Use of Deep Learning for Image/Video Compression

PhD Candidate: Congcong Wang[1]

Supervisors

Prof. Faouzi Alaya Cheikh[1], Prof. Ole Jakob Elle[2,3], Prof. Azeddine Beghdadi[4]

[1]Norwegian University of Science and Technology, Norway

[2]Oslo University Hospital, Norway

[3]University of Oslo, Norway

[4]University Sorbonne Paris Nord , France

Outline

- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

Outline

- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

Image/Video Compression

- Why is Compression Needed?
  - A two-hour standard definition (SD) television movie:

$$30\,\frac{frames}{sec} \times (720 \times 480)\,\frac{pixels}{frame} \times 3\,\frac{bytes}{pixel} = 31{,}104{,}000\ bytes/\sec = 31.104\ MB/sec$$

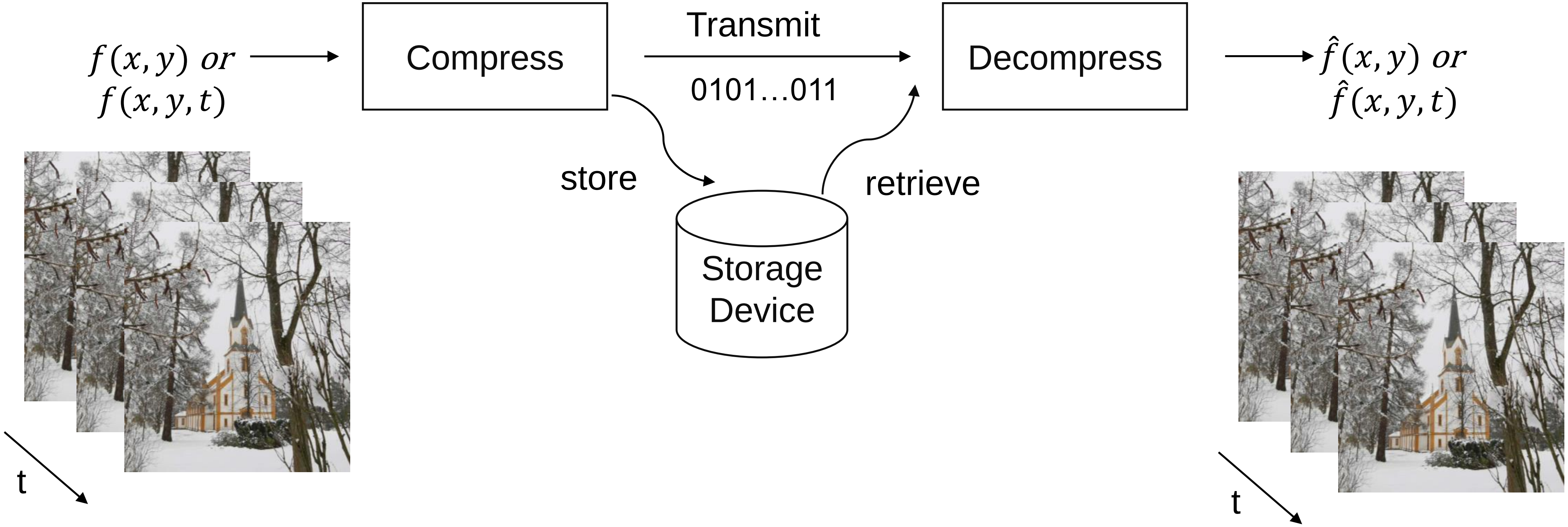
$$31{,}104{,}000\,\frac{bytes}{sec} \times (60^2)\,\frac{sec}{hr} \times 2hrs \cong 2.24 \times 10^{11} = 224\ GB$$

  - Full HD (1080p) 1920×1080 : $1344\ GB$

Compression

Image/Video Compression

- What is image/video compression?
  - The art and science of reducing the amount of data required to represent an image/video

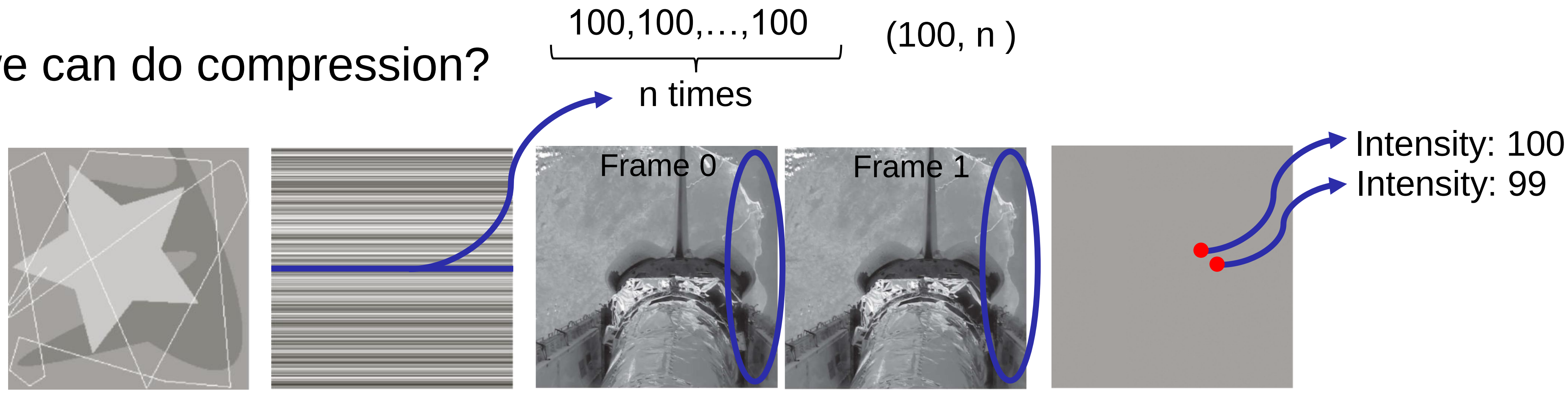


Applications

everywhere

Image/Video Compression

Why we can do compression?

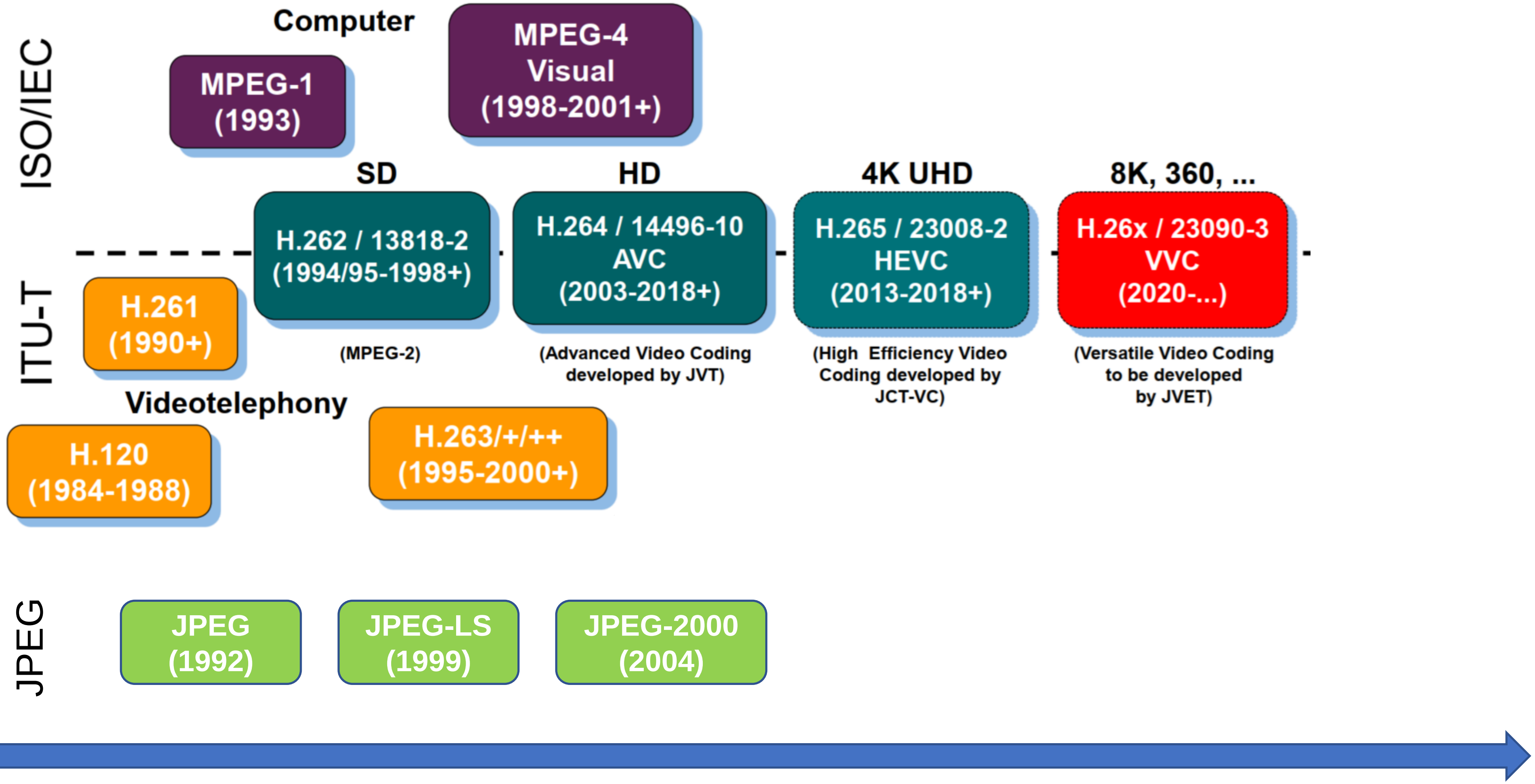


(a) (b) (c) (d)

- (a) Coding redundancy: length of the code words (e.g., 8-bit codes for grey value images) is larger than needed. (Variable length codes)
- (b) Spatial redundancy: correlation between pixels in space is not used in the representation.
- (c) Temporal redundancy: correlation between pixels in time is not used in the representation.
- (d) Irrelevant information: information that is not perceived by the human visual system or not relevant to a given application.

Gonzalez et al., Digital Image Processing

Image/Video Compression Standards



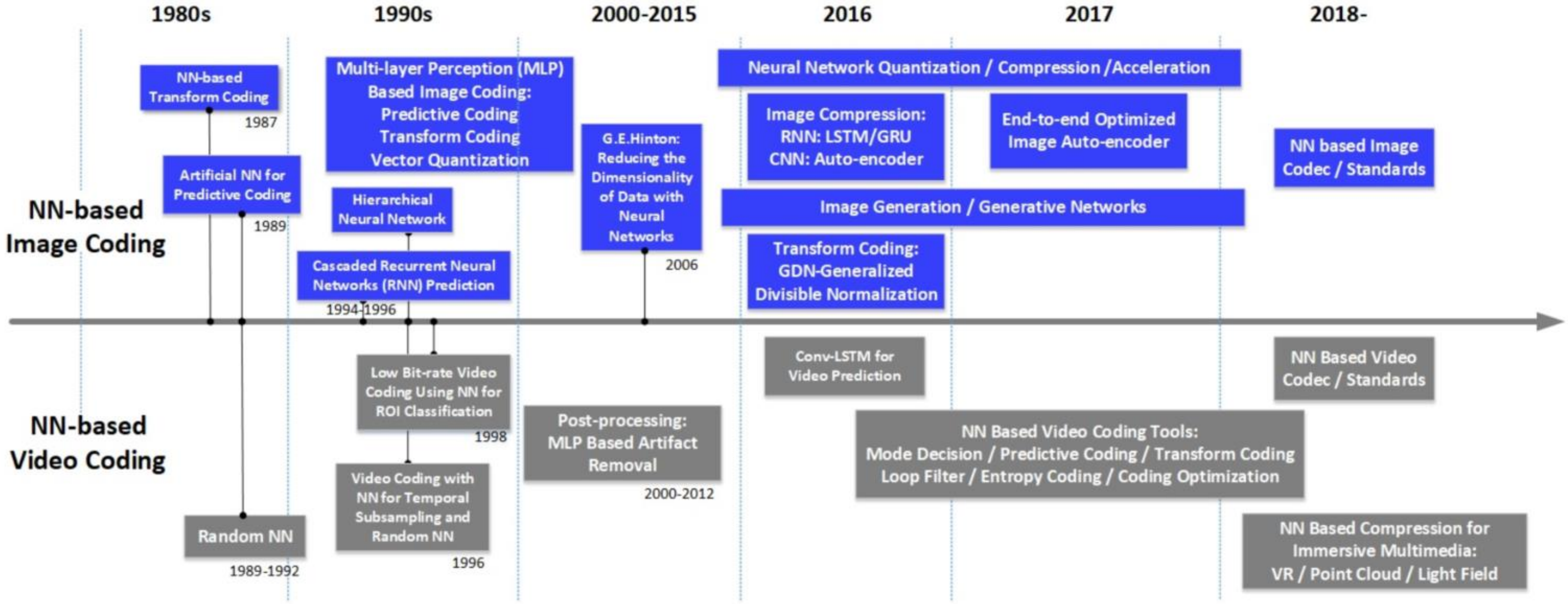
Picture Coding Syposium,1969

1960s

History line of image/video coding standards by ITU-T and ISO/IEC committees

Ohm et al., Trends and Recent Developments in Video Coding Standardization, ICME Tutorial

Neural Network Based Image/Video Compression



The technical roadmap of neural network based compression algorithms

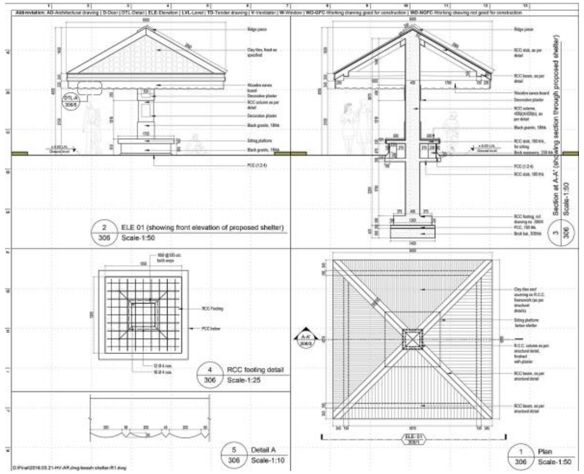
Ma et al., Image and Video Compression with Neural Networks: A Review, 2018

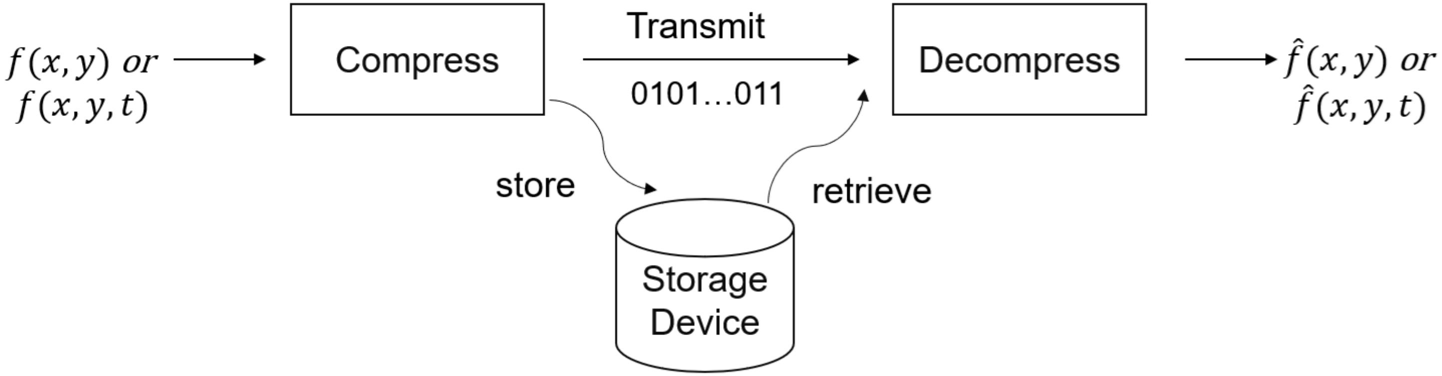
Outline

- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

Lossless Image Compression

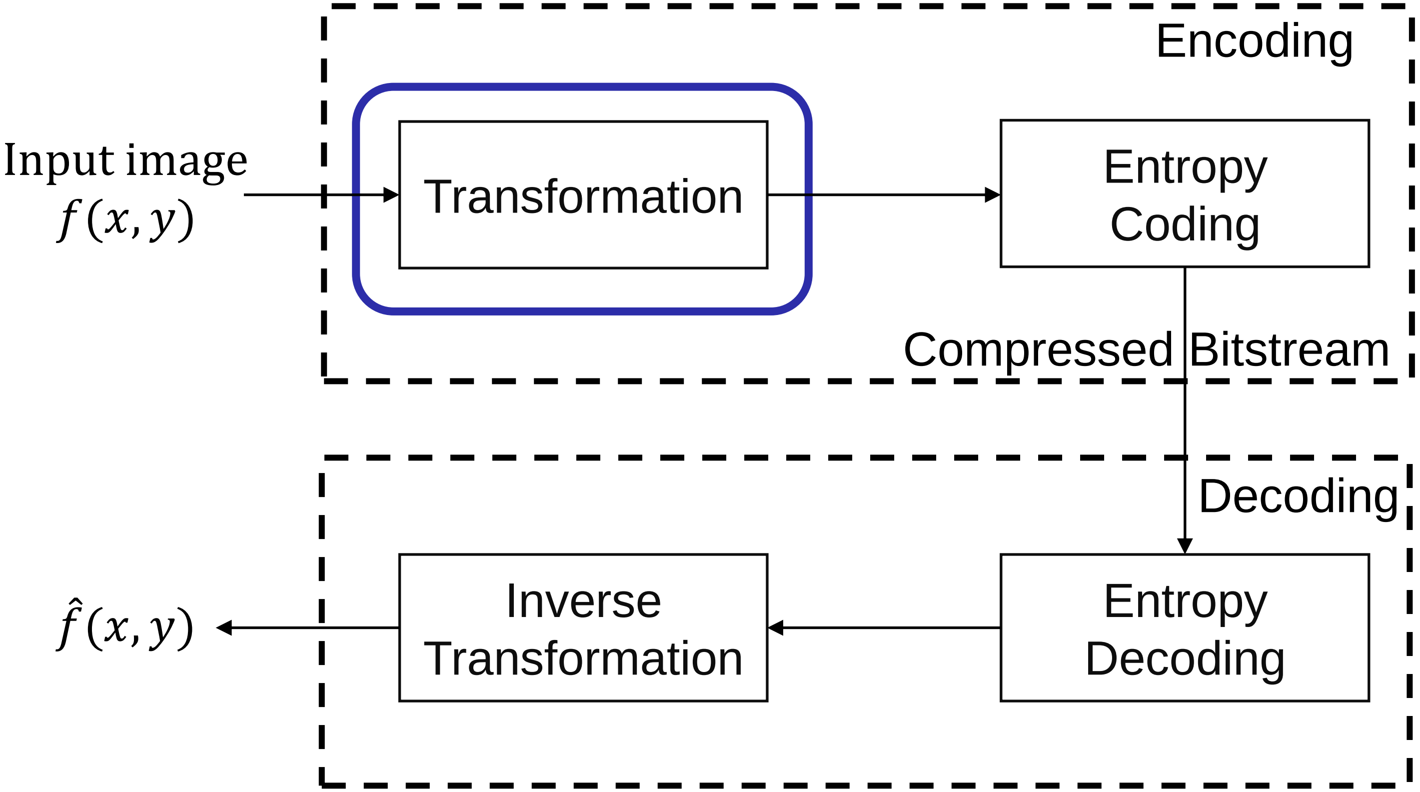
- Goal: To represent an image signal with the smallest possible number of bits without loss of any information
- Applications:
  - digital medical imagery, technical drawings, comics
- Lossless JPEG, JBIG, JBIG2, Lossless JPEG2000





Lossless Image Compression

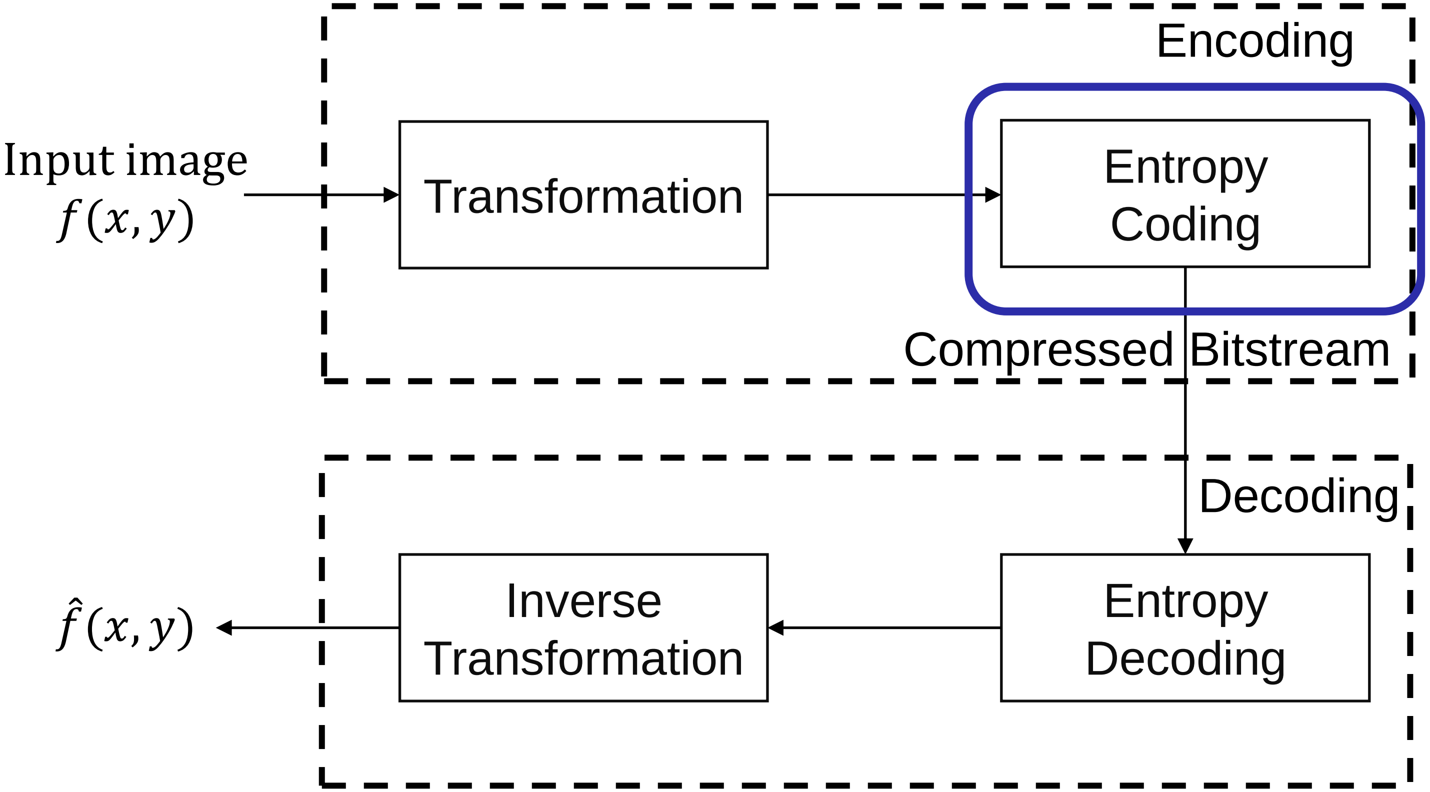
- Reversible transformation, convert $f(\boldsymbol{n})$ to $\hat{f}(\boldsymbol{n})$ that can be compressed more efficently.
- Discrete Cosine Transform (DCT), wavelet transform, color space transform: RGB to luminance-chrominance



A brief transformation based lossless coding system

Lossless Image Compression

- Generates a binary bitstream
- Variable-legnth coding / Entropy coding: Huffman, arithmetic coders, *etc.*



A brief transformation based lossless coding system

Lossless Image Compression

Compression efficiency

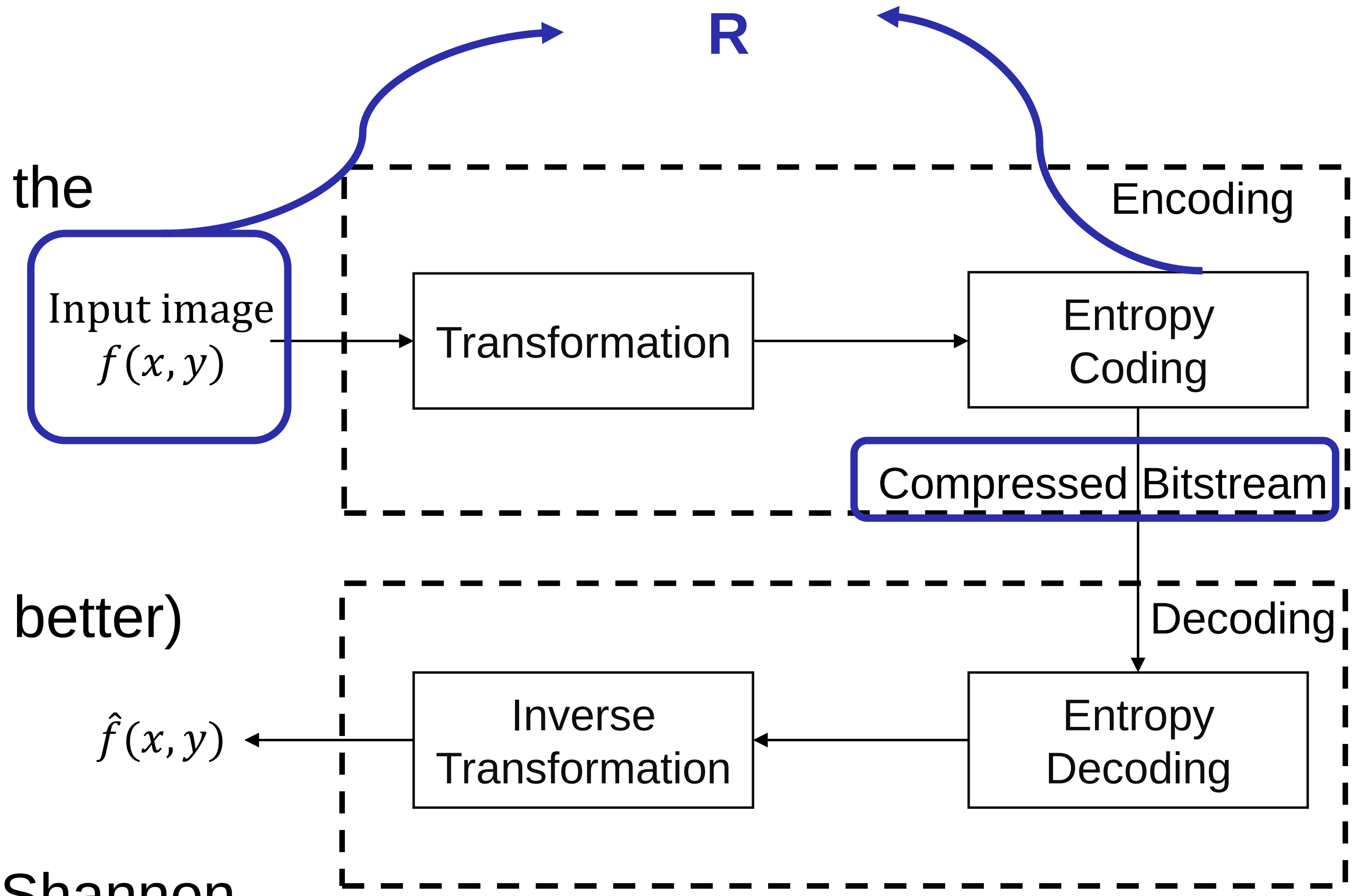
- Compression ratio (R, the higher the better)

$$\frac{Total\ size\ in\ bits\ of\ encoder\ input}{Total\ size\ in\ bits\ of\ encoder\ output}$$

- Bits per pixel (bpp, the lower the better)

$$\frac{Total\ size\ in\ bits\ of\ encoder\ output}{Total\ size\ in\ pixels\ of\ encoder\ input}$$

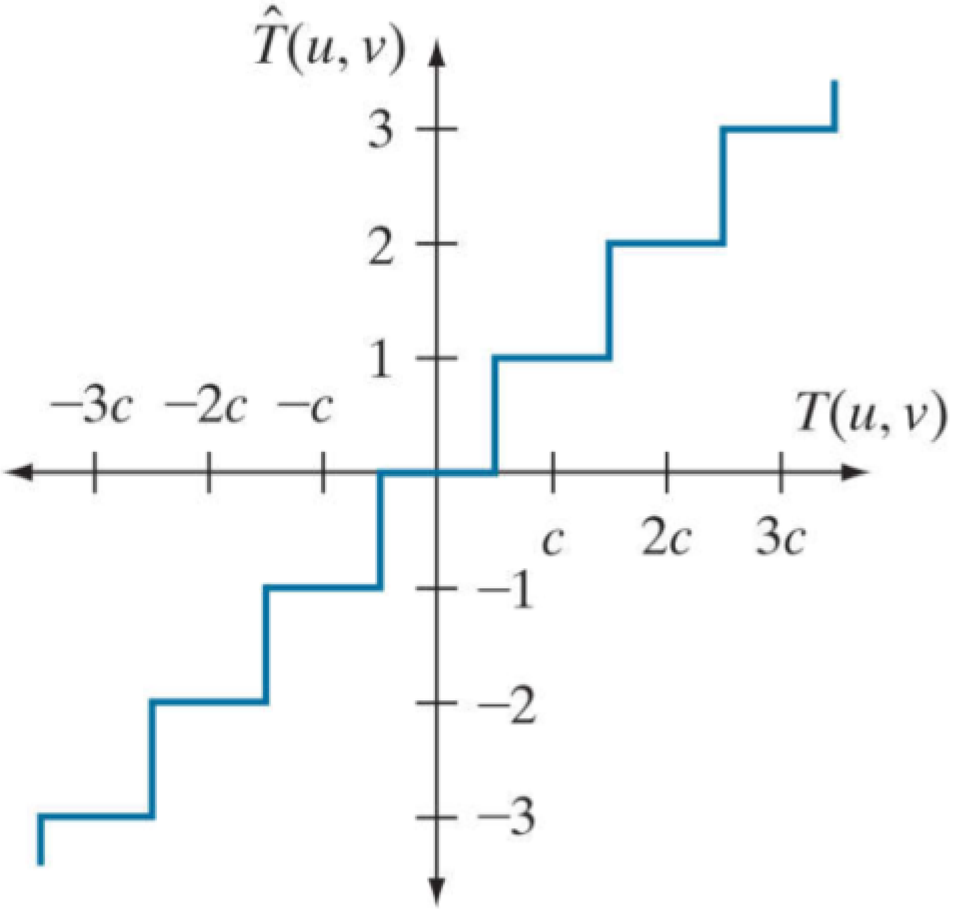
- A compression factor of 1.5-3 – Shannon Theory

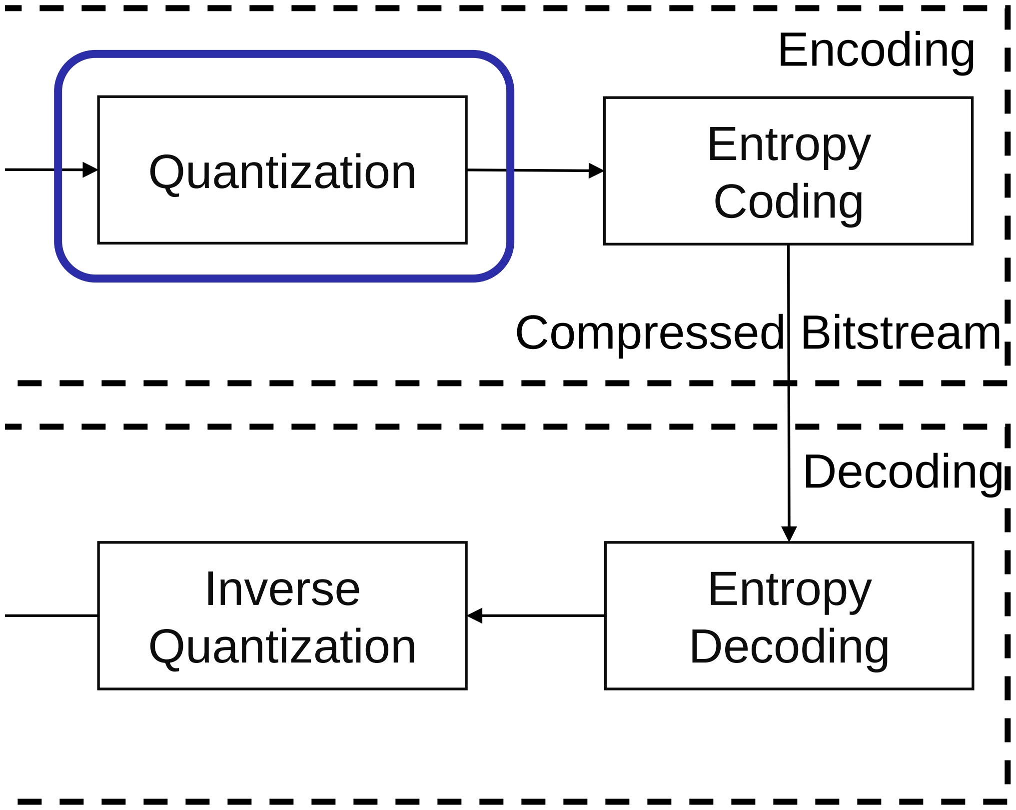


A brief transformation based lossless coding system

Lossy Image Compression

- Compressing a range of values to a single scalar value

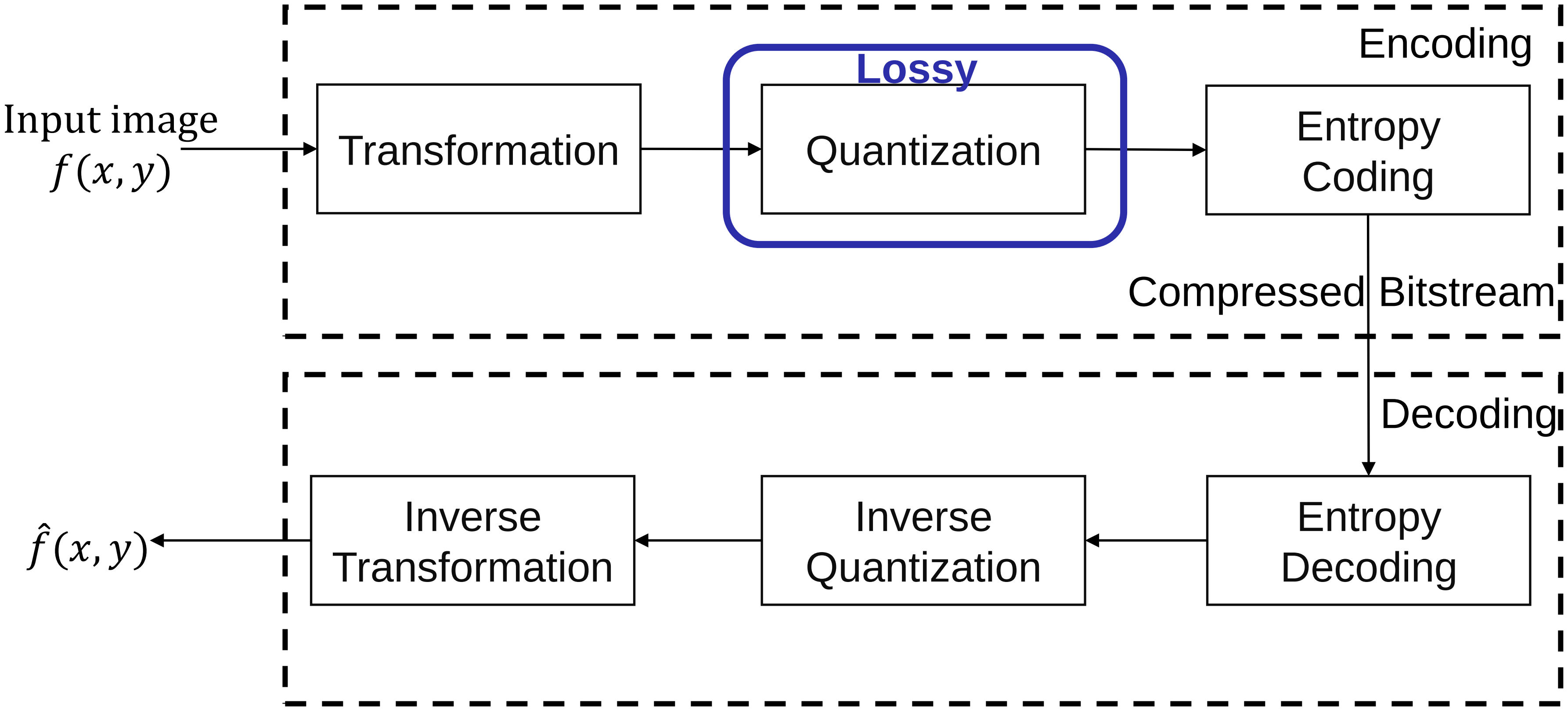




A brief lossy coding system

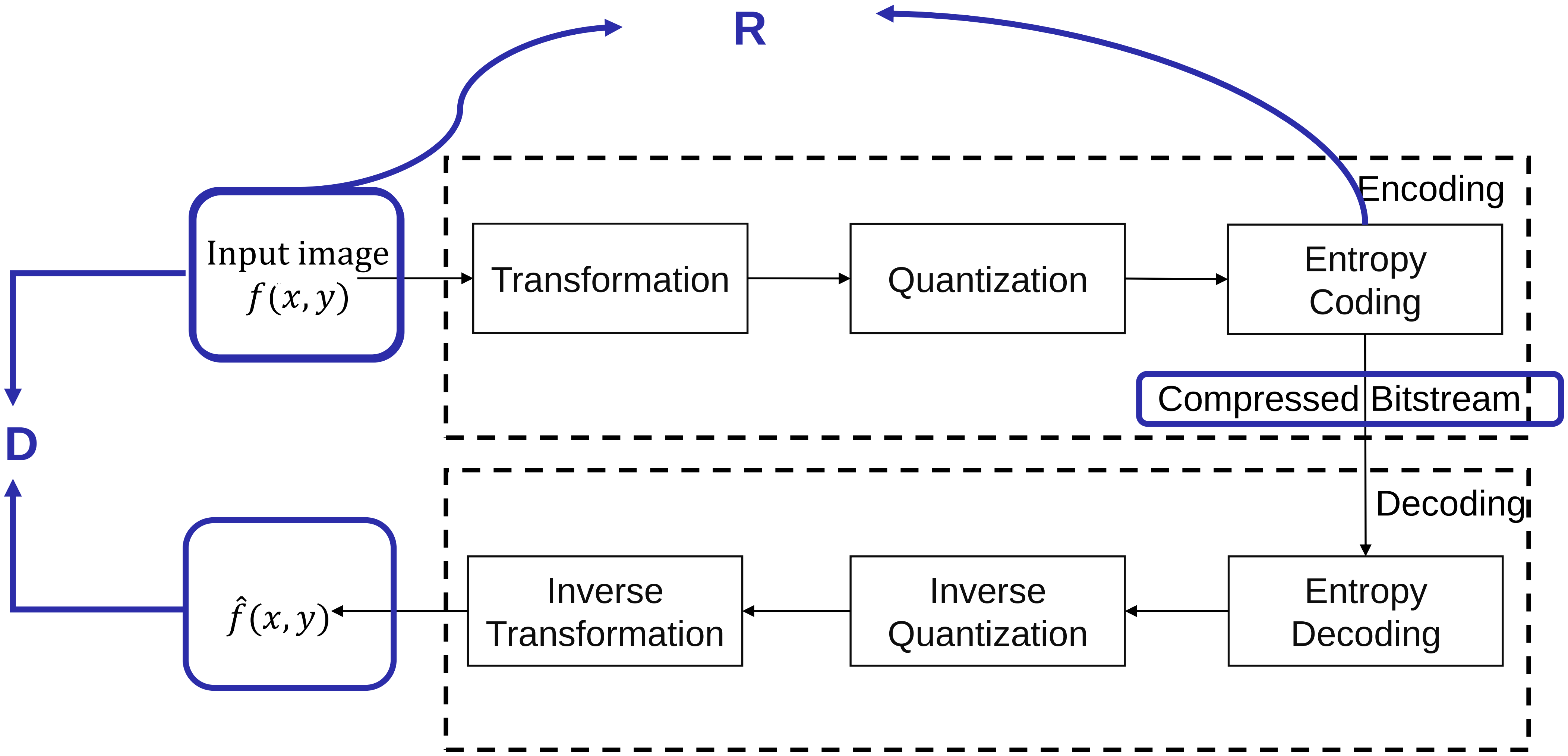
Lossy Image Compression

- Compressing a range of values to a single scalar value



A brief lossy coding system

Lossy Image Compression



A brief lossy coding system

Lossy Image Compression

Compression efficiency

- Compression ratio **R** (the higher the better)

$$\frac{Total\ size\ in\ bits\ of\ encoder\ input}{Total\ size\ in\ bits\ of\ encoder\ output}$$

- Bits per pixel (bpp, the lower the better)

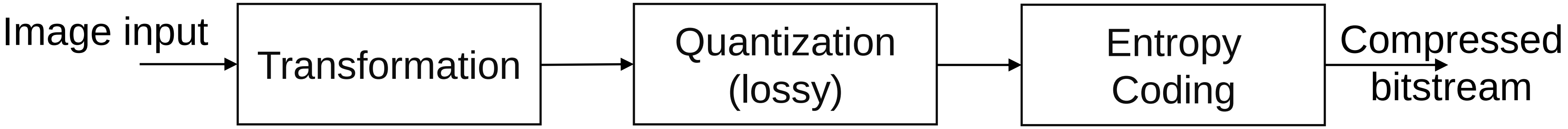
$$\frac{Total\ size\ in\ bits\ of\ encoder\ output}{Total\ size\ in\ pixels\ of\ encoder\ input}$$

- Loss/**D**istortion
  - Mean Square Error (MSE)
  - Peak Signal to Noise Ratio (PSNR)
  - Structural Similarity (SSIM)
  - Multi-scale SSIM (MS-SSIM)
  - etc.

Compression Efficiency + Reconstruction Quality

Rate-Distortion tradeoffs

Deep Image Compression



- Piecemeal Approaches:
  - Learned Transforms
  - Differentiable Quantization
  - Specialized Entropy Models
  - Deep models combined with classical methods
- End to End Approaches
  - ‘Deepen’ the traditional image coding schemes
  - New image coding framework / deep scheme

Traditional framework

Deep Image Compression

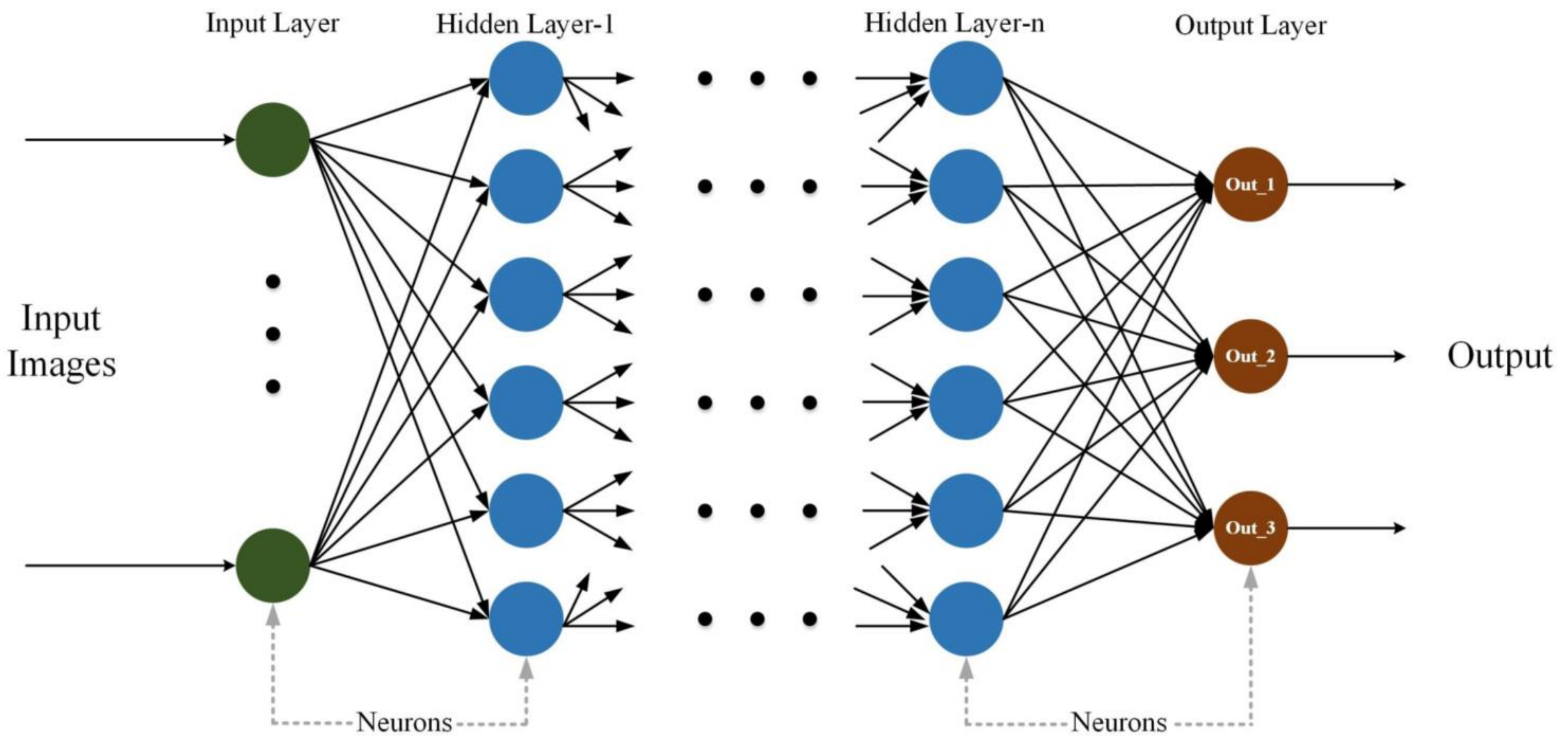
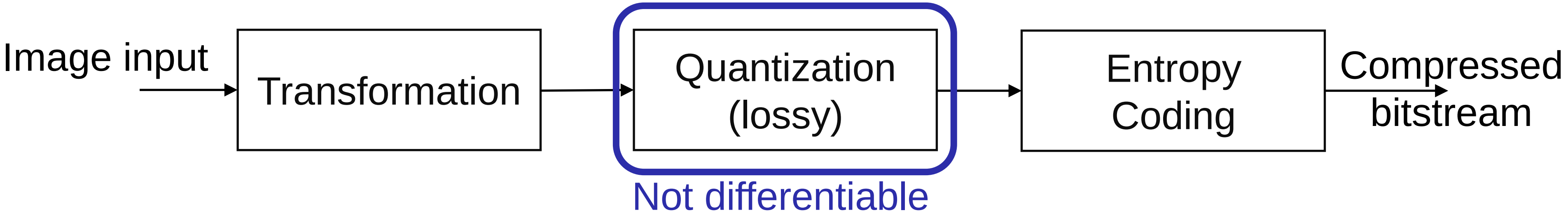


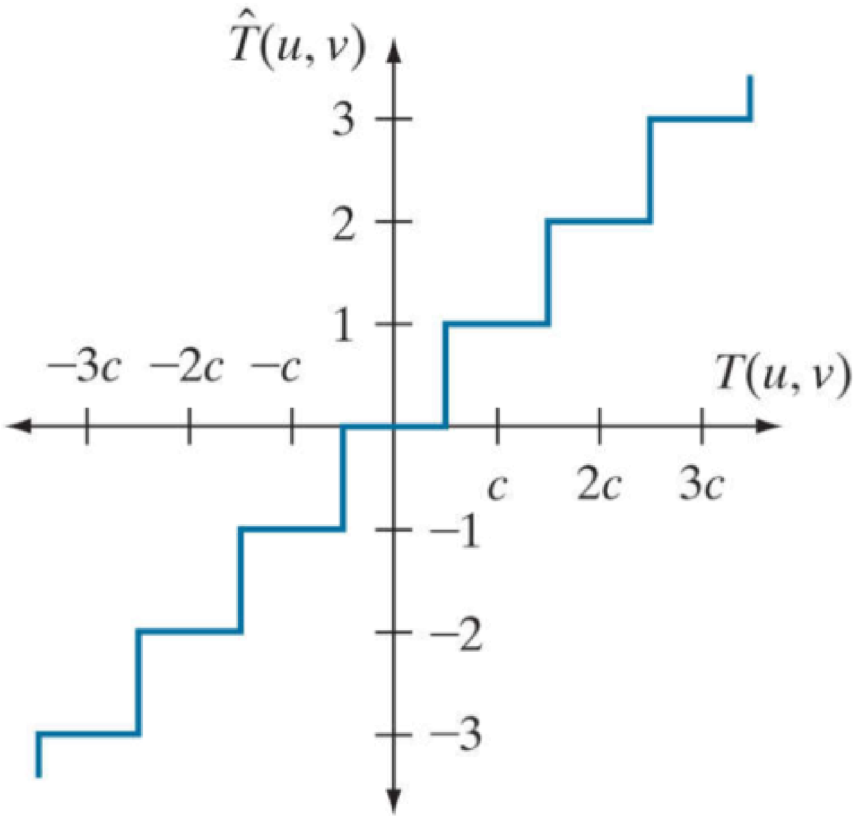
Illustration of a neural network architecture

Ma et al., Image and Video Compression with Neural Networks: A Review, 2018

Deep Image Compression: Differentiable Quantization

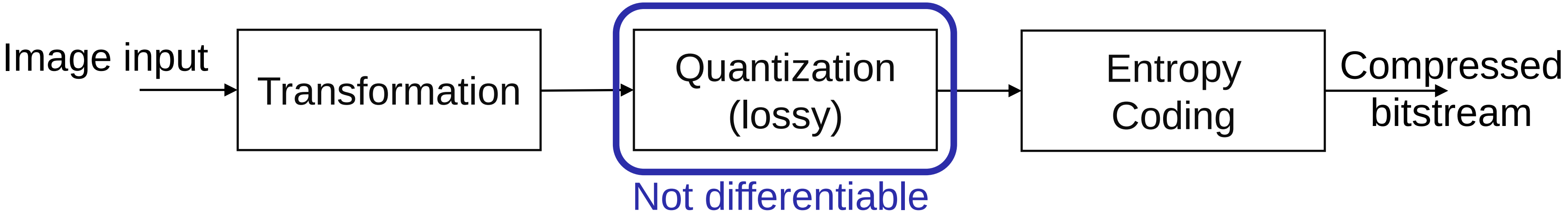


- Quantization is implemented by using a round function, and its derivative is almost zero except at the integers.



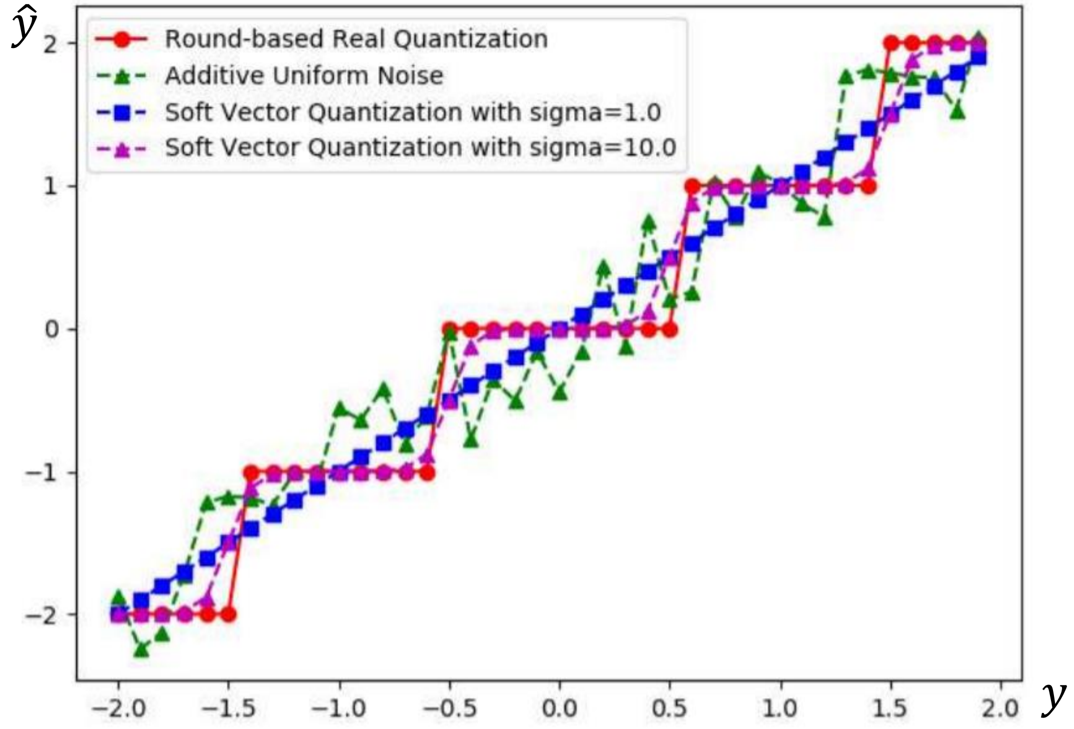
Gonzalez et al., Digital Image Processing

Deep Image Compression: Differentiable Quantization



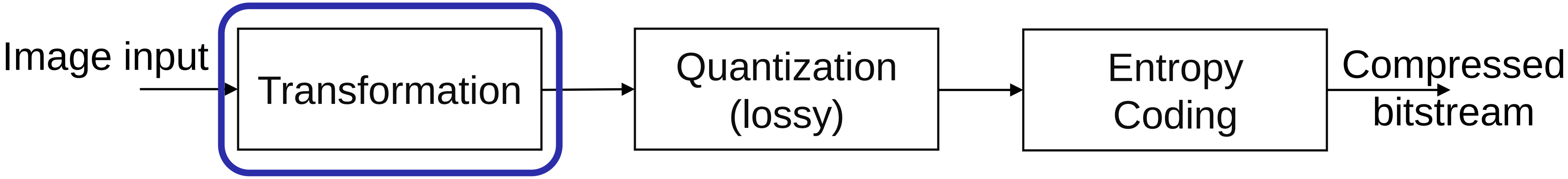
- Quantization is implemented by using a round function, and its derivative is almost zero except at the integers.
  - additive uniform noise
  - Soft-to-hard vector quantization
  - etc.

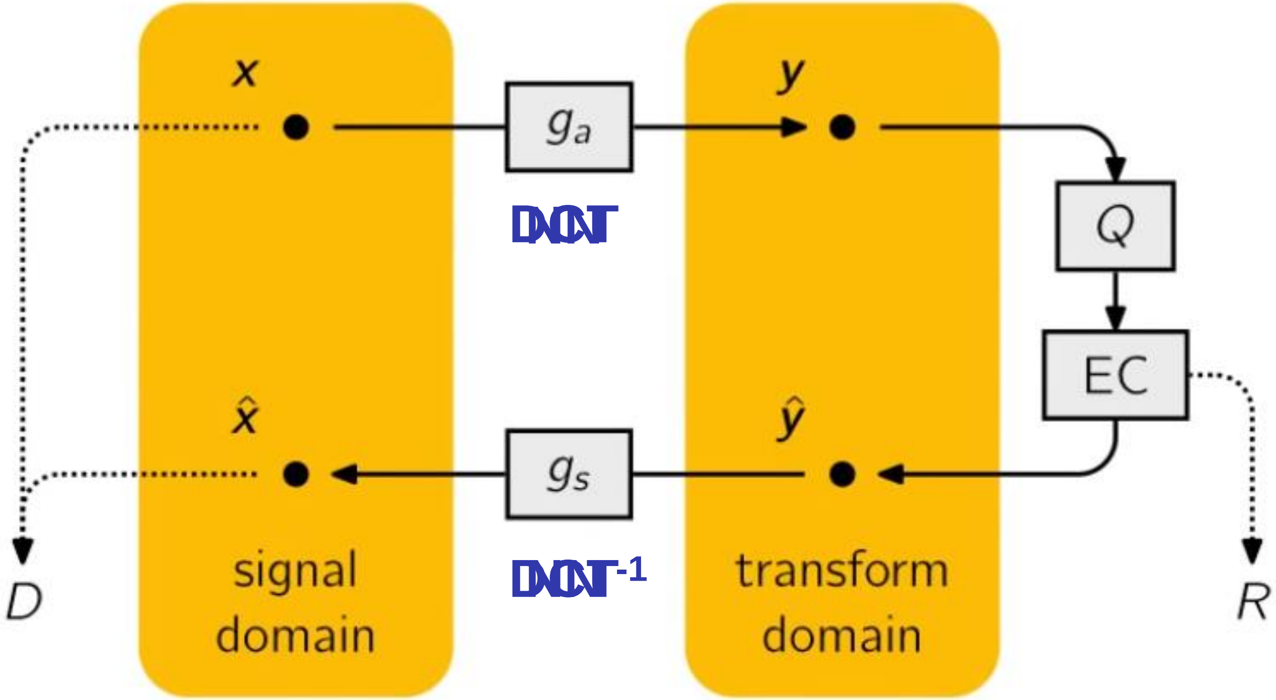
Performance with different quantization methods.

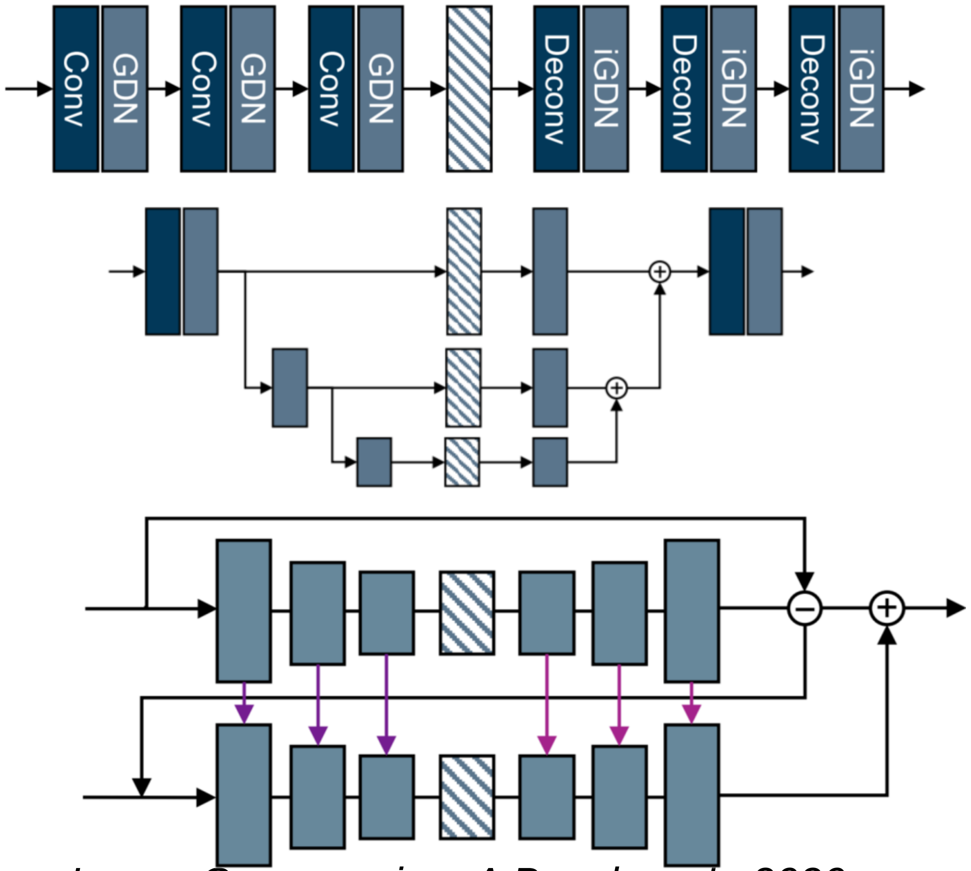


Cheng et al., Learning Image and Video Compression through Spatial-Temporal Energy, 2018, CVPR Compaction

Deep Image Compression: Learned Transforms

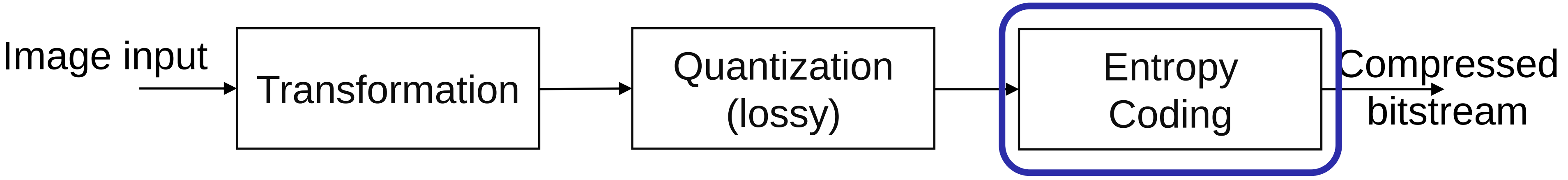


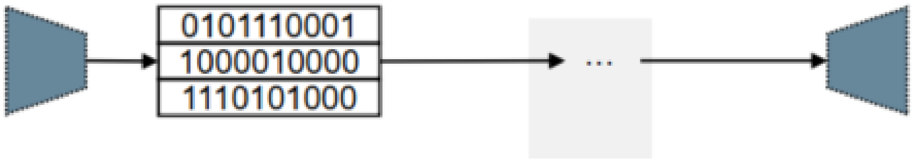




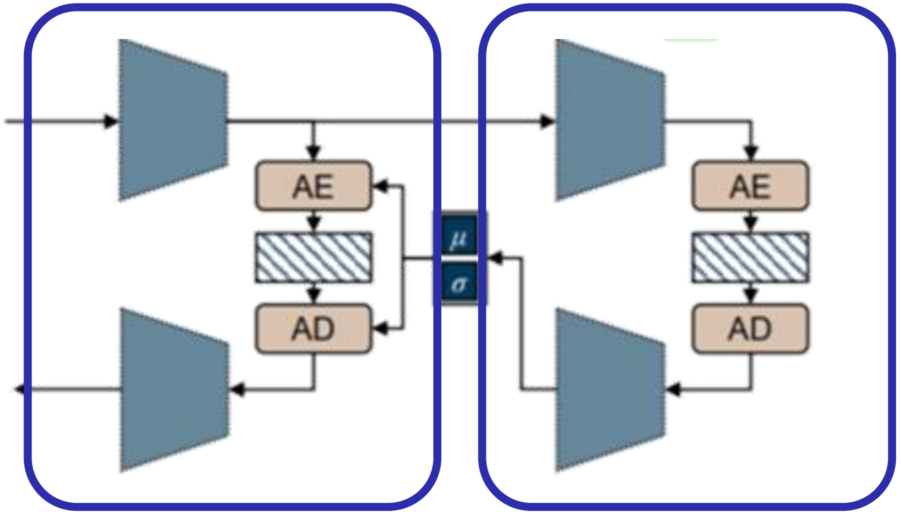
Hu et al., Learning End-to-End Lossy Image Compression: A Benchmark, 2020
Balle, PCS 2018 – Learned Image Compression

Deep Image Compression: Entropy Coding



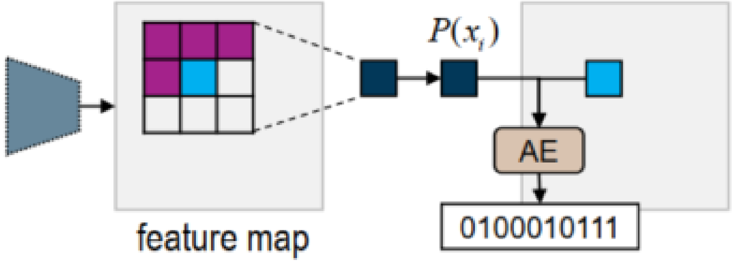


Direct modeling



side information

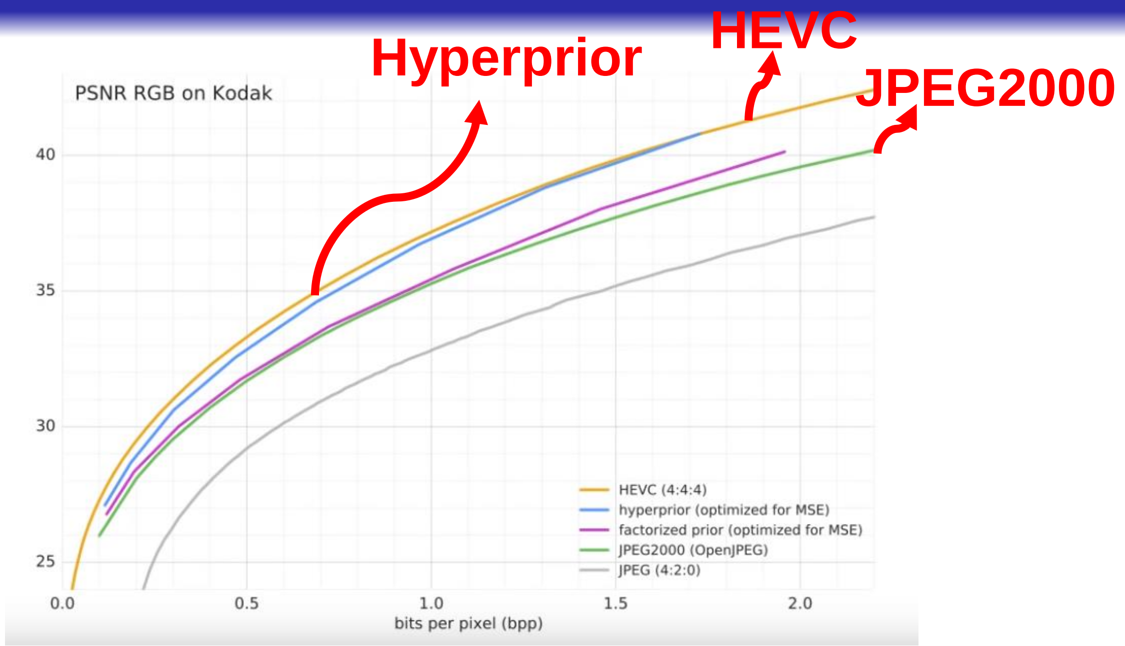
Hyperprior entropy model

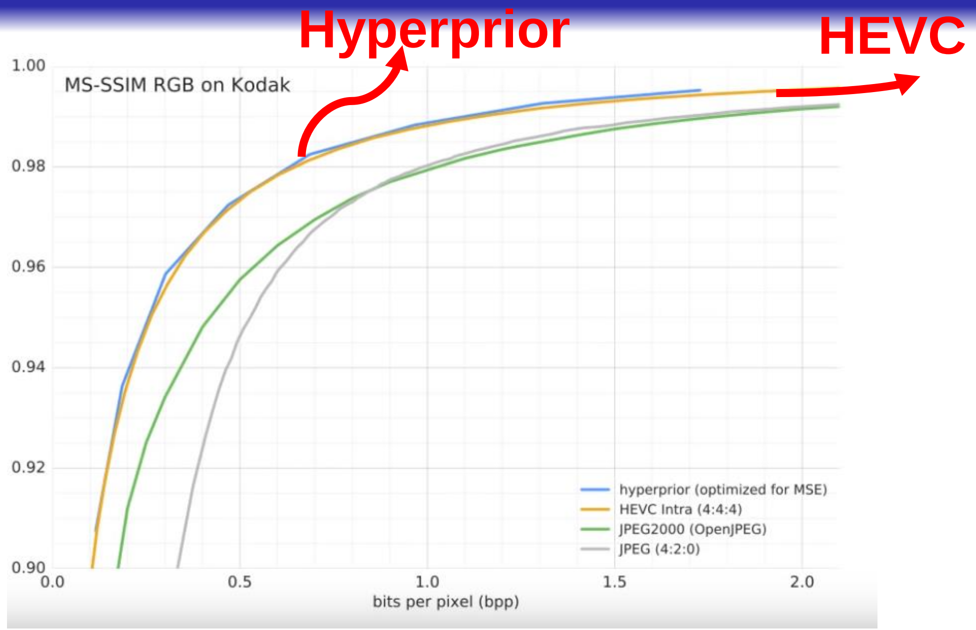


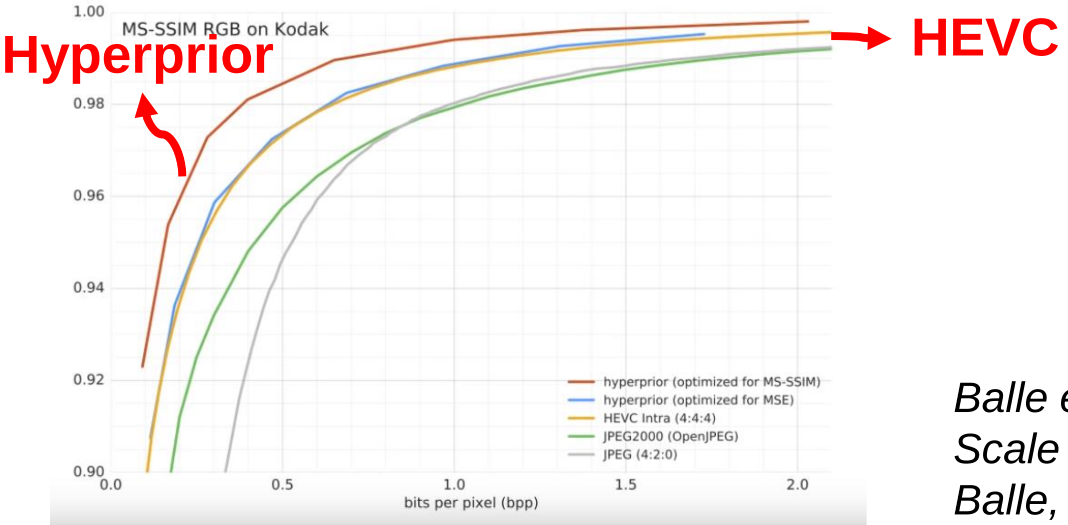
Spatial context model for latent code maps

Hu et al., Learning End-to-End Lossy Image Compression: A Benchmark, 2020
Balle et al., Variational Image Compression with a Scale Hyperprior, 2018, ICLR

Deep Image Compression: Entropy Coding

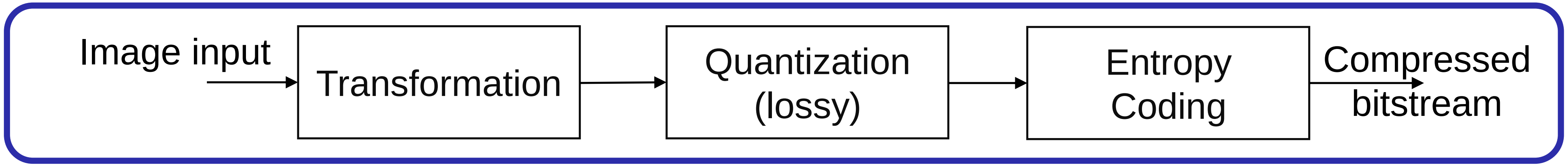


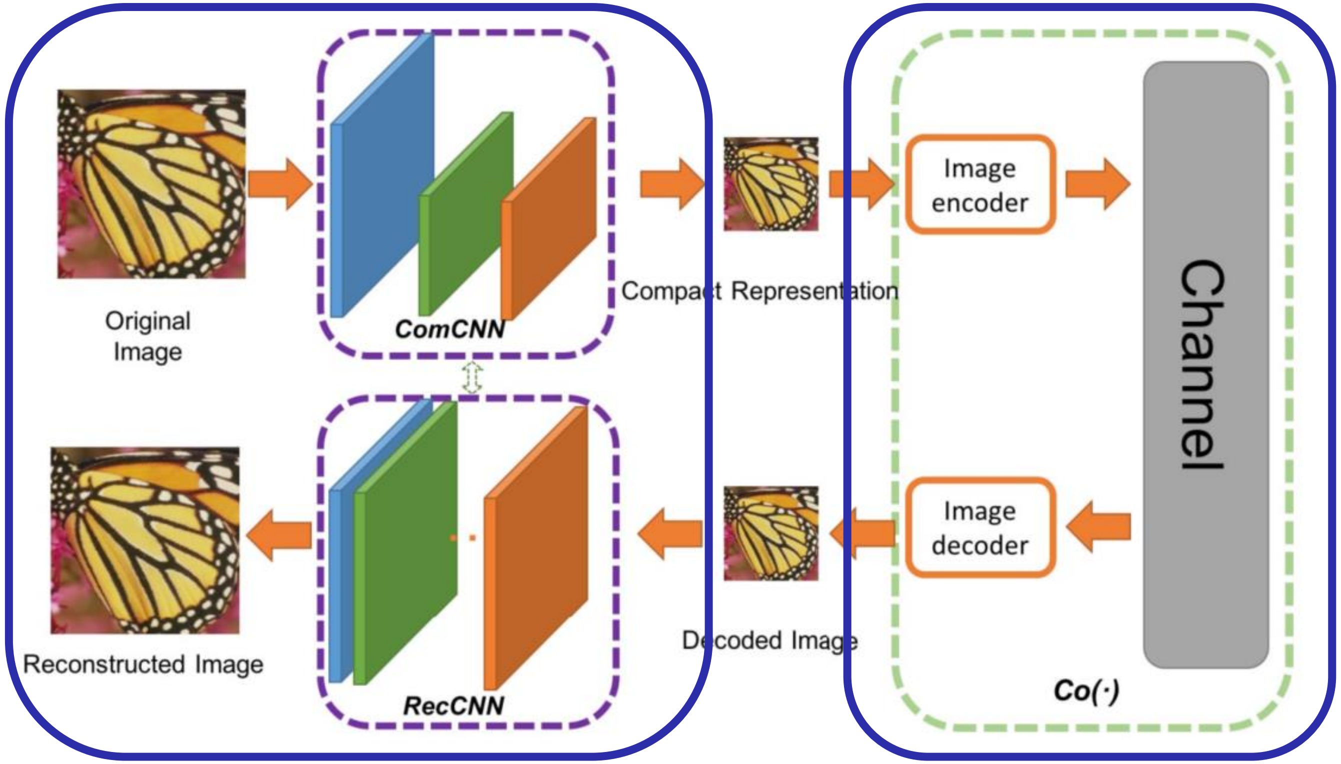




Balle et al., Variational Image Compression with a Scale Hyperprior, ICLR 2018
Balle, PCS 2018 – Learned Image Compression

Deep Image Compression: Deep Models Combined with Classical Methods





- Compatible with Traditional image codes: JPEG, JPEG2000, or BPG
- Outperform JPEG,JPEG2000
- Slightly better or comparable to original BPG

Jiang et al. An End-to-End Compression Framework Based on Convolutional Neural Networks, 2018

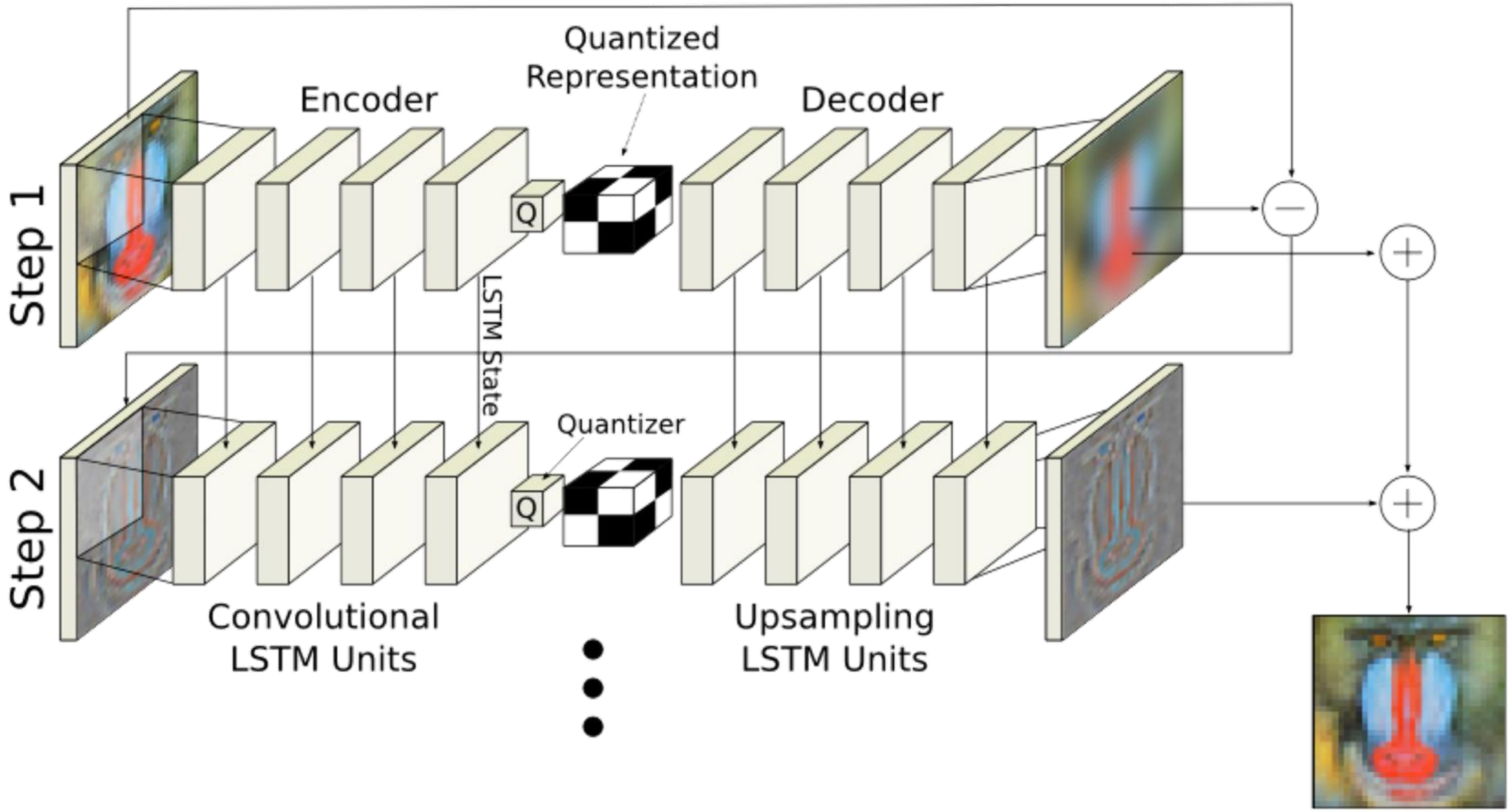
Deep Image Compression

Image input → Transformation → Quantization (lossy) → Entropy Coding → Compressed bitstream

- Piecemeal Approaches:
  - Learned Transforms
  - Differentiable Quantization
  - Specialized Entropy Models
  - Deep models combined with classical methods
- End to End Approaches
  - ‘Deepen’ the traditional image coding schemes
  - New image coding framework / deep scheme

Traditional framework

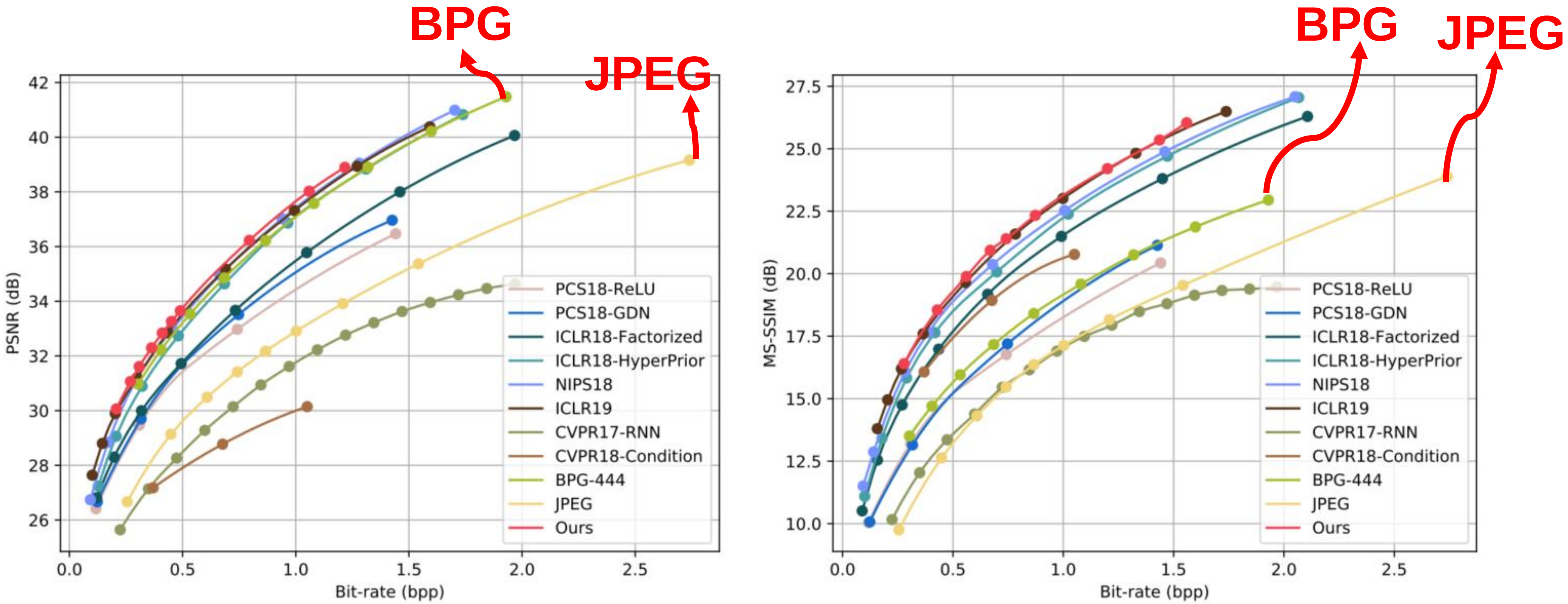
Deep Image Compression: Deep Scheme



- First “feasible” neural compression method
- Ambitious (transform coding & replace entropy coding)
- Computationally intensive
- Better than JPEG, but not competitive with H.265

Toderici et al., Variable rate image compression with recurrent neural networks, 2016, ICLR
PCS 2019, Toderici, Neural Image Compression: Recent Developments and Opportunities, Keynote

Deep Image Compression: Performance



(a) Kodak, PSNR

(b) Kodak, MS-SSIM

- State of the art performance
- Not superior compared to HEVC in terms of PSNR
- High computational cost

Hu et al., Learning End-to-End Lossy Image Compression: A Benchmark, 2020

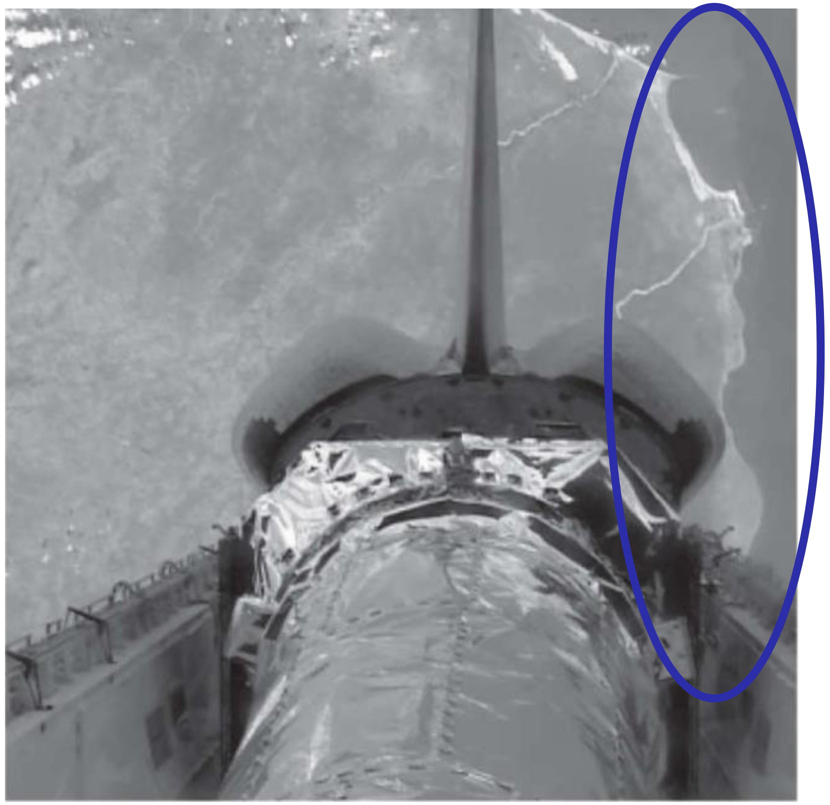
Outline

- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

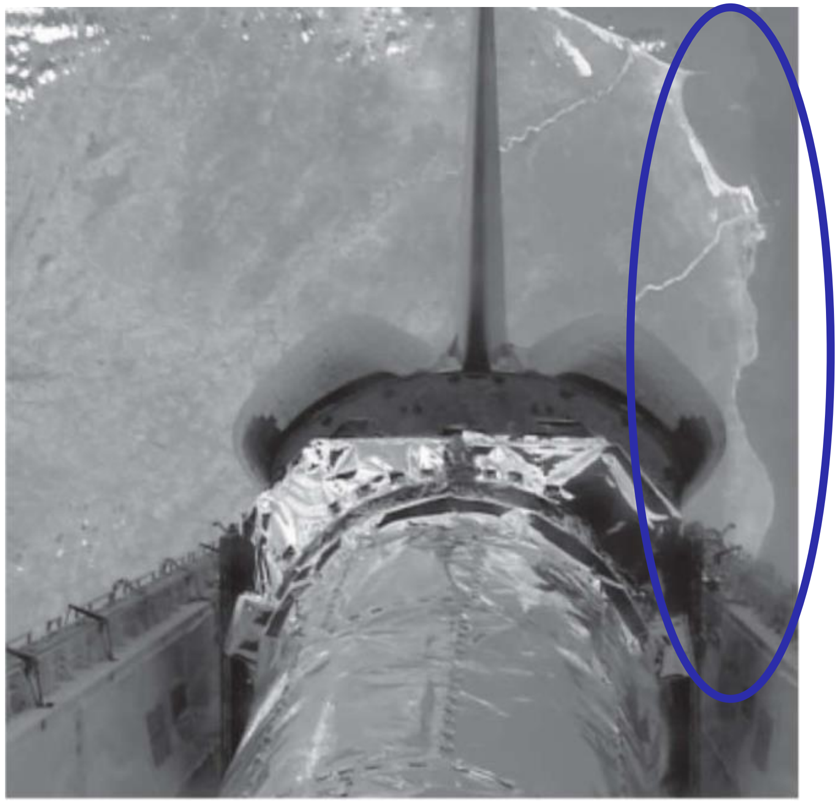
Video Compression

- Temporal Redundancy
  - Take advantage of similarity between successive frames

Frame 0



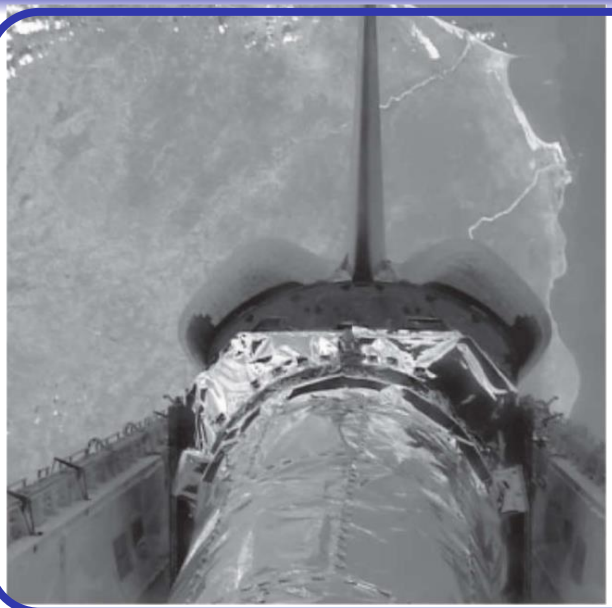
Frame 1



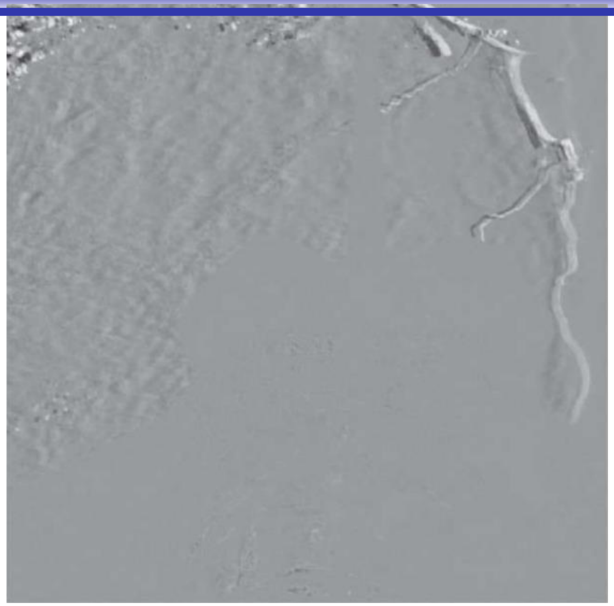
Gonzalez et al., Digital Image Processing

Video Compression

Frame 0



Frame 1

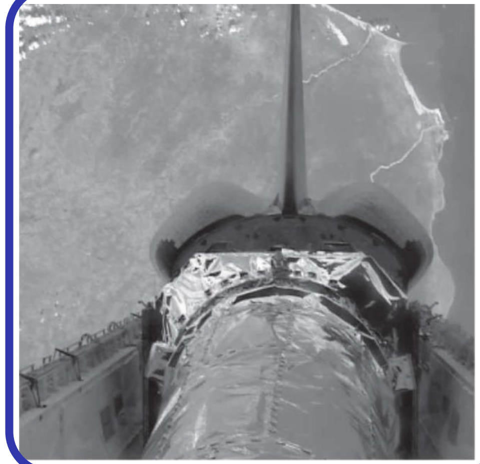


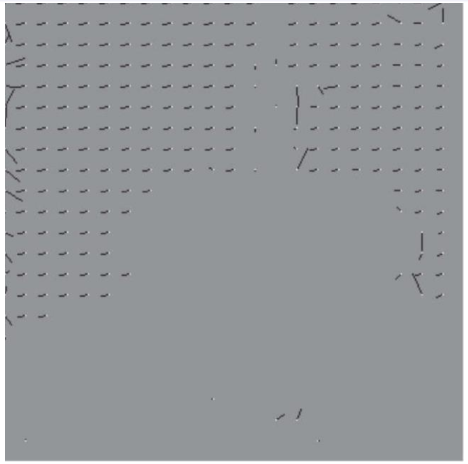
Frame difference of frame 0 and frame 1c

- The compression ratio can be 1.92

Video Compression

Frame 0

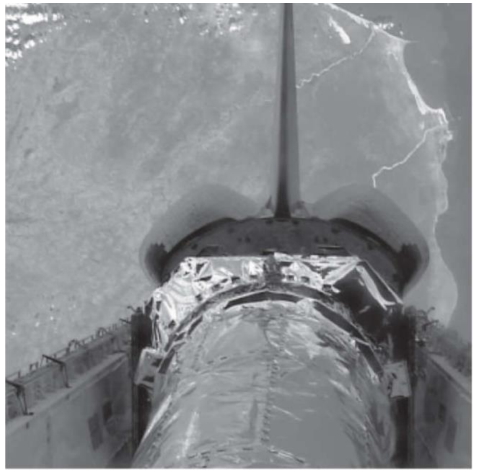


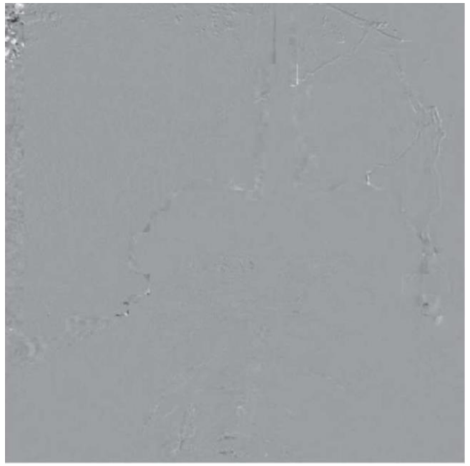


Motion Vector between Frame 0 and 1

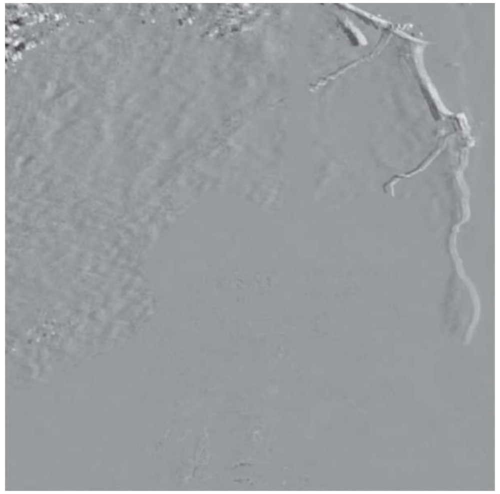
- The compression ratio can be 2.63

Frame 1



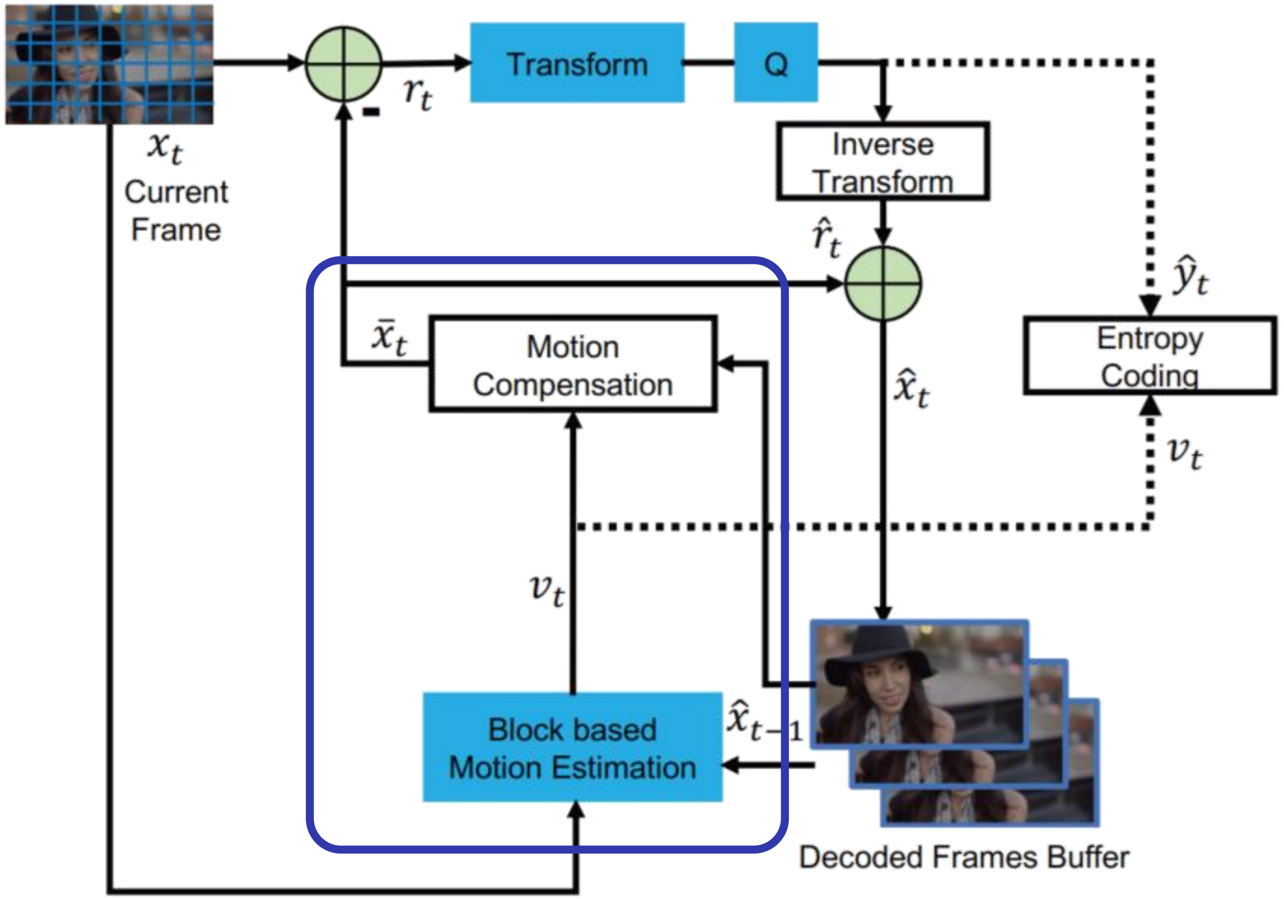


Frame difference after motion compensation



Original Frame difference

Video Compression



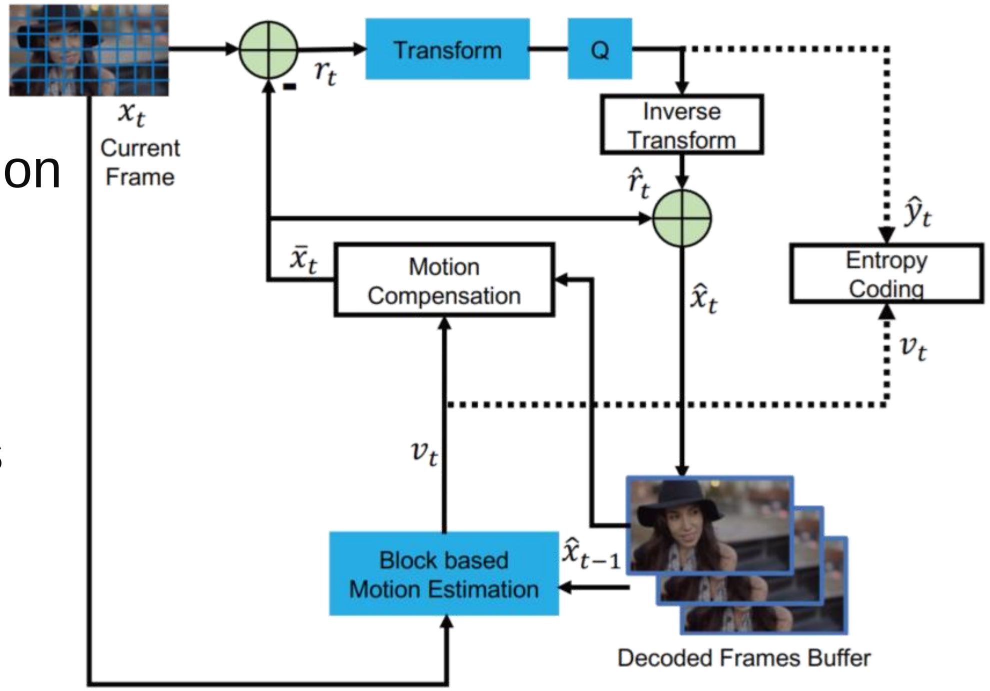
Traditional Video Compression Framework

- Motion estimation
- Motion compensation
- Transform
- Quantization
- Entropy coding
- Inverse transform
- Frame reconstruction

Luo et al., DVC: An End-to-end Deep Video Compression Framework, 2019, CVPR

Deep Video Compression

- Piecemeal Approaches:
  - Learned Motion estimation, motion compensation
  - Learned Transforms
  - Differentiable Quantization
  - Specialized Entropy Models
  - Deep models combined with classical methods
  - etc.



- End to End Approaches:
  - 'Deepen' the traditional video coding schemes
  - New video coding framework / deep scheme

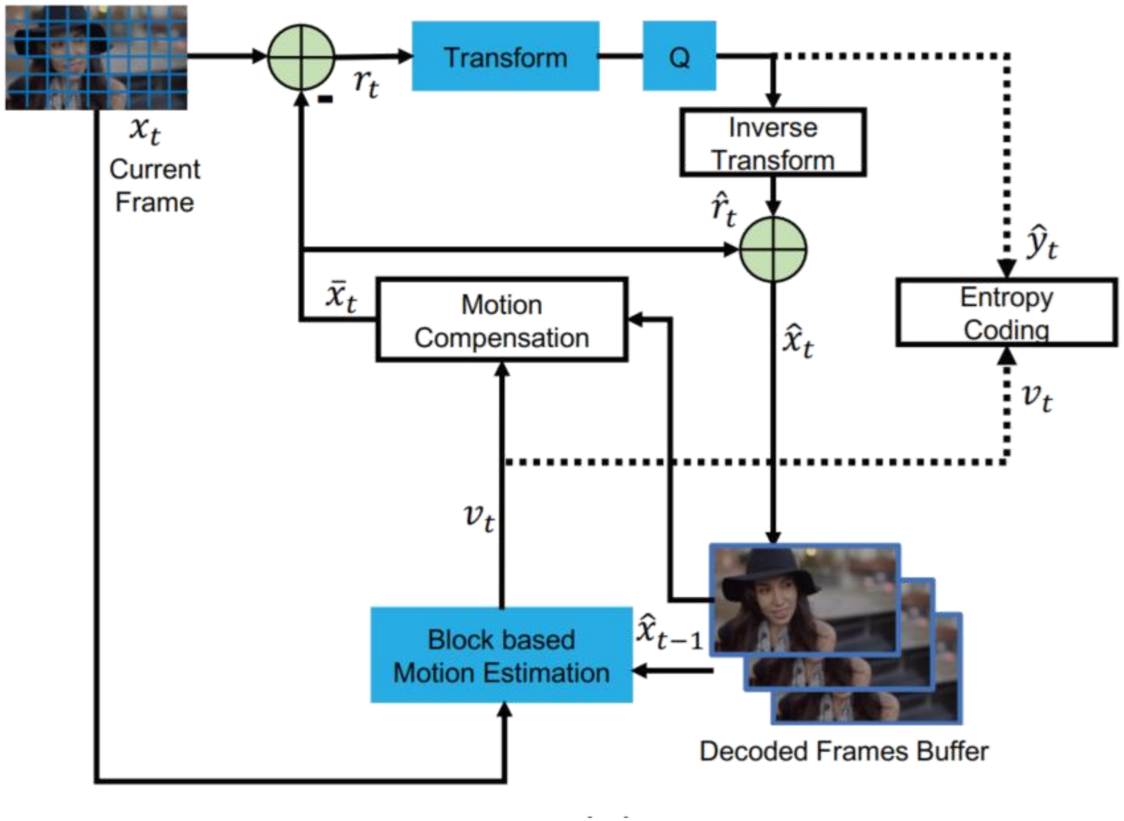
Deep Video Compression

- Piecemeal Approaches:
  - Learned Motion Estimation, Motion Compensation
  - Learned Transforms
  - Differentiable Quantization
  - Specialized Entropy Models
  - Deep models combined with classical methods
  - etc.

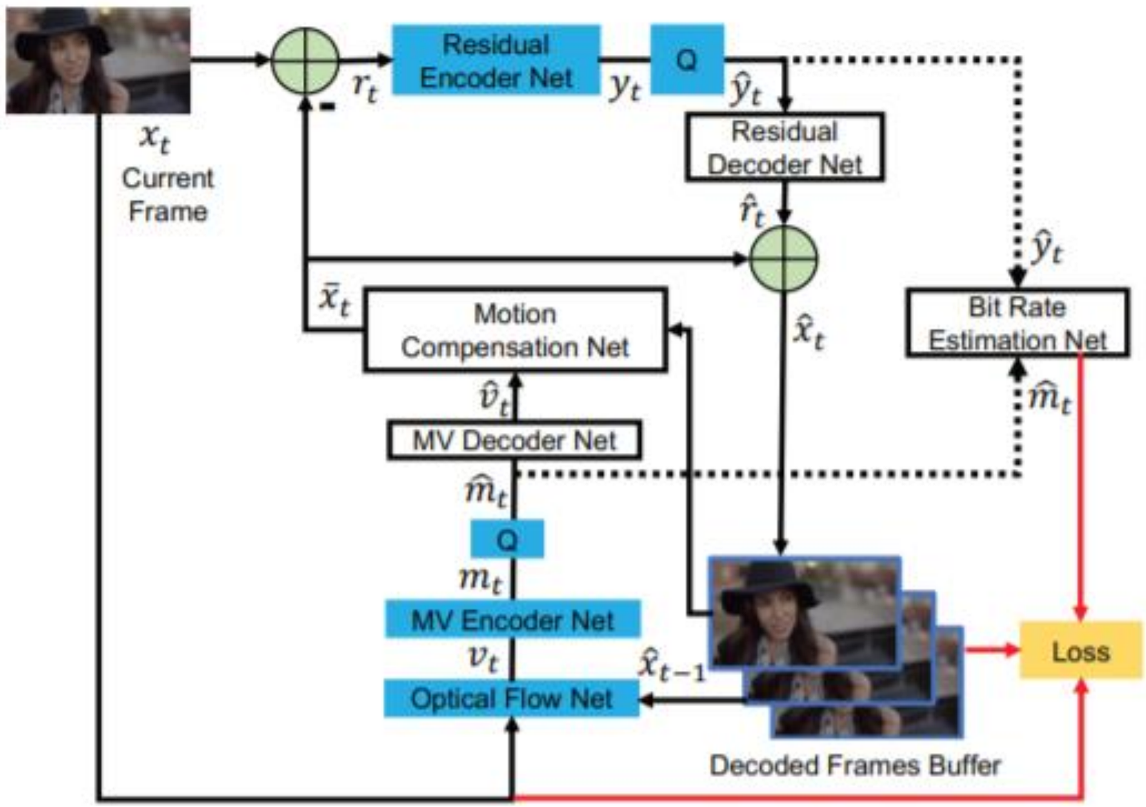
- End to End Approaches
  - ‘Deepen’ the traditional video coding schemes
  - New video coding framework / deep scheme

Deep Video Compression: ‘Deepen’

Traditional Video Compression Framework



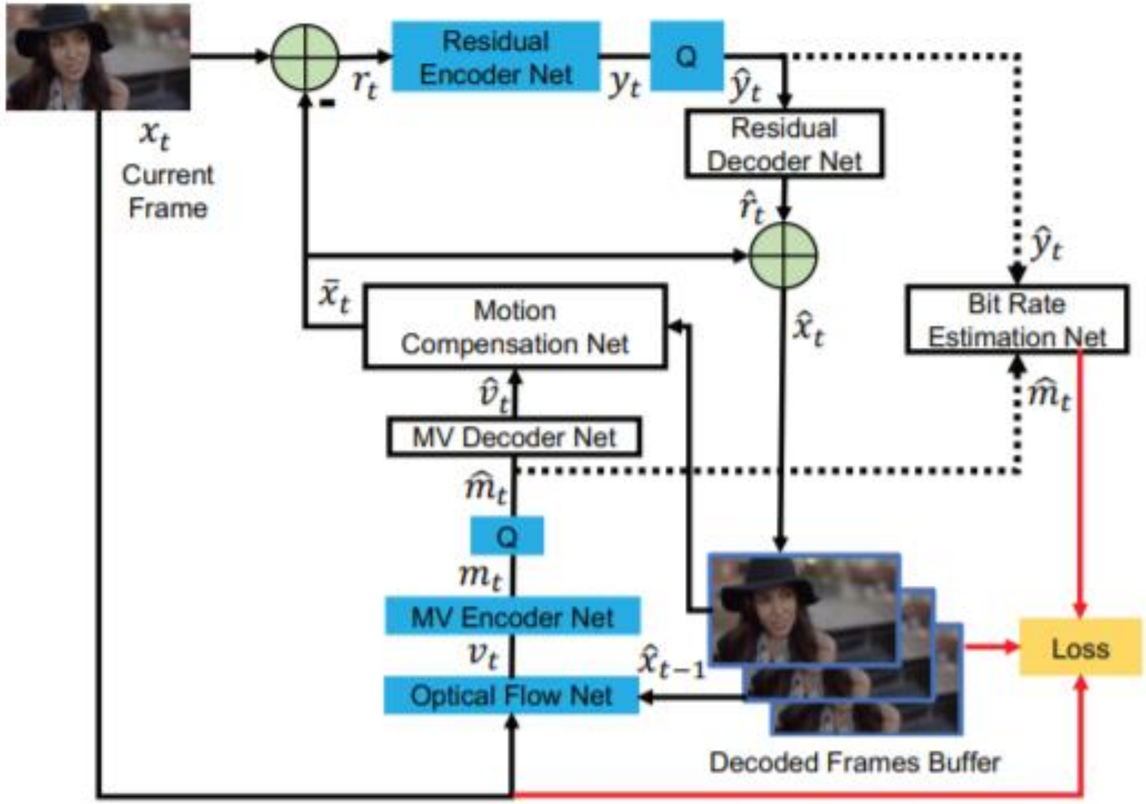
An end-to-end video compression network



Luo et al., DVC: An End-to-end Deep Video Compression Framework, 2019, CVPR

Deep Video Compression: 'Deepen'

An end-to-end video compression network



Loss Function: rate-distortion optimization

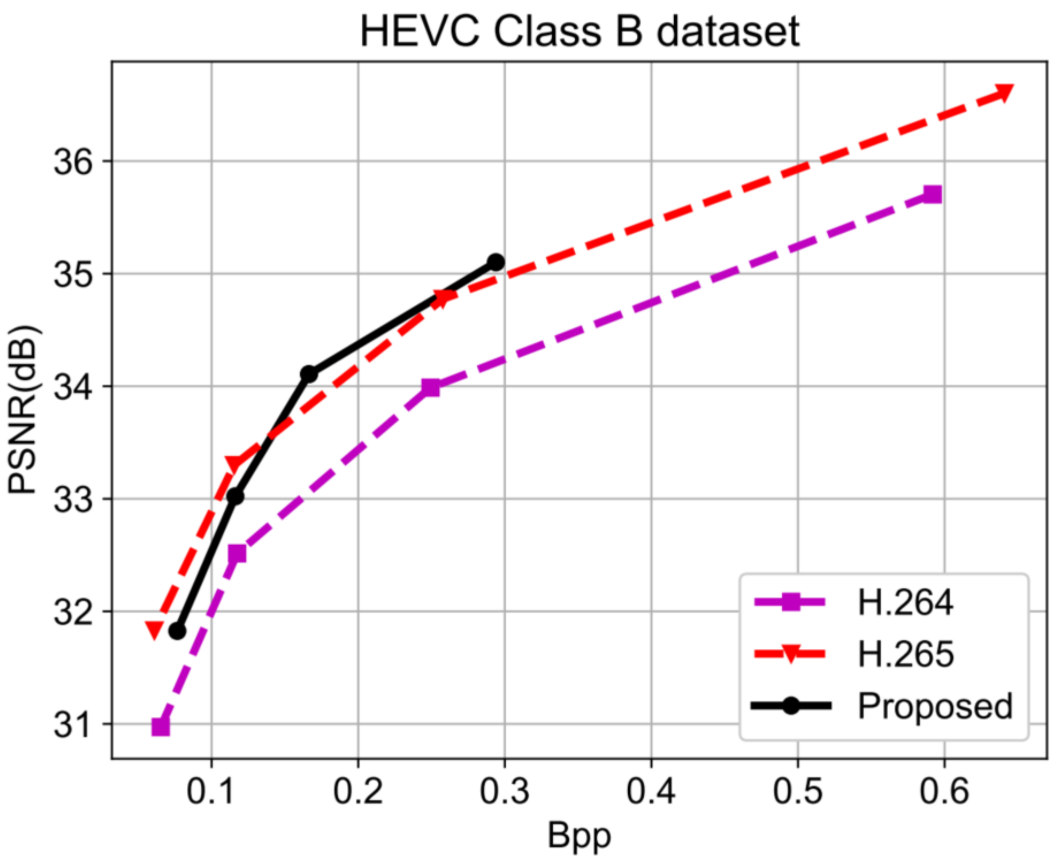
$$\lambda D + R = \lambda d(x_t, \hat{x}_t) + (H(\hat{m}_t) + H(\hat{y}_t))$$

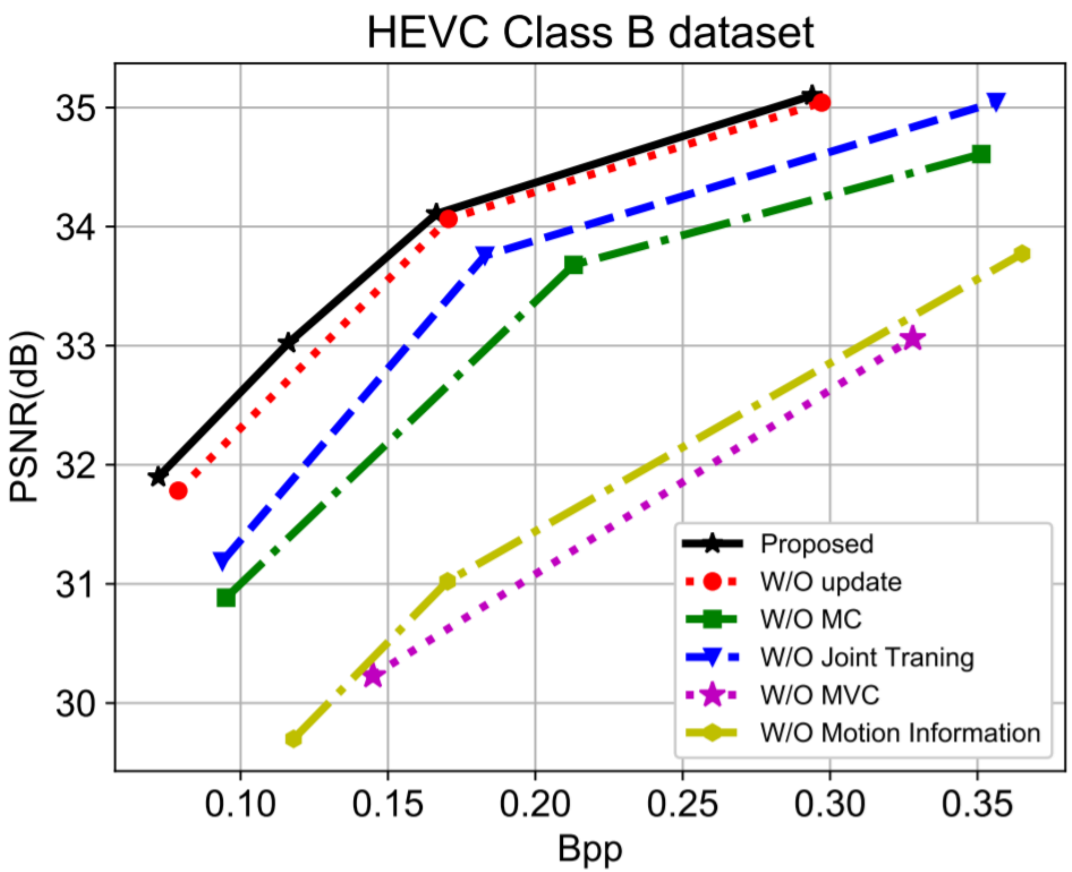
d: MSE error

H: the number of bits used for encoding the representations, including motion $\hat{m}_t$ and residual $\hat{y}_t$

Luo et al., DVC: An End-to-end Deep Video Compression Framework, 2019, CVPR

Deep Video Compression: ‘Deepen’





- Outperform H.264 in terms of PSNR
- Similar or better compared to H.265

- Joint training the ‘deepen’ framework achieves the best performance

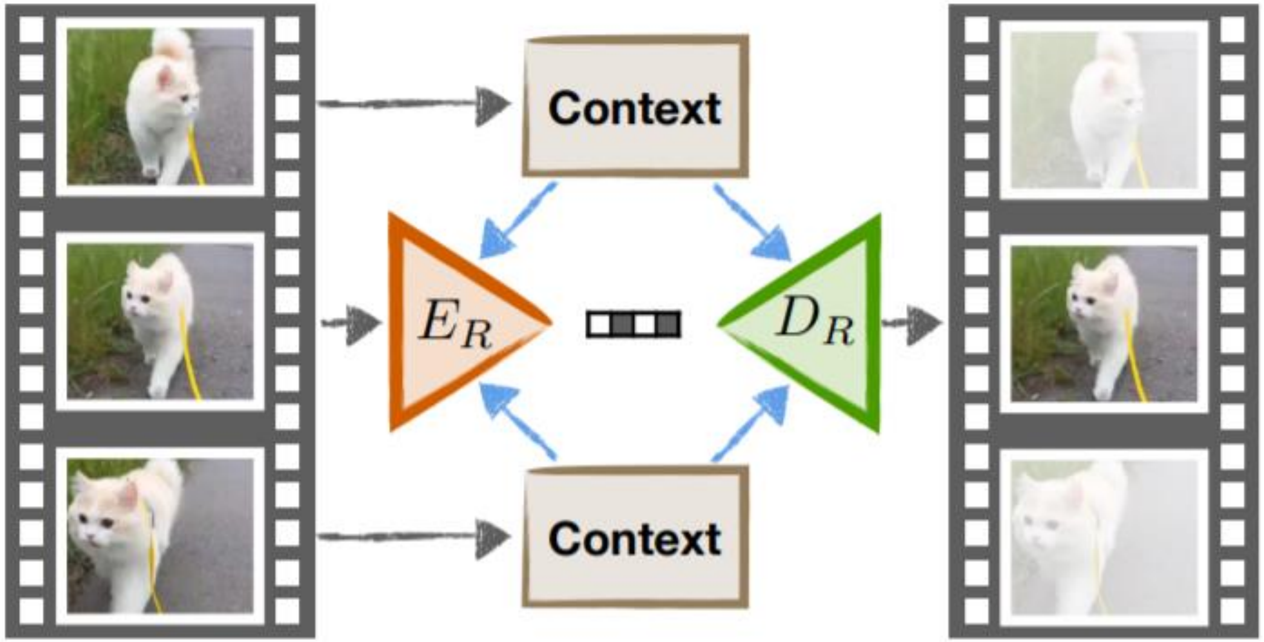
Luo et al., DVC: An End-to-end Deep Video Compression Framework, 2019, CVPR

Deep Video Compression: Deep Scheme

- Piecemeal Approaches:
  - Learned Motion estimation, motion compensation
  - Learned Transforms
  - Differentiable Quantization
  - Specialized Entropy Models
  - Deep models combined with classical methods
  - etc.

- End to End Approaches
  - ‘Deepen’ the traditional video coding schemes
  - New video coding framework / deep scheme

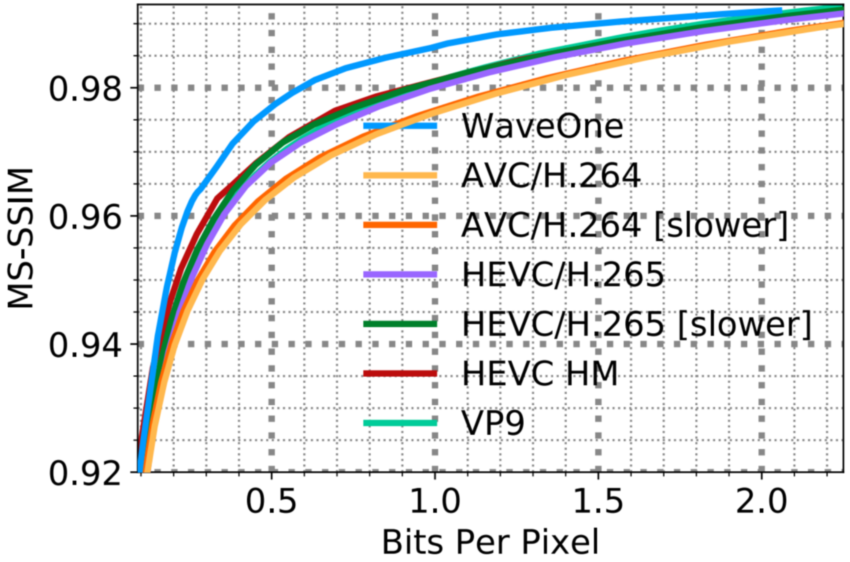
Deep Video Compression: Deep Scheme

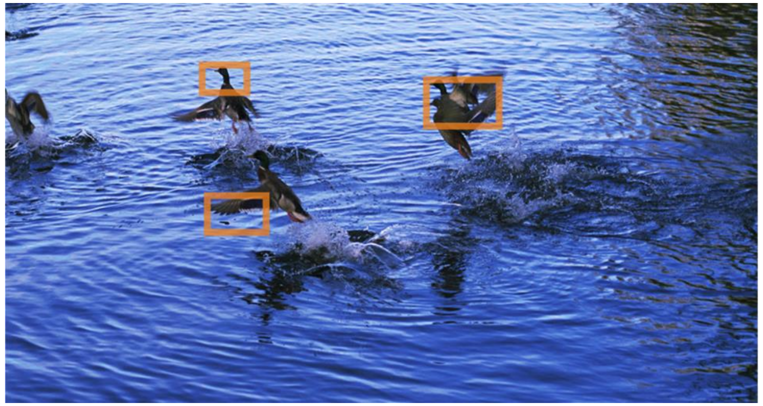


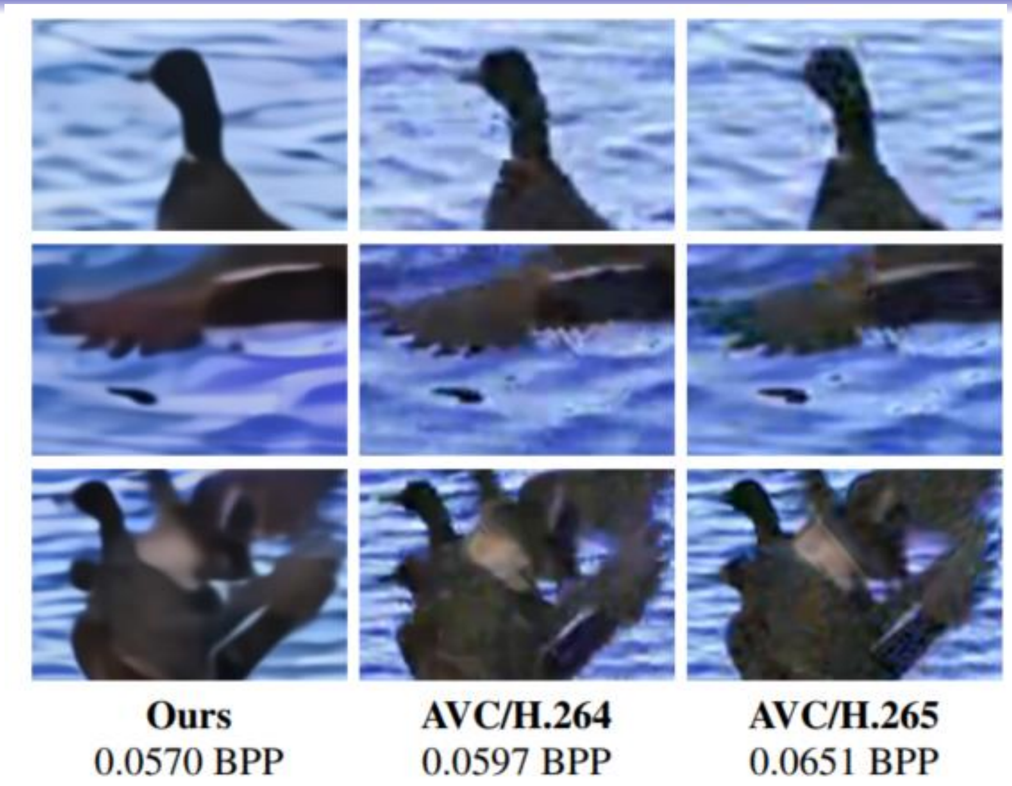
- Video compression -- Repeated image interpolation
- Encodes key frames via NN
- Reconstructs remaining frames by interpolating
- On par with H.264

Wu et al. Video Compression through Image Ianterpolation,2018, ECCV

Deep Video Compression: Performance







- State of the art performance (MS-SSIM)
- Not sufficient for real-time deployment

Rippel et al., Learned Video Compression, 2019, ICCV

Outline

- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

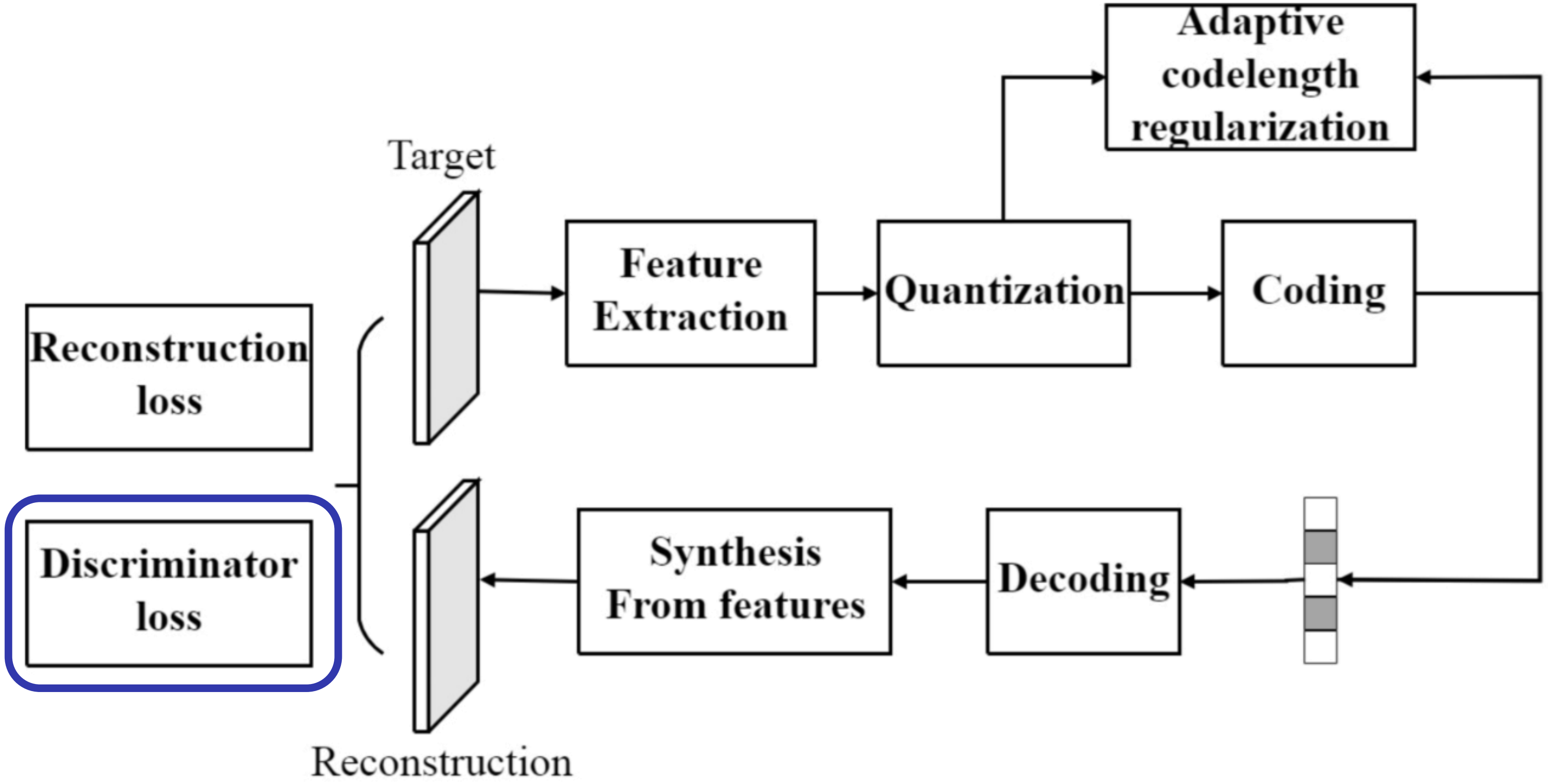
Special Purpose Coding

- Distortion: MSE, SSIM, PSNR, etc.
- Perceptual Naturalness of the reconstructed image/video
- Extreme Image Compression, e.g. targeting bitrates below 0.1 bpp

Generative modelling:

GAN (generative adversarial network), VAE (variational auto-encoder)

Special Purpose Coding



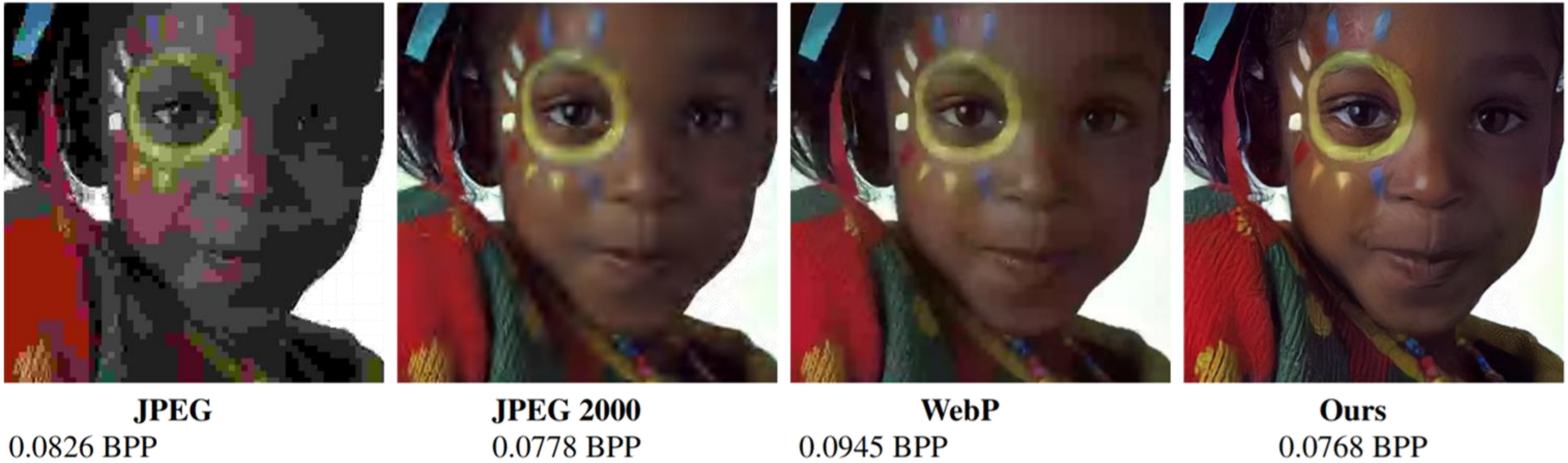
- Discriminator loss: encourages visually pleasing reconstructions

Rippel et al., Real-Time Adaptive Image Compression, 2017, ICML

Special Purpose Coding

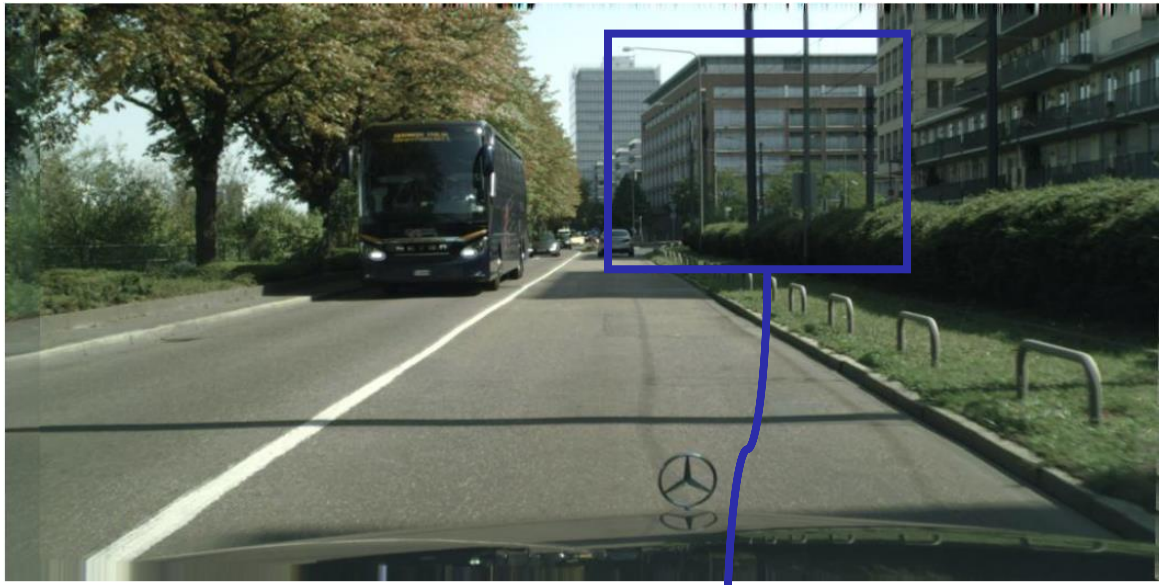
Original
24 BPP

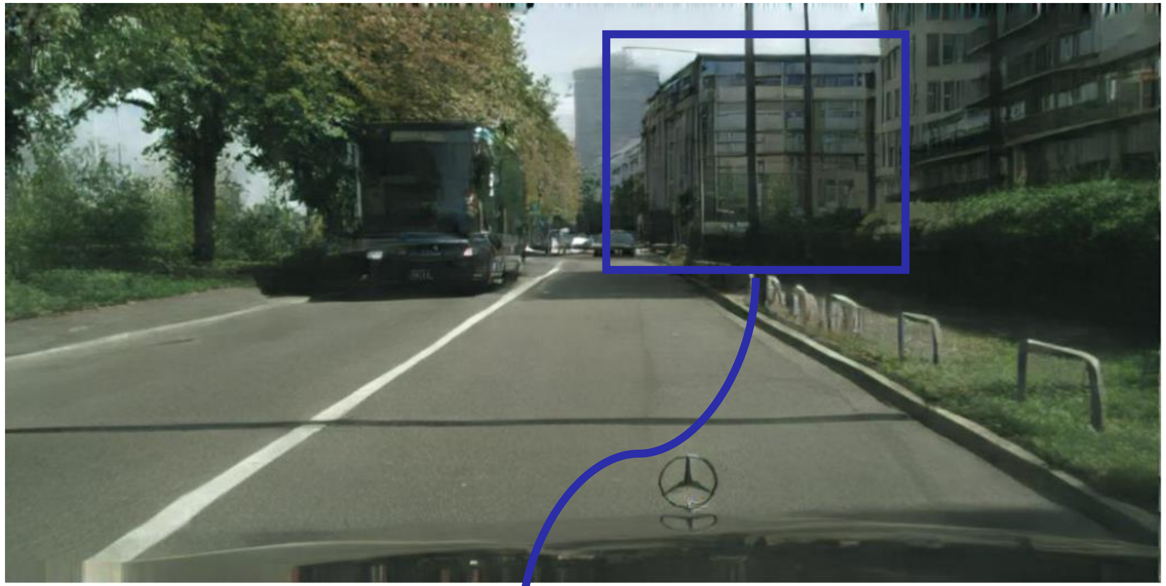
- Run in real-time
- Across different quality levels
  - 2.5 times smaller than JPEG, JPEG2000
  - 1.7 times smaller than BPG



Rippel et al., Real-Time Adaptive Image Compression, 2017 ICML

Special Purpose Coding



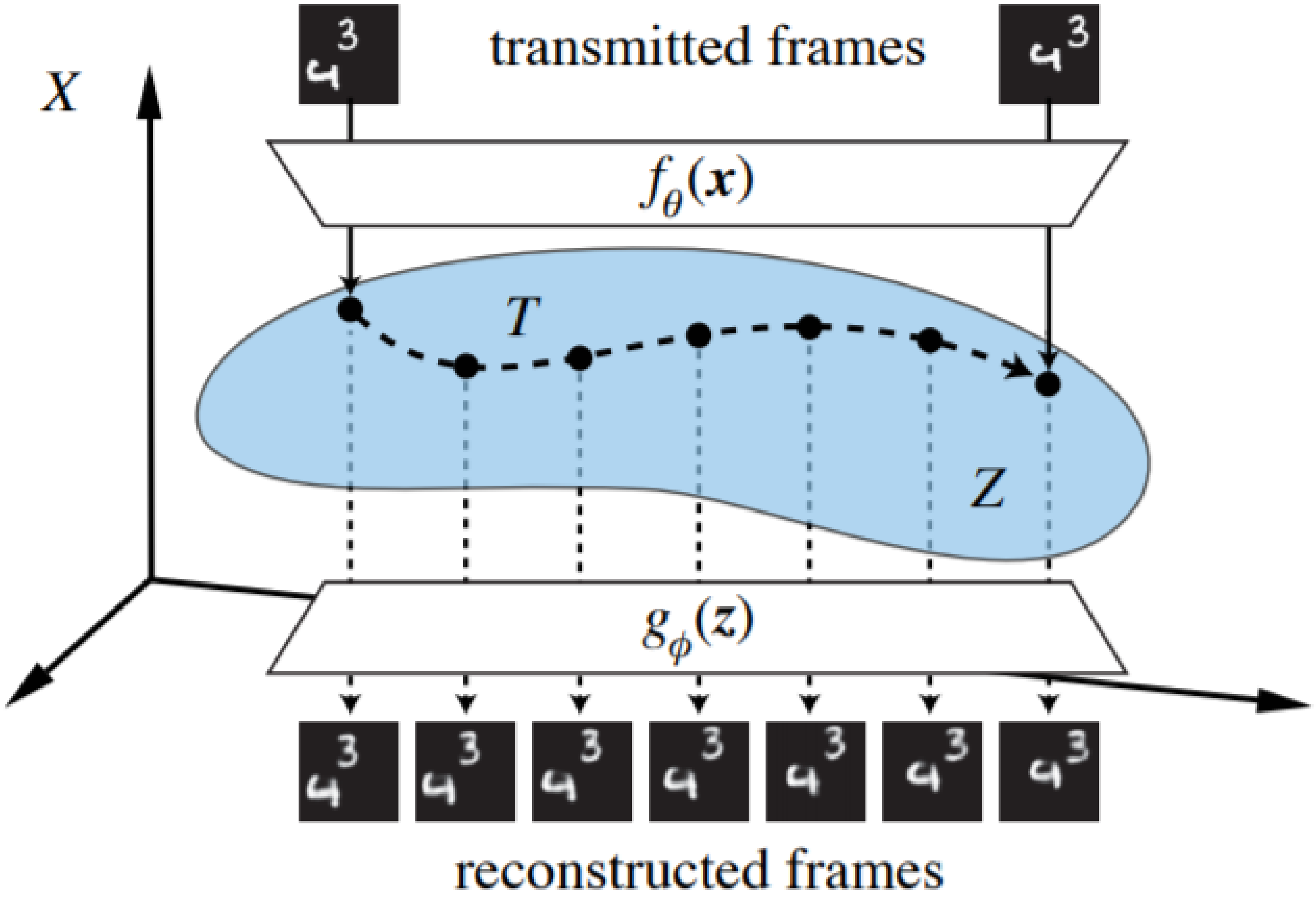


- Original

- Decoded image by GAN based method

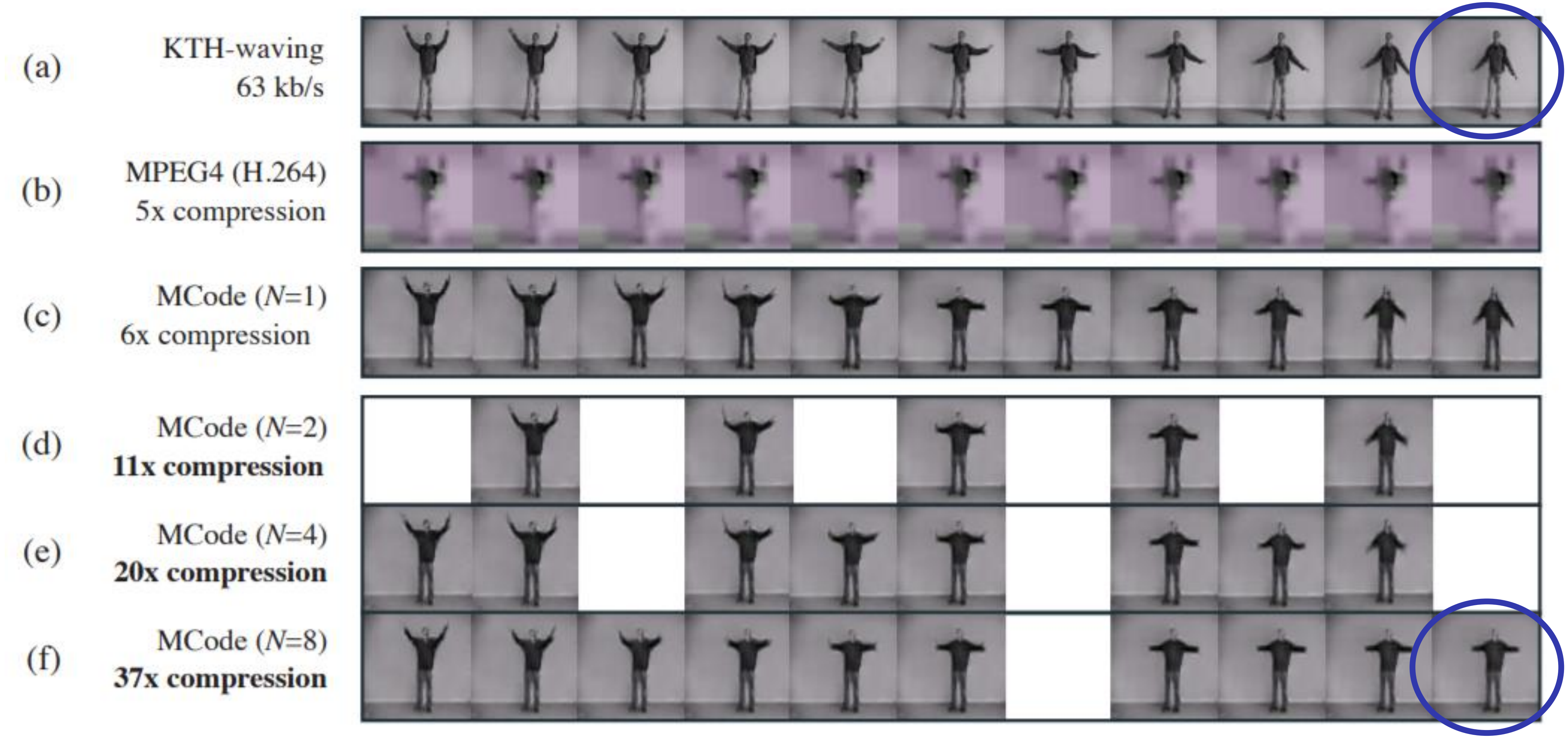
Agustsson et al., Generative Adversarial Networks for Extreme Learned Image Compression , 2019, ICCV

Special Purpose Coding



Santurkar et al. Generative Compression, 2018, PCS

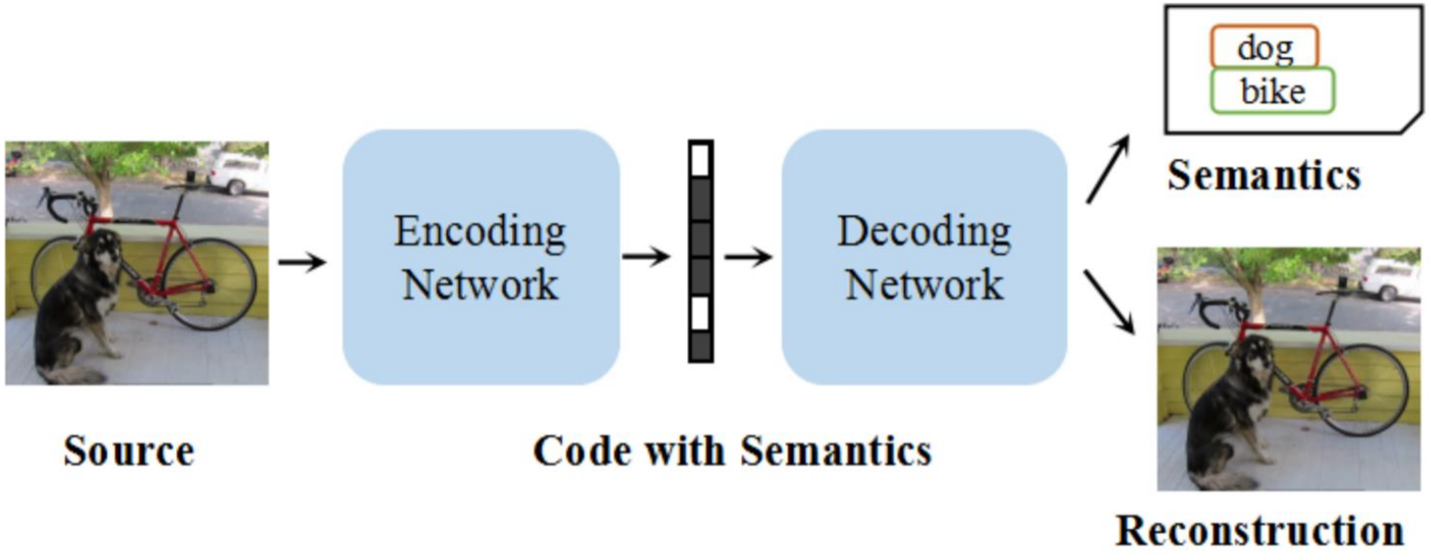
Special Purpose Coding



Santurkar et al. Generative Compression, 2018, PCS

Special Purpose Coding

- Semantic coding



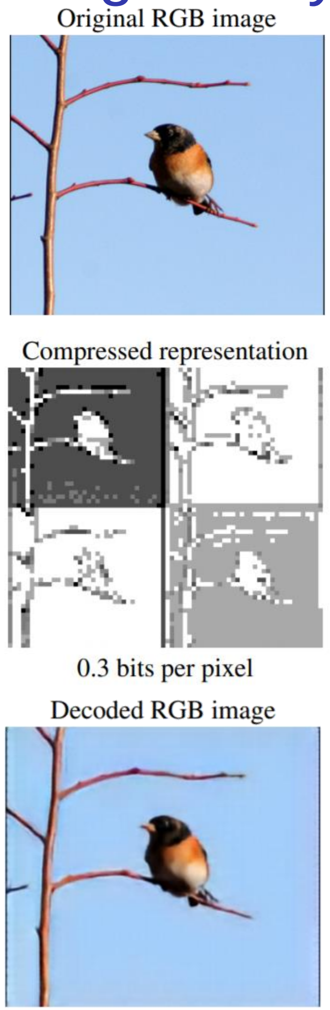
- Carry semantic information during storage and transmission
- Reduce computation of semantic analysis (such as object recognition) in client-side applications.

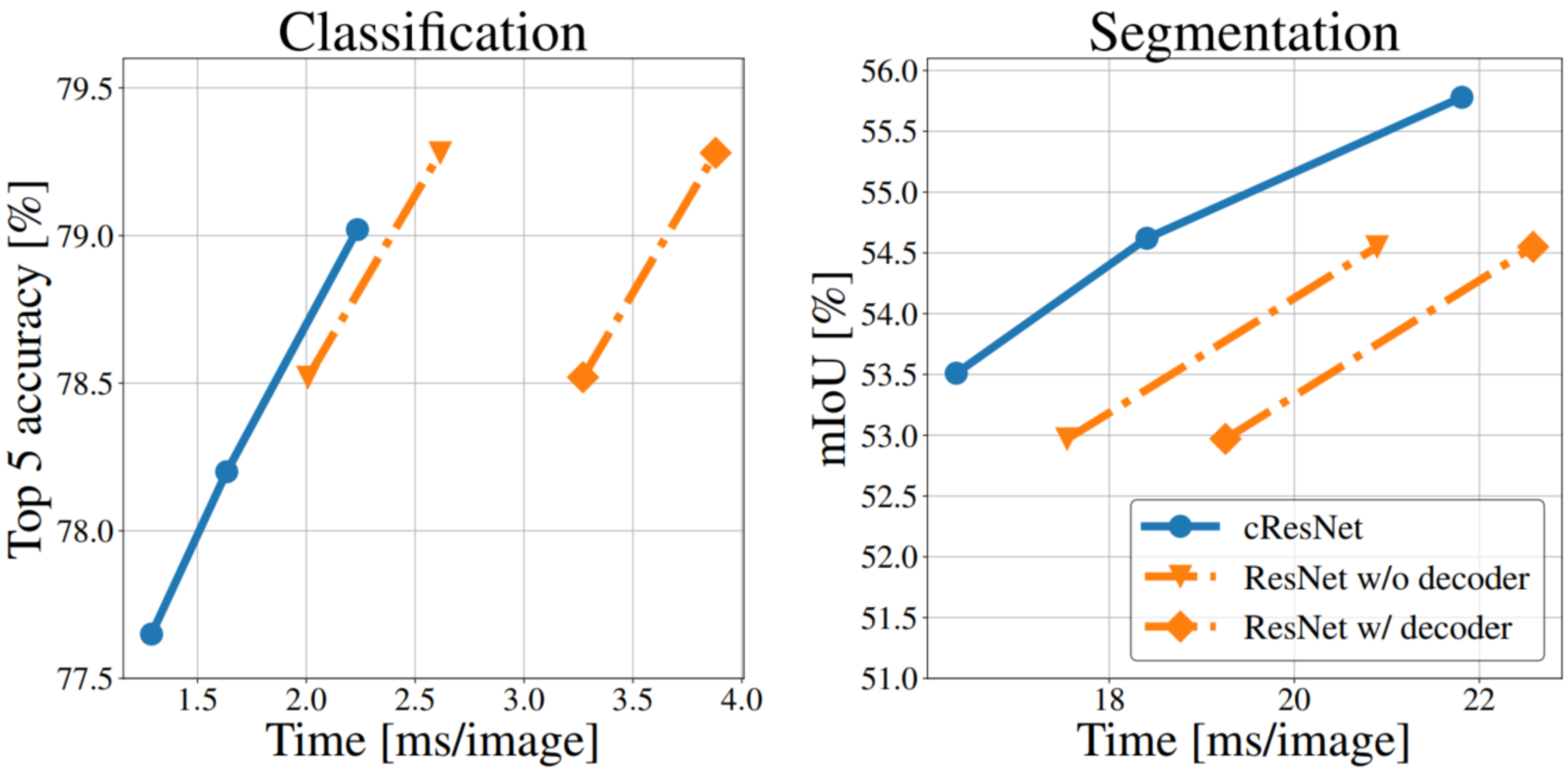
Loss: compression ratio + distortion + semantic analysis

Luo et al., DeepSIC: Deep Semantic Image Compression, 2018, ICONIP

Special Purpose Coding

- Image analysis in the compressed domain





Torfason et al., Towards Image Understanding from Deep Compression without Decoding, 2018, ICLR

Dataset

- Training dataset
  - Many existing image sets： ImageNet, DIV2K, *etc*.
  - Vimeo-90k dataset
    - 89,800 independent clips that are different from each other in content.
- Testing dataset
  - Kodak: 24 images with resolution 512x768
  - Tecnick: 100 images with resolution 1200x1200
  - UVG dataset
  - HEVC Standard Test Sequences
- **CLIC** -- CVPR Workshop and **C**hallenge on **L**earned **I**mage **C**ompression
    - On average with resolution of 1913x1361 for mobile photos
    - On average with resolution of 1803x1175 for professional photos
    - Updated year by year, since 2018
  - 2020
    - Predicted Frame Encoding track
    - Low bitrate track

Self supervised
No manual lable is needed

Outline

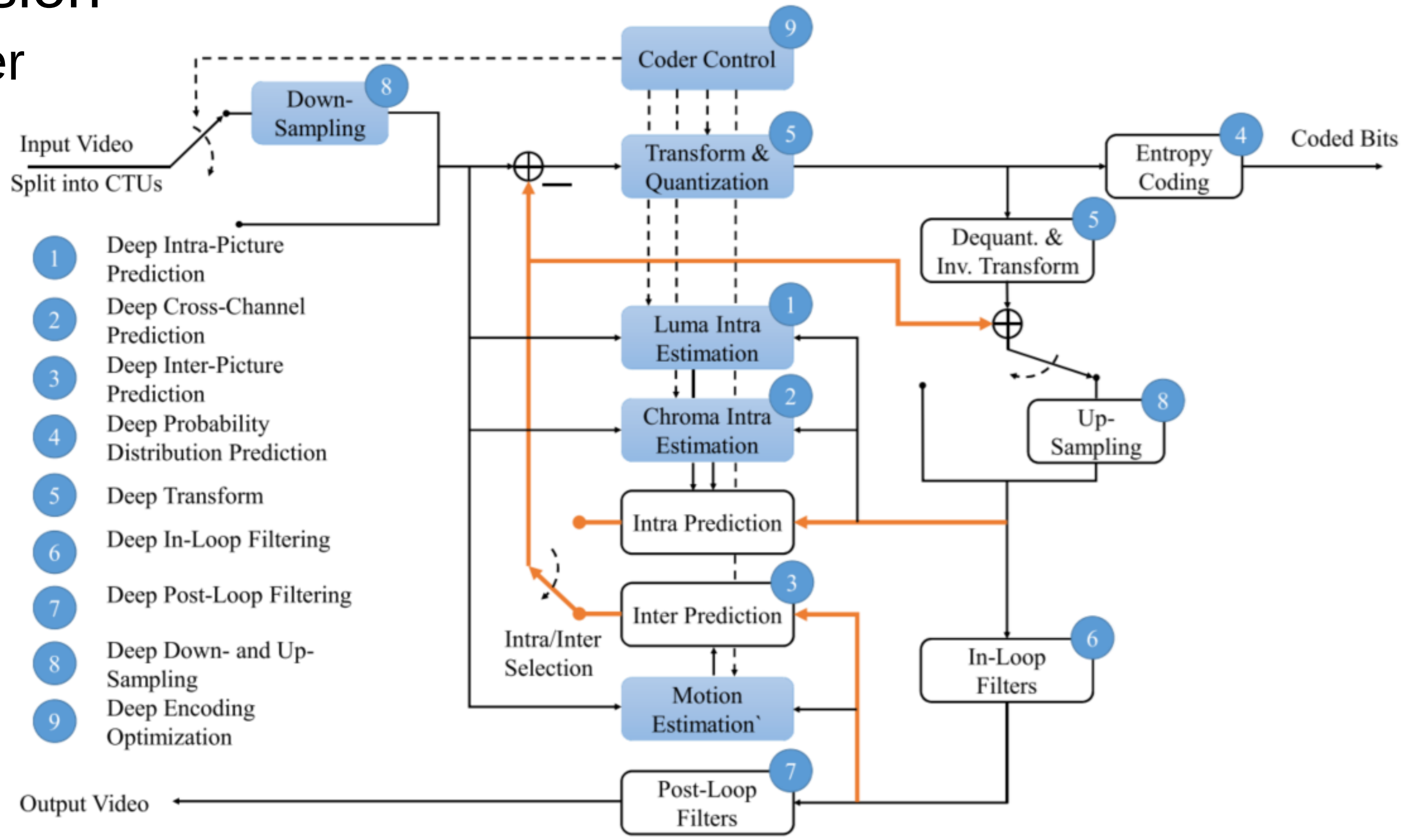
- Introduction of Image/Video Compression
- Image Compression
- Video Compression
- Special Purpose Coding
- Conclusion

Summary: Use of Deep Learning for Image/Video Compression

- Image Compression
  - Piecemeal Approaches
  - End to End Approaches
- Video Compression
  - Piecemeal Approaches
  - End to End Approaches
- Special purpose coding
  - Perceptual Naturalness
  - Extreme image compression
  - Semantic coding
  - Image analysis in the compressed domain

Discussion

- Neural to Classical Compression
  - Some blocks can be ported over
    - Learned transforms
    - Better entropy models
      - e.g. Hyperprior
    - Learned motion estimation
    - etc.
  - ‘Deepen’ traditional coding schemes
  - End to end deep schemes

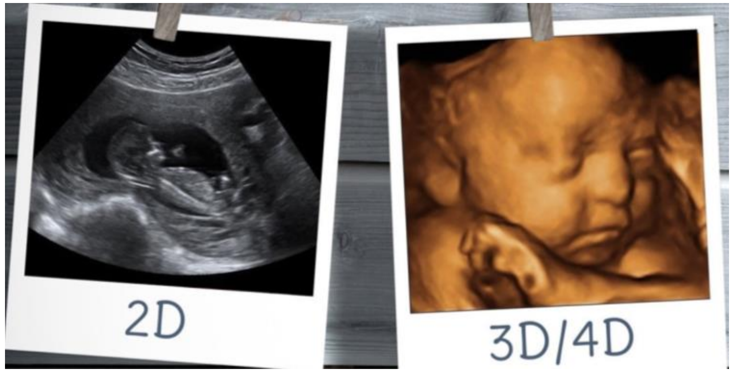


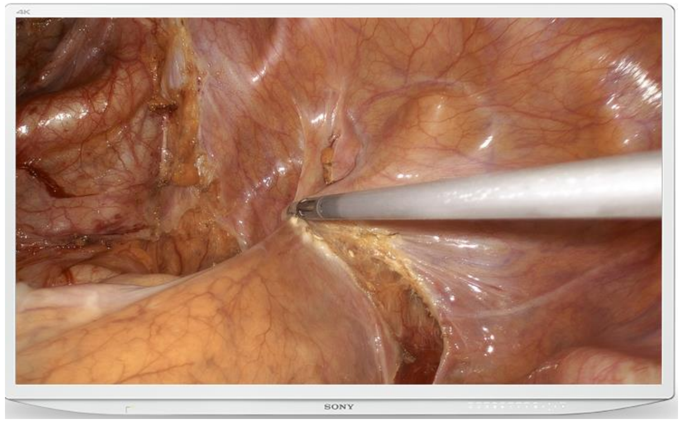
Liu et al., Deep Learning-Based Video Coding: A Review and a Case Study, 2020

Discussion

- Whole framework can be jointly optimized
- Flexible
  - MSE, SSIM, other differentiable objective metrics
- Optimization for special purpose:
  - perceptual naturalness
  - extreme image compression
  - semantic coding
  - ...
- Chance for new and more flexible schemes
- May help solving new challenges, e.g. medical data

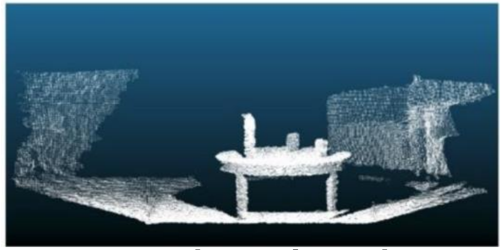
Medical data





4K,8K

Panorama



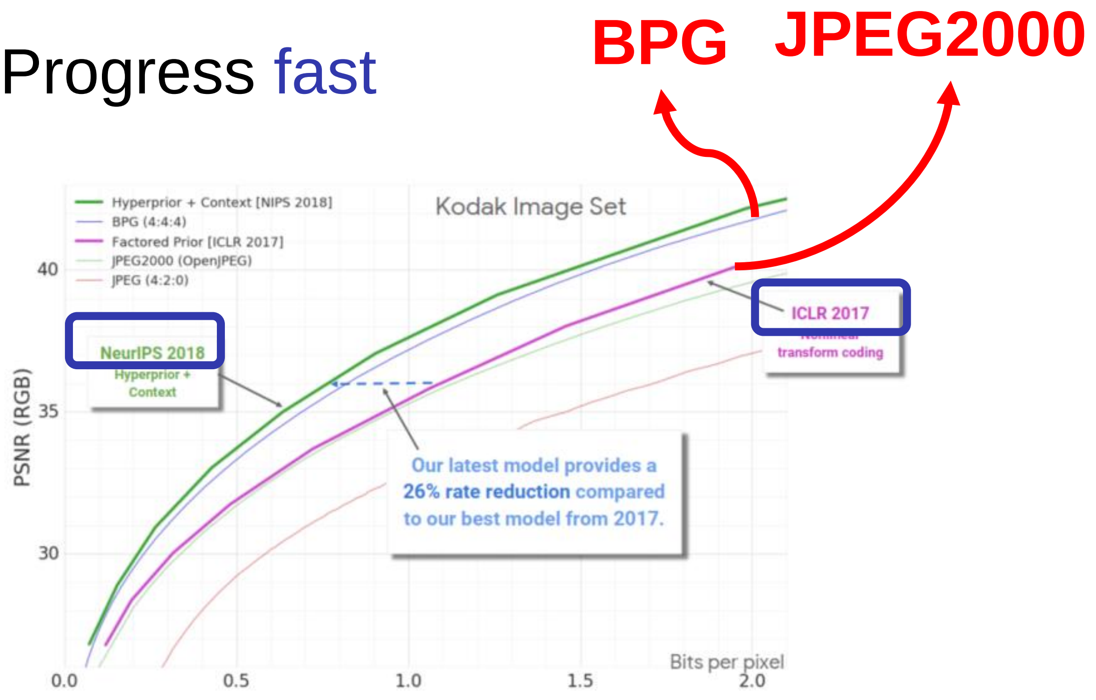
Point cloud

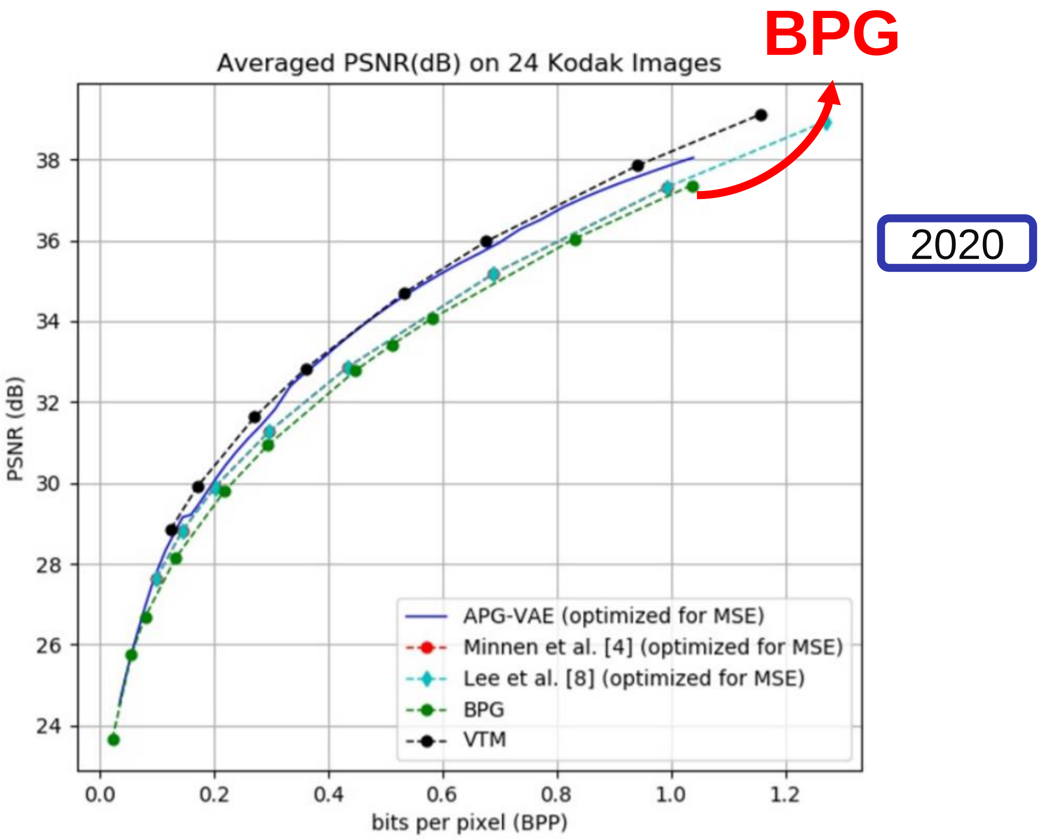
Discussion

Although,

- In its infancy, outperforms the JPEG2000 and slightly better/start to outperforms HEVC
- Computational demanding

Progress fast





PCS 2019, Toderici, Neural Image Compression: Recent Developments and Opportunities, Keynote

Cui et al., G-VAE: A Continuously Variable Rate Deep Image Compression Framework, 2020, arxiv

Future Considerations

- Deep compression for new data: point cloud, AR/VR data, medical data
- Computational efficient compression
- Energy efficient compression
- Better quality metrics: more perceptual related

Thank you