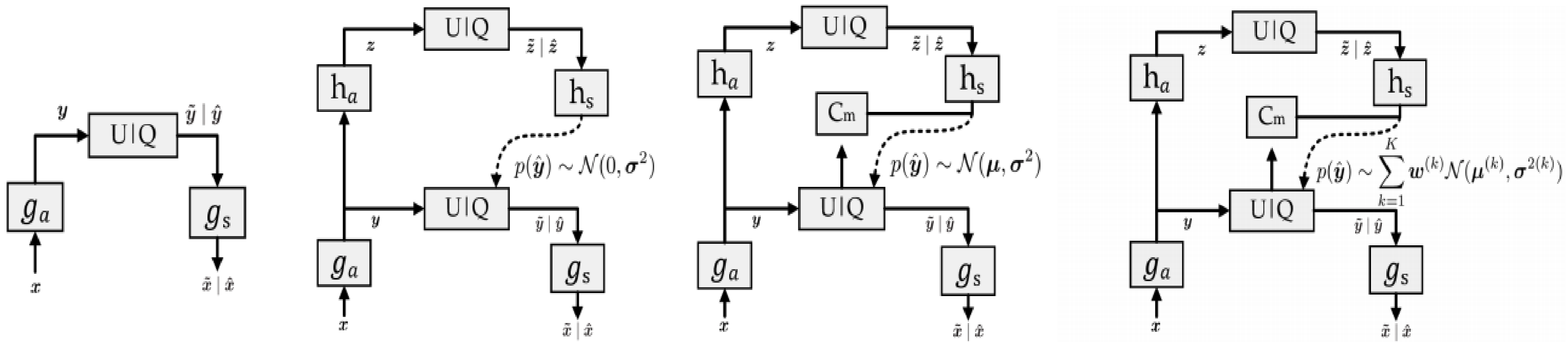
- Introduction
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- **CNN-Transformer Image Compression**
- RNN Image Compression
- Variable Rate RNN Image Compression

CNN-Transformer Image Compression



CNN based image compression [HU2020].

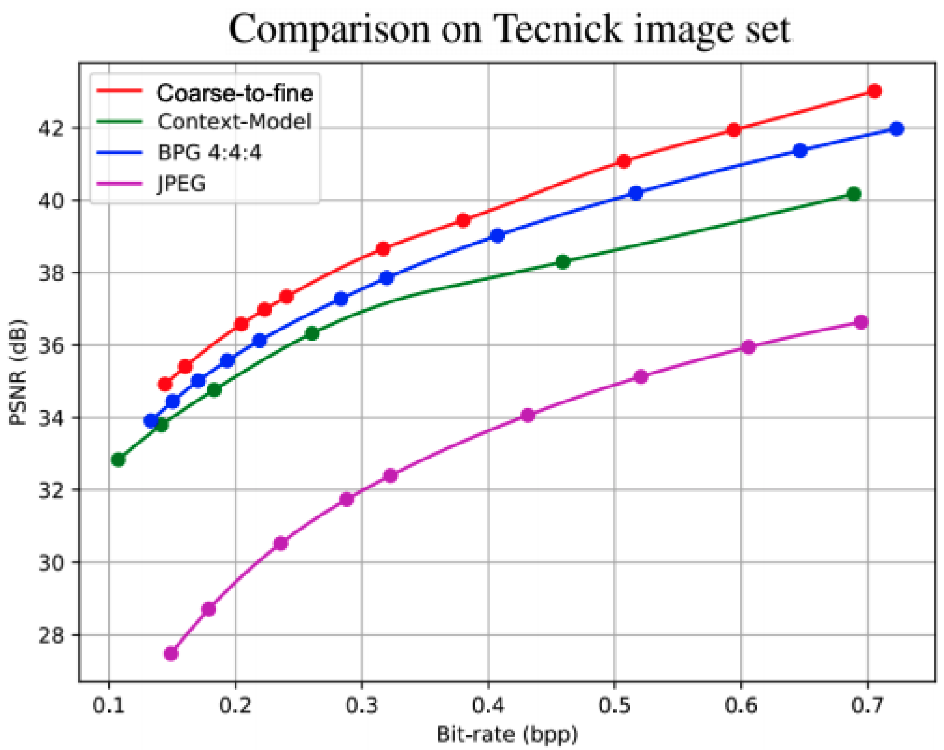
CNN-Transformer Image Compression

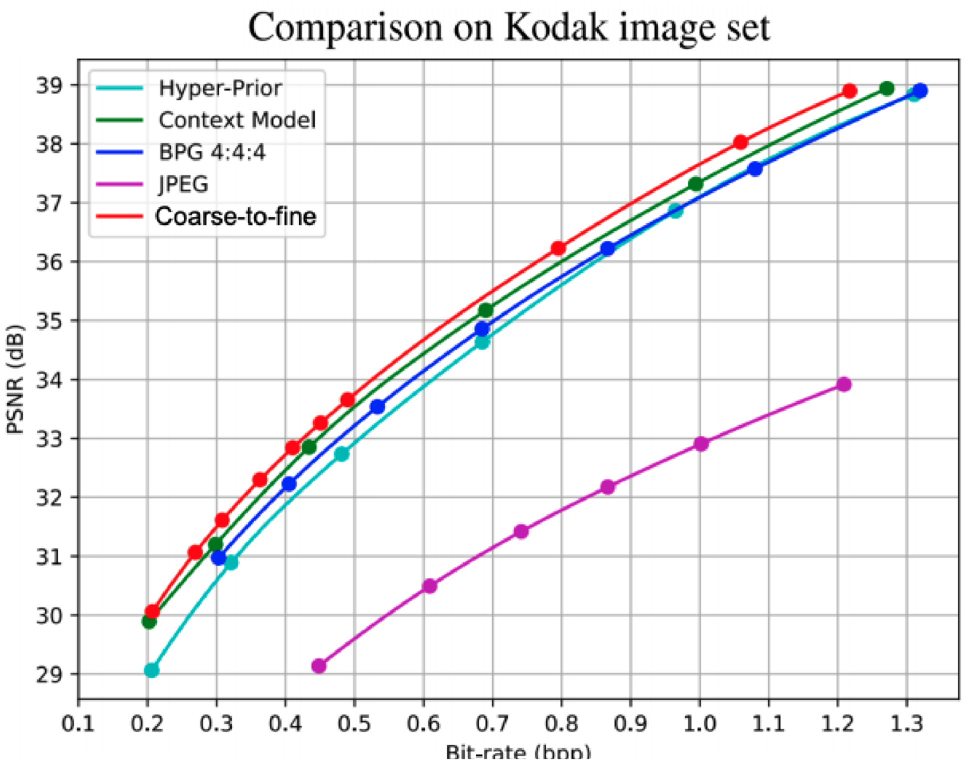


(a) Baseline (b) Hyperprior (c) Joint (d) Proposed Model

Quantization and entropy coding (U|Q) [HU2020].

CNN-Transformer Image Compression





Performance Evaluation [HU2020].

Outline

- Introduction
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- **RNN Image Compression**
- Variable Rate RNN Image Compression

RNN Image Compression

Different types of RNN architectures can be used:

- LSTMs
- Associative Long Short-Term Memory
- Gated Recurent Units (GRU).

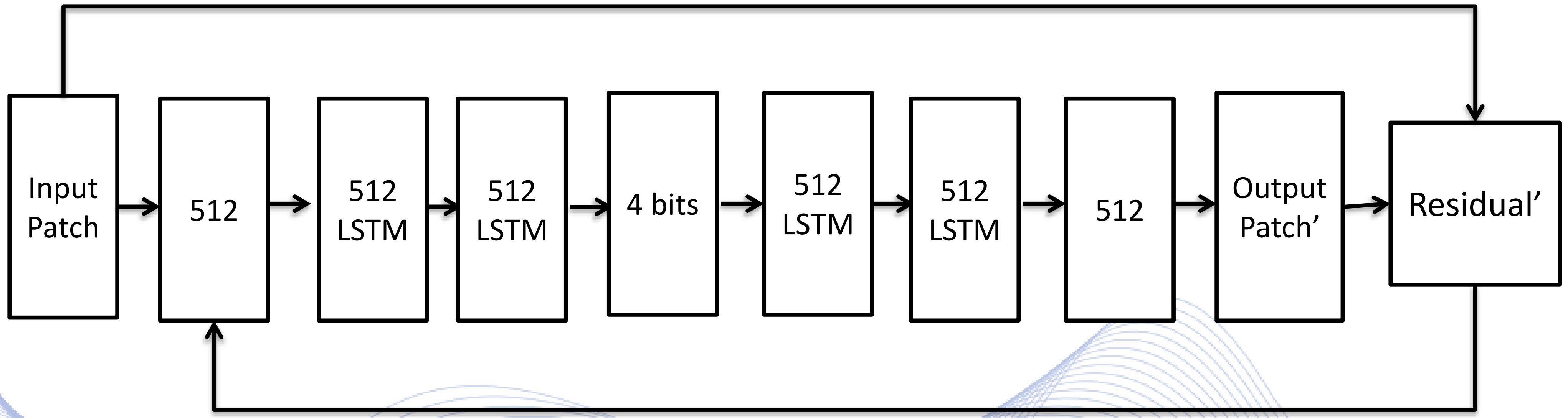
RNN Image Compression

- Feed-forward neural networks with autoencoders can be combined with Recurrent Neural Networks (RNN) for image compression.

- RNN models can be used in various ways to achieve better compression results.

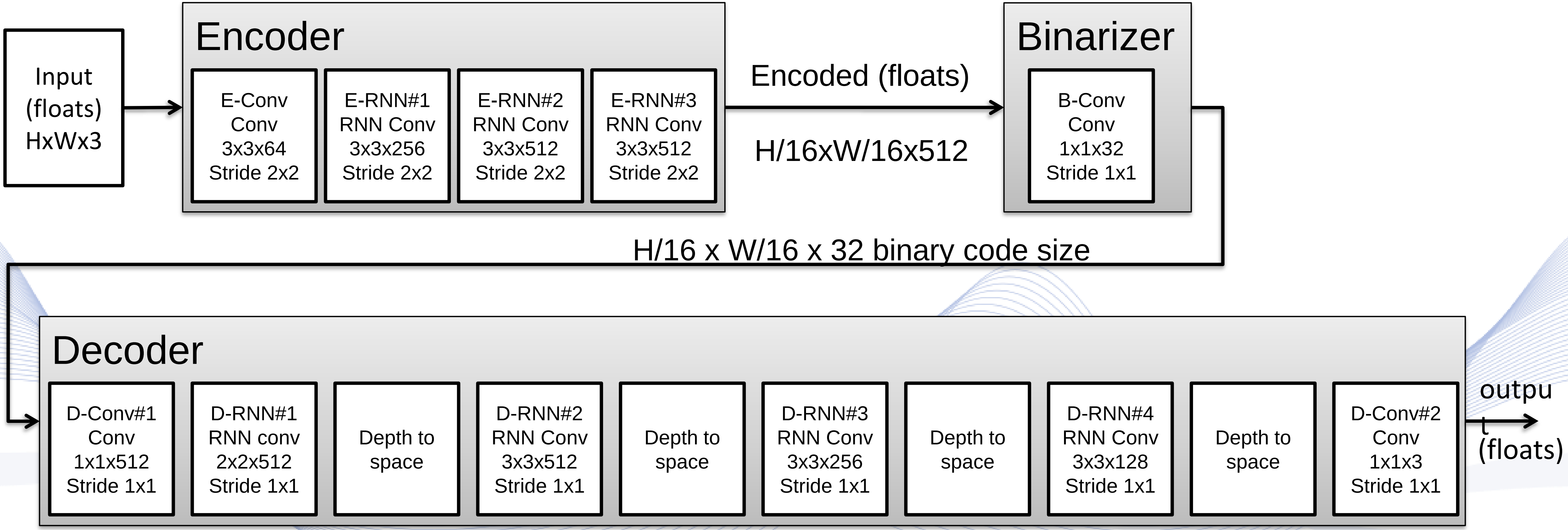
- They can be combined with CNNs to better capture image information.

RNN Image Compression



Fully connected LSTM residual encoder. LSTM layers have 512 fully connected units [TOD2016] .

RNN Image Compression



RNN image compression architecture.

RNN Image Compression

Trained on the 32×32 dataset.				
Model	Rank	MS-SSIM AUC	Rank	PSNR-HVS AUC
GRU (One Shot)	**1**	**1.8098**	**1**	**53.15**
LSTM (Residual Scaling)	**2**	**1.8091**	4	52.36
LSTM (One Shot)	3	1.8062	3	52.57
LSTM (Additive Reconstruction)	4	1.8041	6	52.22
Residual GRU (One Shot)	5	1.8030	**2**	**52.73**
Residual GRU (Residual Scaling)	6	1.7983	8	51.25
Associative LSTM (One Shot)	7	1.7980	5	52.33
GRU (Residual Scaling)	8	1.7948	7	51.37
Baseline		1.7225		48.36
JPEG				
YCbCr 4:4:4		1.7748		51.28
YCbCr 4:2:0		1.7998		52.61

Performance measured on the Kodak dataset.

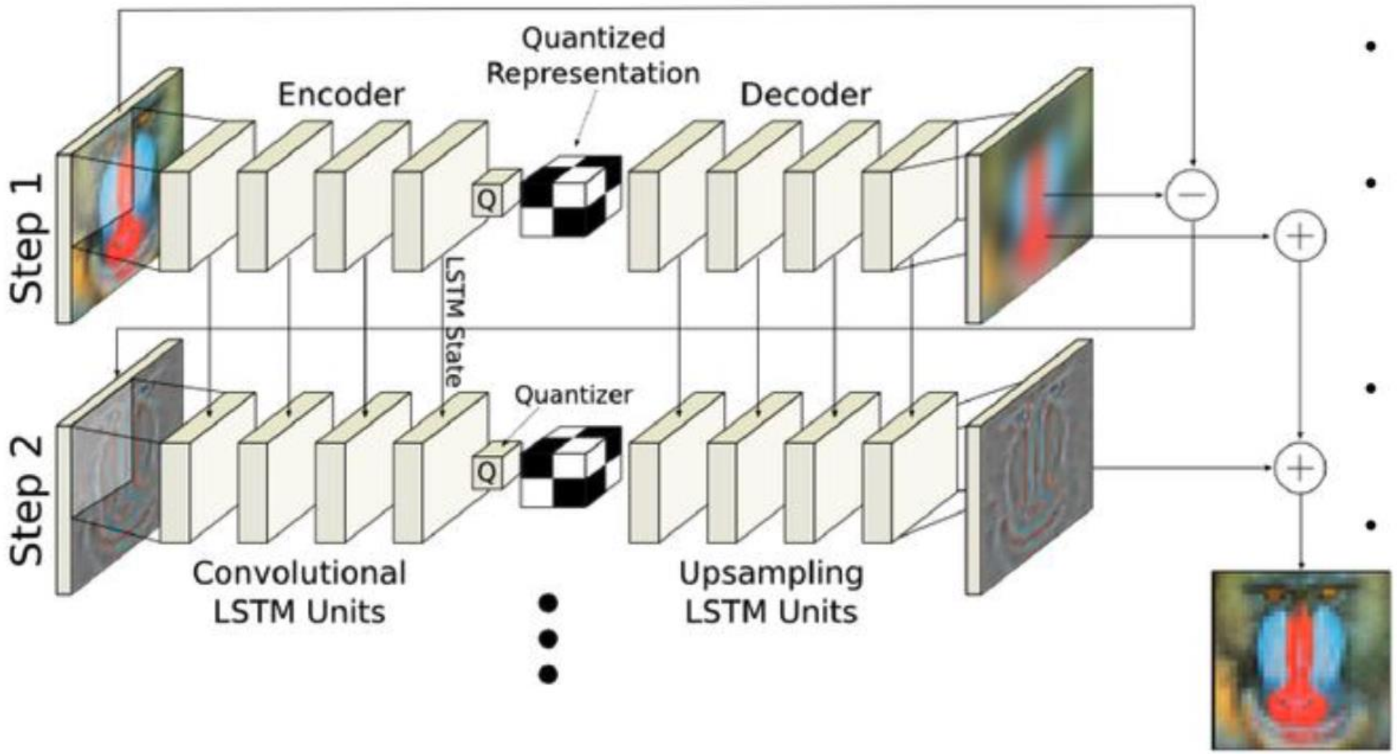
RNN Image Compression

a) JPEG 420 and b) Residual GRU(one-shot) compressed images.

Outline

- Introduction
- Image Compression Types
- Image Compression Evaluation
- Transform Image Compression
- Neural Predictive Image Coding
- Neural Image Autoencoding
- CNN-Transformer Image Compression
- RNN Image Compression
- **Variable Rate RNN Image Compression**

Variable Rate RNN Image Compression

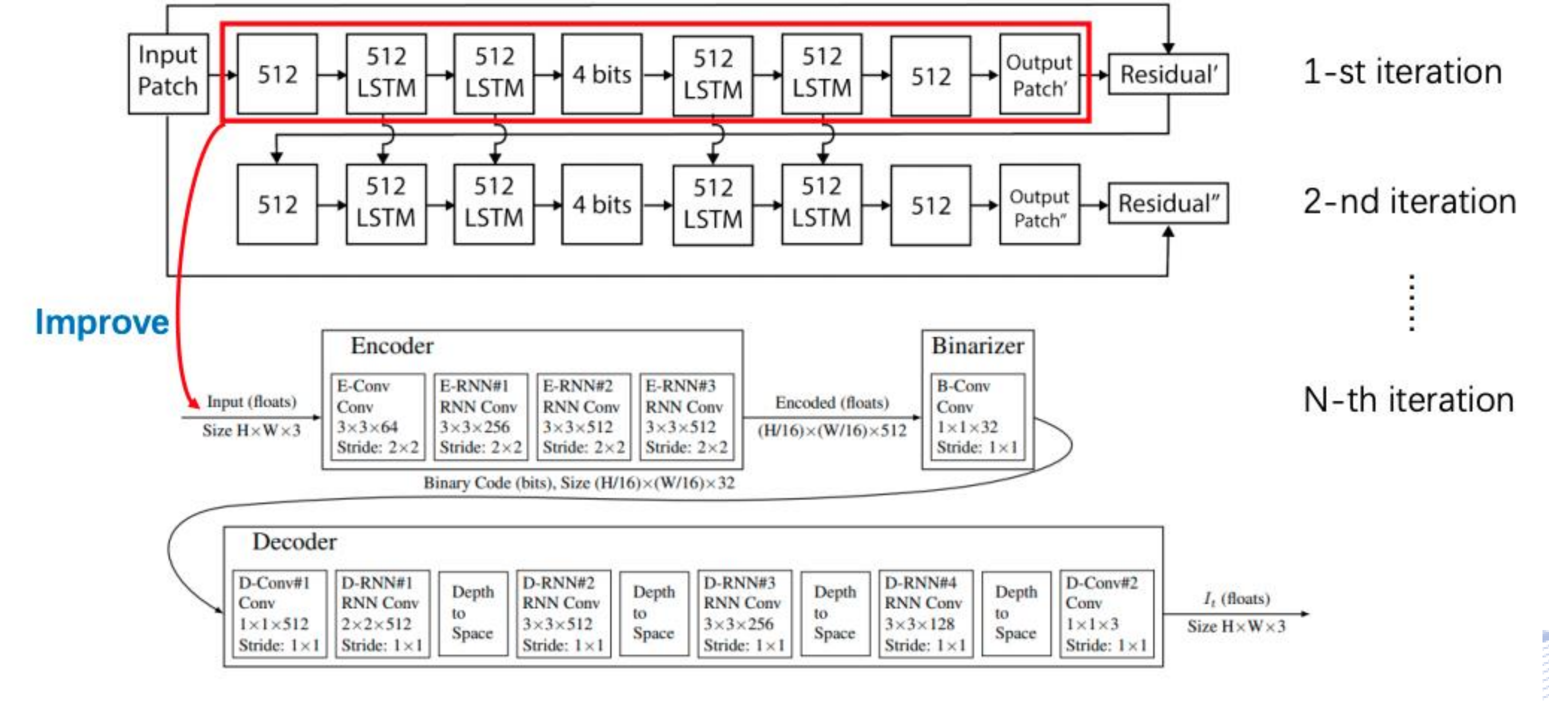


Conv/Deconv LSTM architecture [TOD2016].

Variable Rate RNN Image Compression

- In fixed rate image compression, a separate model has to be trained for each point along the rate–distortion curve.
- ***Variable rate RNN image compression*** can be performed based on the autoencoder architecture and residual encoding.
- A single network can recurrently be used for different bitrates, depending on the number of times the input passes through the network.
- It extracts hierarchical latent representation with increasing levels of detail, which are focused on global information first.

Variable Rate RNN Image Compression



Convolutional / deconvolutional residual encoder [TOD2016].

Variable Rate RNN Image Compression

Left: Original image, I = R[0]. **Center:** Reconstructed image, P[1]. **Right:** the residual, R[1], which represents the error introduced by compression.

The second pass through the network. **Left:** R[1] is given as input. **Center:** A higher quality reconstruction, P[2]. **Right:** A smaller residual R[2] is generated by subtracting P[2] from the original image.

Residual image in iterative variable rate image compression [TOD2016].

Standardization of Learning-based image compression

JPEG initiates standardisation of image compression based on AI

The 89th JPEG meeting was held online from 5 to 9 October 2020.

During this meeting multiple JPEG standardisation activities and explorations were discussed and progressed. Notably, the call for evidence on learning-based image coding was successfully completed and evidence was found that this technology promises several new functionalities while offering at the same time superior compression efficiency, beyond the state of the art.

JPEG AI

At the 89th meeting the submissions to the Call for Evidence on learning-based image coding were presented and discussed. Four submissions were received in response to the Call for Evidence. The results of the subjective evaluation of the submissions to the Call for Evidence were reported and discussed in detail by experts. It was agreed that there is strong evidence that learning-based image coding solutions can outperform the already defined anchors in terms of compression efficiency, when compared to state-of- the-art conventional image coding architecture. Thus, it was decided to create a new standardisation activity for a JPEG AI on learning-based image coding system, that applies machine learning tools to achieve substantially better compression efficiency compared to current image coding systems, while offering unique features desirable for an efficient distribution and consumption of images. This type of approach should allow to obtain an efficient compressed domain representation not only for visualisation, but also for machine learning based image processing and computer vision. JPEG AI releases to the public the results of the objective and subjective evaluations as well as a first version of common test conditions for assessing the performance of leaning-based image coding systems.

JPEG statement on AI solutions [JPE2020].

Bibliography

[PIT2021] I. Pitas, “Computer vision”, Createspace/Amazon, in press.

[PIT2017] I. Pitas, “Digital video processing and analysis” , China Machine Press, 2017 (in Chinese).

[PIT2013] I. Pitas, “Digital Video and Television” , Createspace/Amazon, 2013.

[NIK2000] N. Nikolaidis and I. Pitas, “3D Image Processing Algorithms”, J. Wiley, 2000.

[PIT2000] I. Pitas, “Digital Image Processing Algorithms and Applications”, J. Wiley, 2000.

Bibliography

[TOD2015] G. Toderici, S.M. O'Malley, S.J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, R. Sukthankar, “Variable rate image compression with recurrent neural networks” arXiv preprint arXiv:1511.06085, 2015.

[TOD2016] G. Toderici, S.M. O'Malley, S.J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, R. Sukthankar, “Variable Rate Image Compression with Recurrent Neural Networks”, Proc. ICLR 2016.

[TOD2017] G. Toderici, D. Vincent, N. Johnston, S.J. Hwang, D. Minnen, J. Shor, M. Covell “Full Resolution Image Compression with Recurrent Neural Networks” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017), pp. 5306-5314, 2017.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr