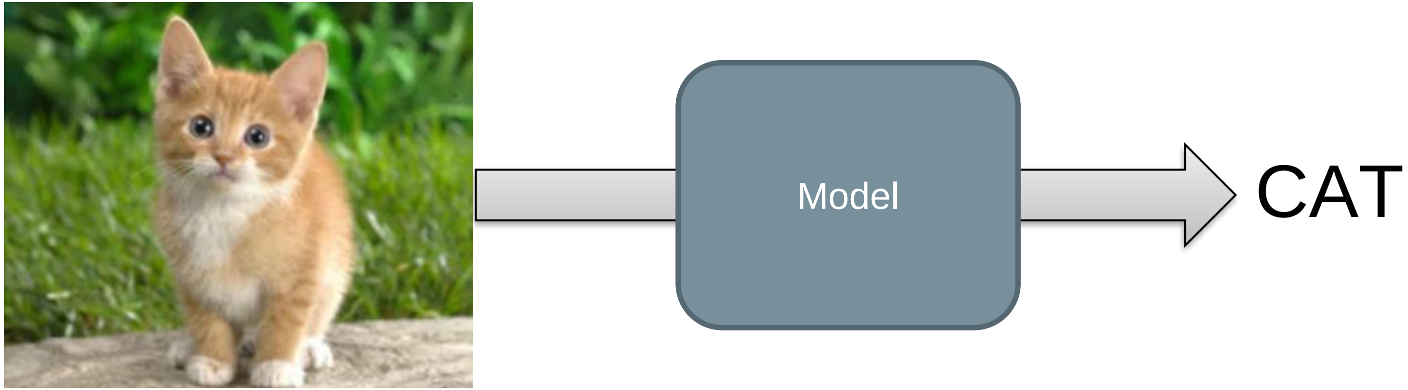
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

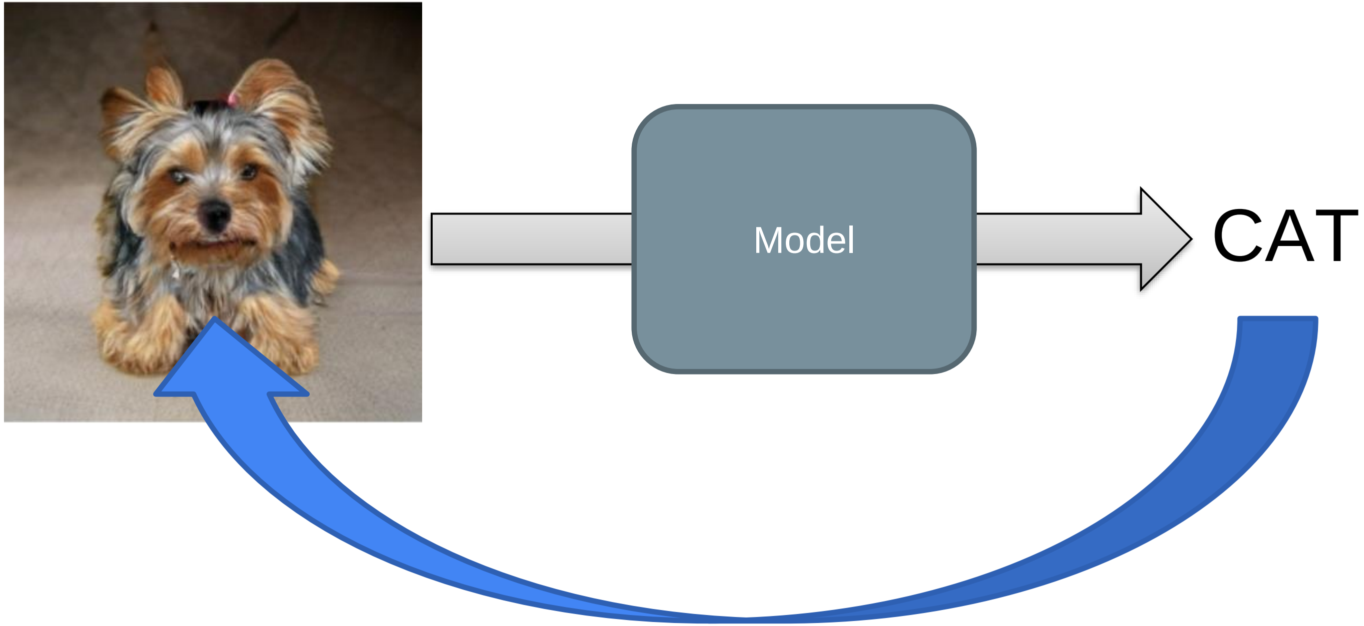
Exploring Invertibility in Image Processing and Restoration

Qifeng Chen
Visual Intelligence Lab
The Hong Kong University of Science and Technology

Visual Tasks are Two-fold: Discriminative



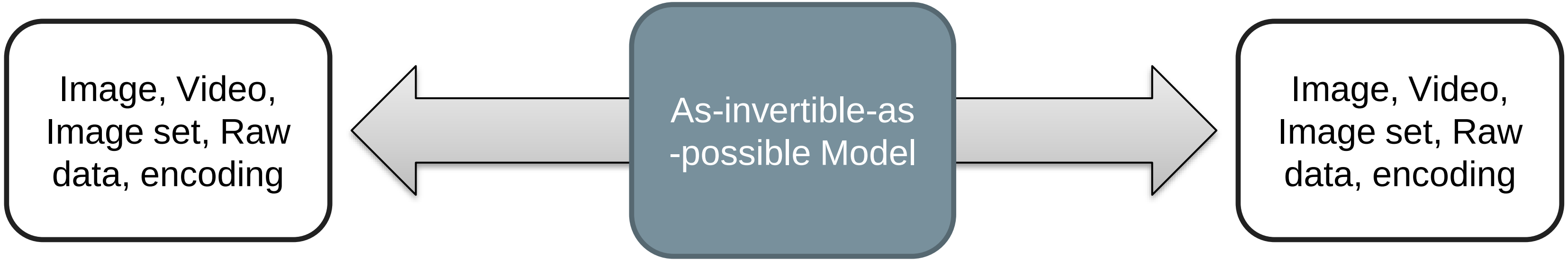
Visual Tasks are Two-fold: Generative



Class-conditional image synthesis

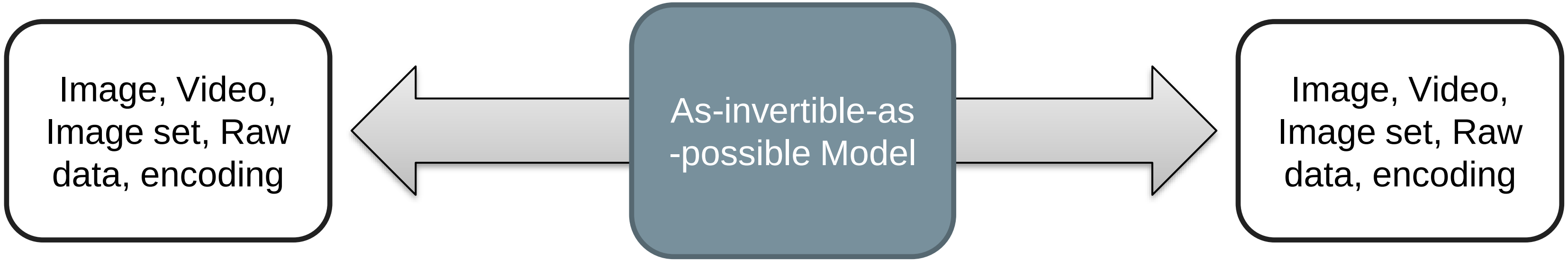
What is Invertibility?

- Joint learning of y=f(x) and x=g(y)
    - $g \approx f^{-1}$
    - x and y can have different dimensions



Exploring Invertibility in Practical Applications

- Invertible Image Signal Processing
- Image Compression
- Reversible Image Conversion
- Novel Views in a JPEG image



Invertible Image Signal Processing

Yazhou Xing*, Zian Qian*, Qifeng Chen
The Hong Kong University of Science and Technology
CVPR 2021

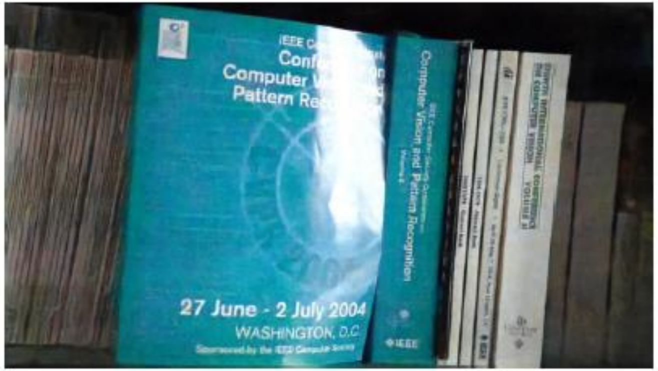
Code and data are available at
https://github.com/yzxing87/Invertible-ISP

Motivation

Why RAW data:

- Professional photographers choose to edit RAW images.
- Raw sensor data is also a better choice than RGB images for computer vision tasks, **due to its linear relationship with scene irradiance.**

Sony α7S II output



Enhanced results on camera RAW (Chen et al.)

Motivation

However, accessing RAW data can be hard:

- RAW is memory-demanding.
- RAW may be discarded during the process of data storing, transferring and sharing.

Our objective: enable users to get access to the real RAW data **without explicitly storing it.**

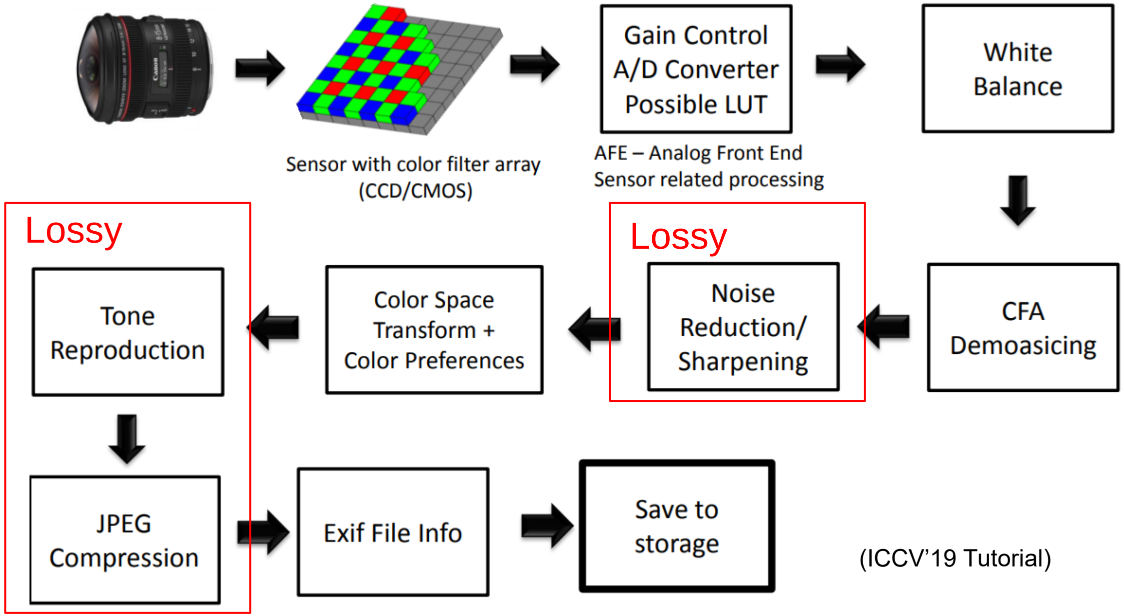
Results at a glance

Ground-truth RGB Ground-truth RAW UPI RAW [7] CycleISP RAW [44] Our RAW

Previous works

Synthesize RAW from RGB images: inaccurate results

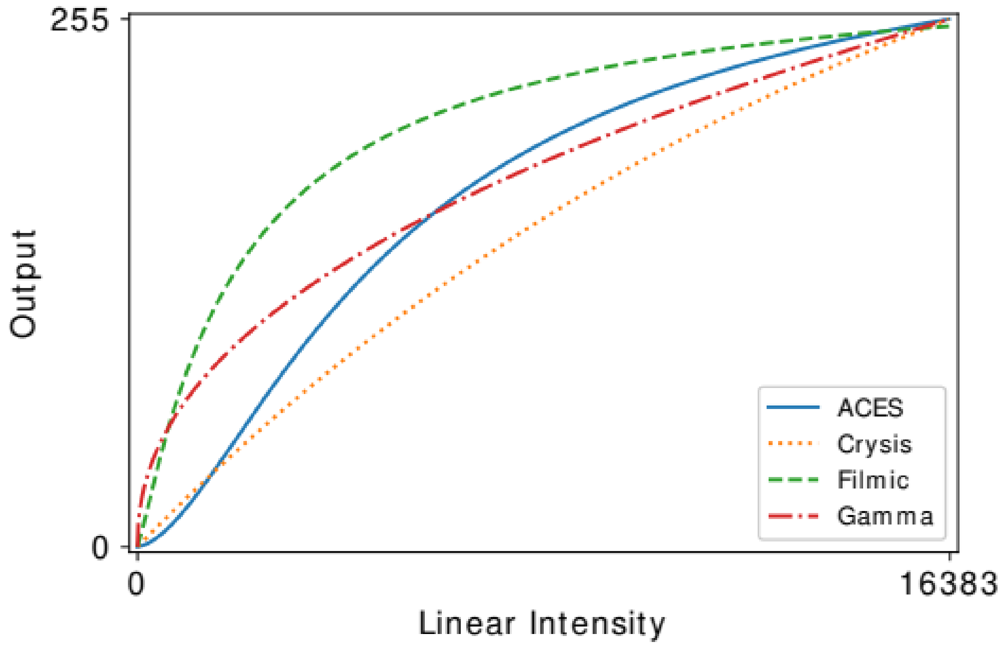
- key problem: rely on the underlying lossy in-camera ISP pipeline



(ICCV'19 Tutorial)

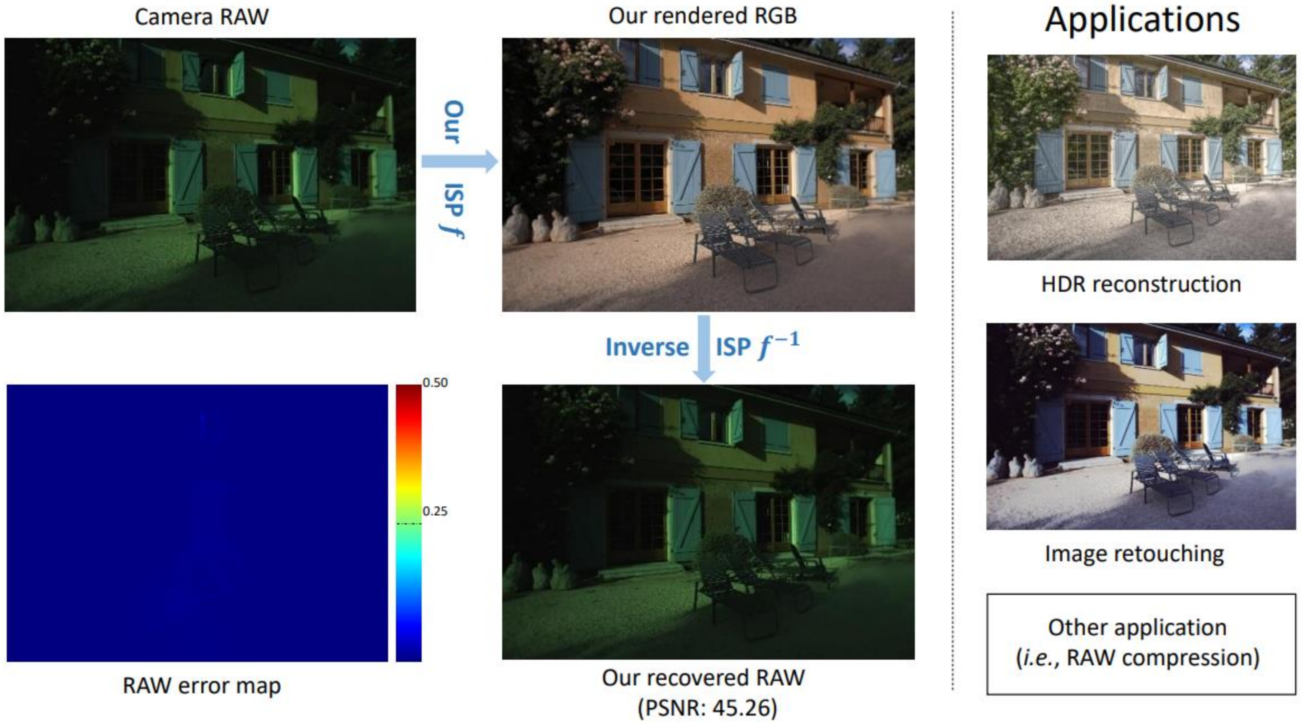
Traditional ISP analysis

- Quantization and tone mapping: may cause a 0.004 RMSE error at a single pixel
- Out-of-range clipping
- JPEG compression

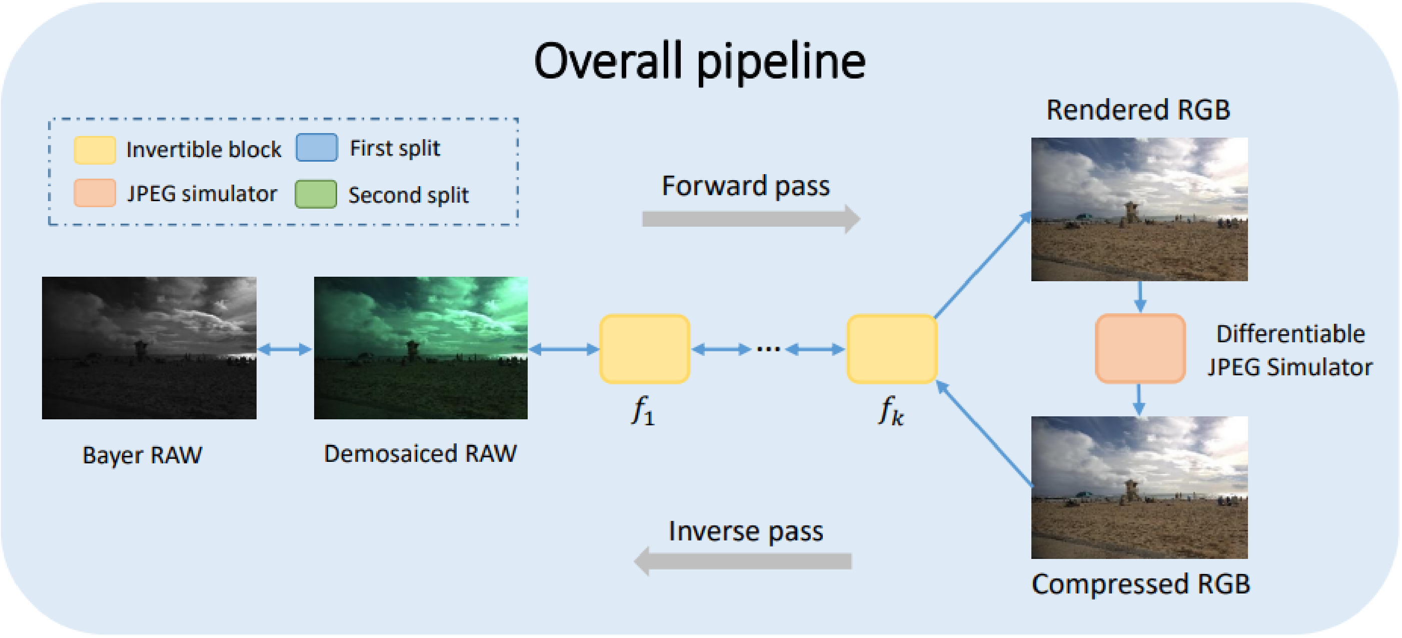


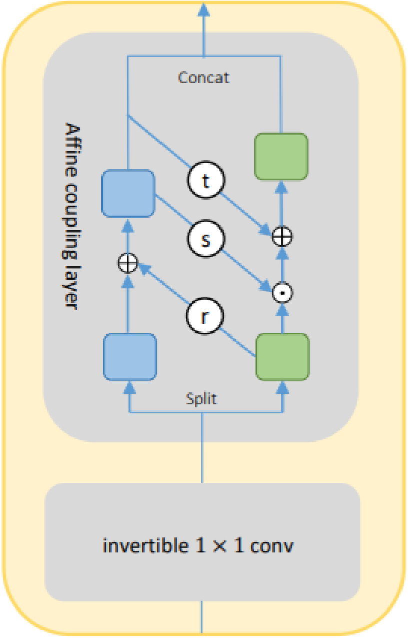
Our solution

- Design an invertible ISP for both RGB rendering and RAW reconstruction



Our solution





Invertible block f_i

- Invertible and bijective function: affine couling layers

$$\mathbf{y} = f_0 \circ f_1 \circ f_2 \circ \cdots \circ f_k(\mathbf{x}),$$

$$\mathbf{x} = f_k^{-1} \circ f_{k-1}^{-1} \circ \cdots \circ f_0^{-1}(\mathbf{y}).$$

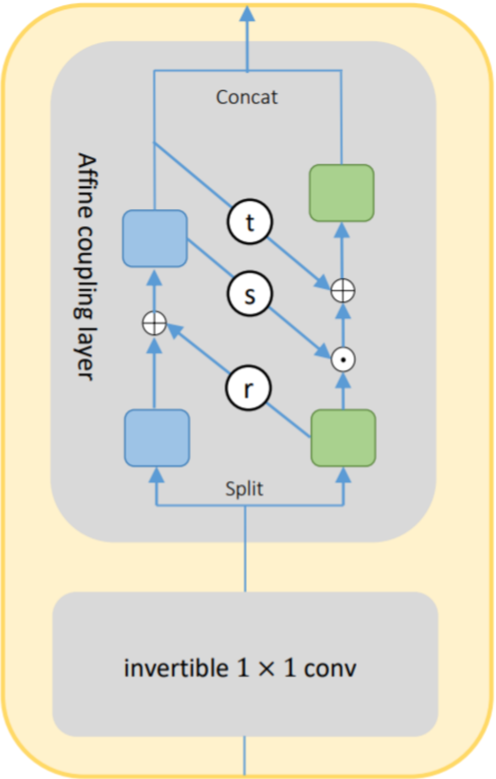
Our solution

- Affine coupling layer: given a D dimensional input m, n is calculated by

$$\mathbf{n}_{1:d} = \mathbf{m}_{1:d} + r(\mathbf{m}_{d+1:D}),$$
$$\mathbf{n}_{d+1:D} = \mathbf{m}_{d+1:D} \odot \exp\left(s\left(\mathbf{m}_{1:d}\right)\right) + t\left(\mathbf{m}_{1:d}\right),$$

- Its inverse function

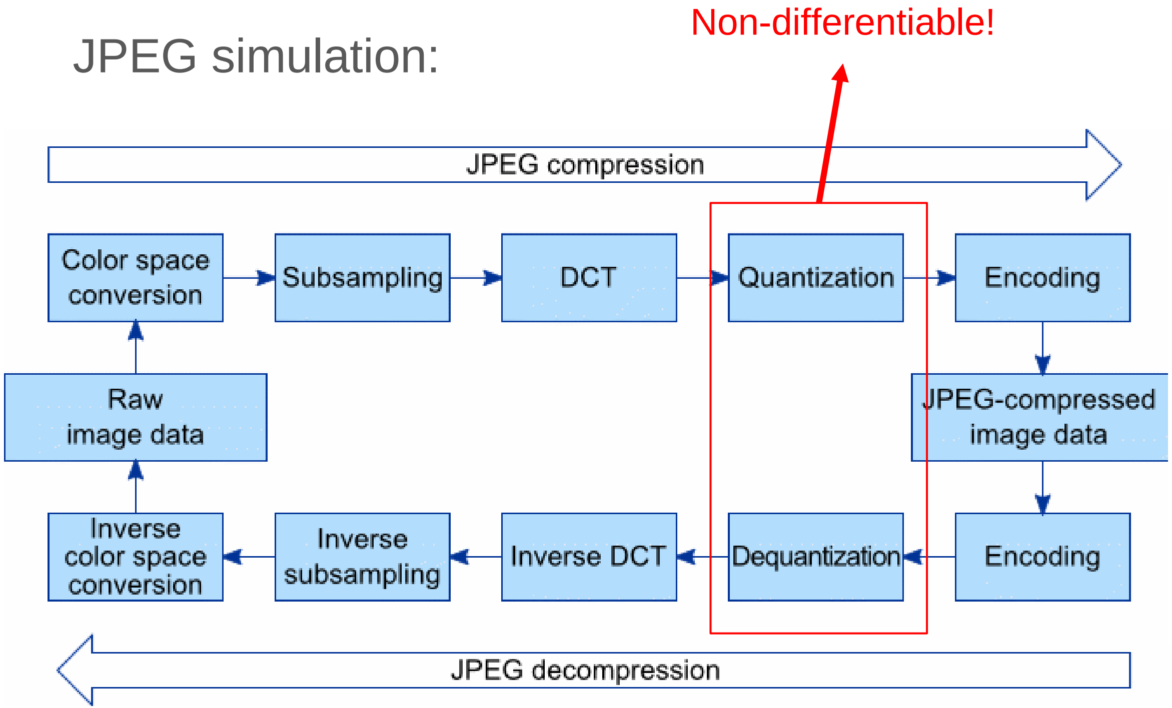
$$\mathbf{m}_{d+1:D} = \left(\mathbf{n}_{d+1:D} - t\left(\mathbf{n}_{1:d}\right)\right) \odot \exp\left(-s\left(\mathbf{n}_{1:d}\right)\right),$$
$$\mathbf{m}_{1:d} \quad = \mathbf{n}_{1:d} - r(\mathbf{m}_{d+1:D}).$$



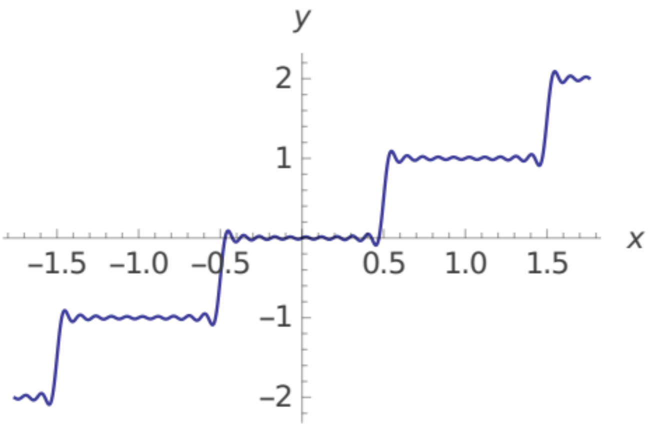
Invertible block f_i

Our solution

JPEG simulation:



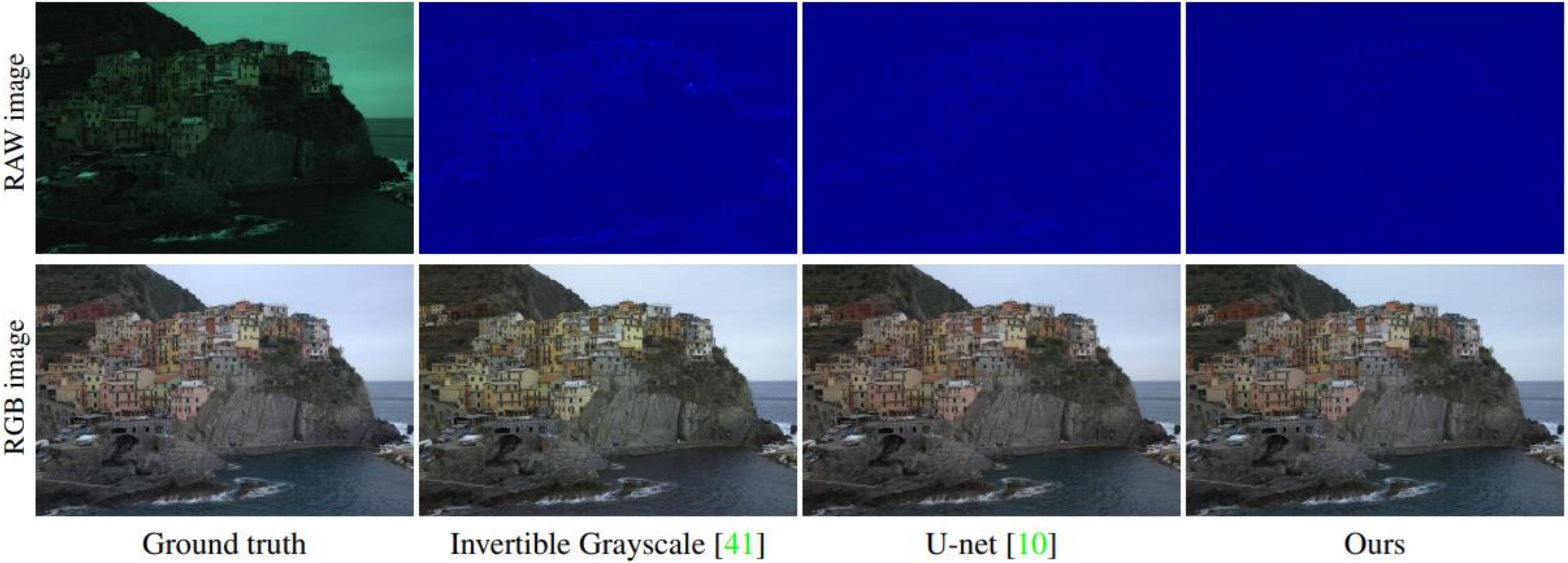
(graphicsmill.com/)



Fourier transformation for differentiable quantization

Experiments

Qualitative evaluation



Experiments

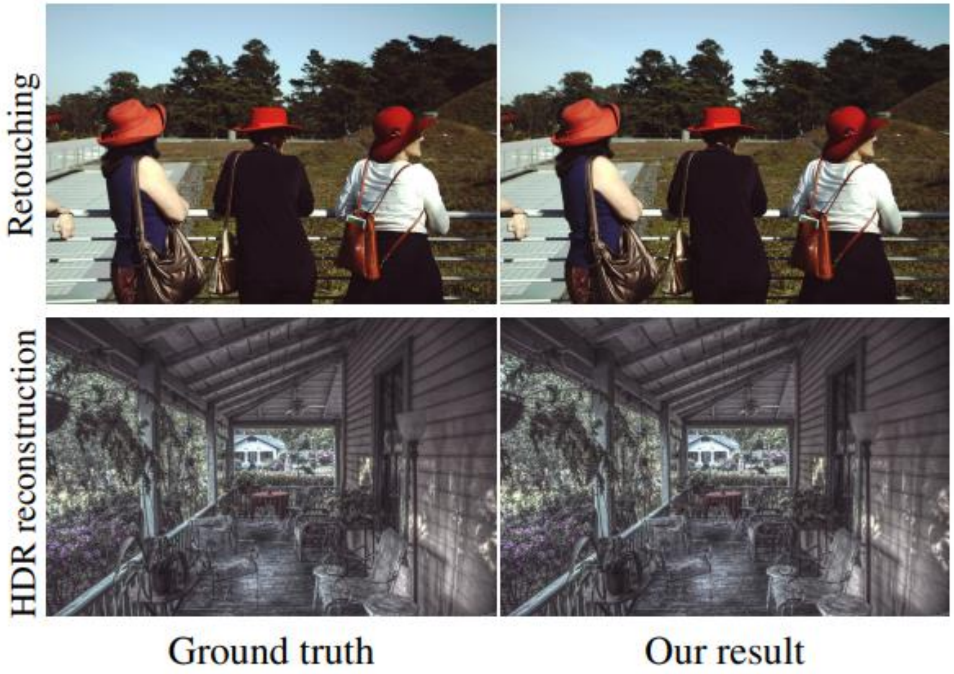
Quantitative evaluation

Method	NIKON D700			Canon EOS 5D		
	RGB		RAW	RGB		RAW
	PSNR	SSIM	PSNR	PSNR	SSIM	PSNR
UPI [7]	-	-	30.12	-	-	26.31
CycleISP [44]	-	-	30.19	-	-	34.48
InvGrayscale [41]	24.13	0.8258	33.28	28.22	0.8714	38.00
U-net [10]	36.48	0.9342	41.17	33.44	0.8893	41.14
Ours (w/o JPEG simulation)	37.44	0.9309	44.19	33.45	0.8923	45.73
Ours (JPEG with DSQ [17])	37.44	0.9467	**45.25**	33.15	0.8946	48.22
Ours (JPEG with Fourier)	**37.47**	**0.9473**	45.23	**33.61**	**0.9007**	**48.57**

Applications

- RAW compression
- Retouching and HDR

Dataset	Compression ratio ↑		BPP ↓	
	Lossy DNG	Ours	Lossy DNG	Ours
NIKON D700	1.61	34.98	8.73	0.4655
Canon EOS 5D	1.52	27.37	6.56	0.5237

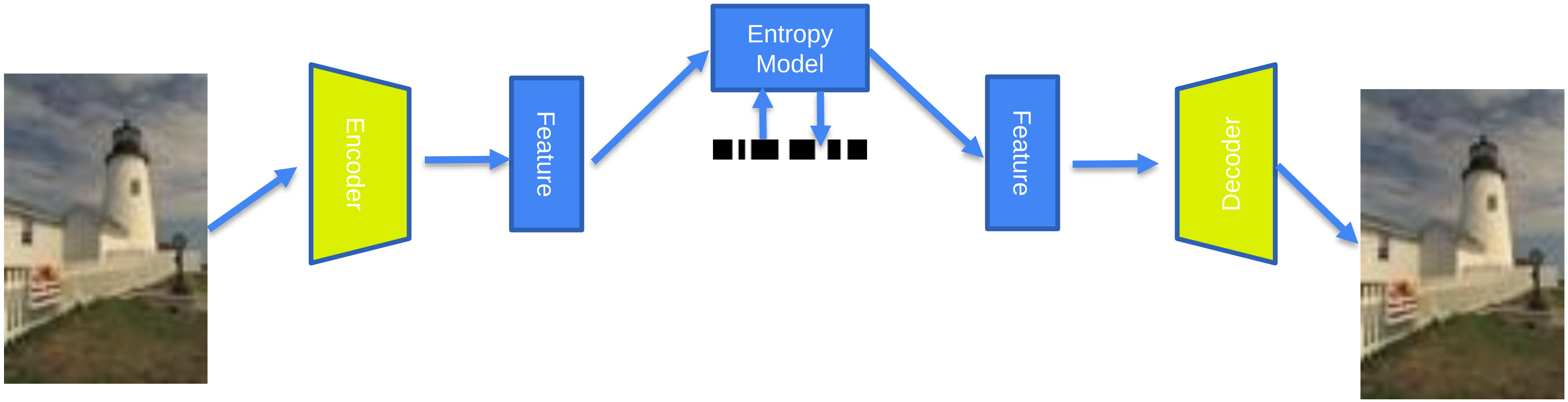


Code and data available at
https://github.com/yzxing87/Invertible-ISP

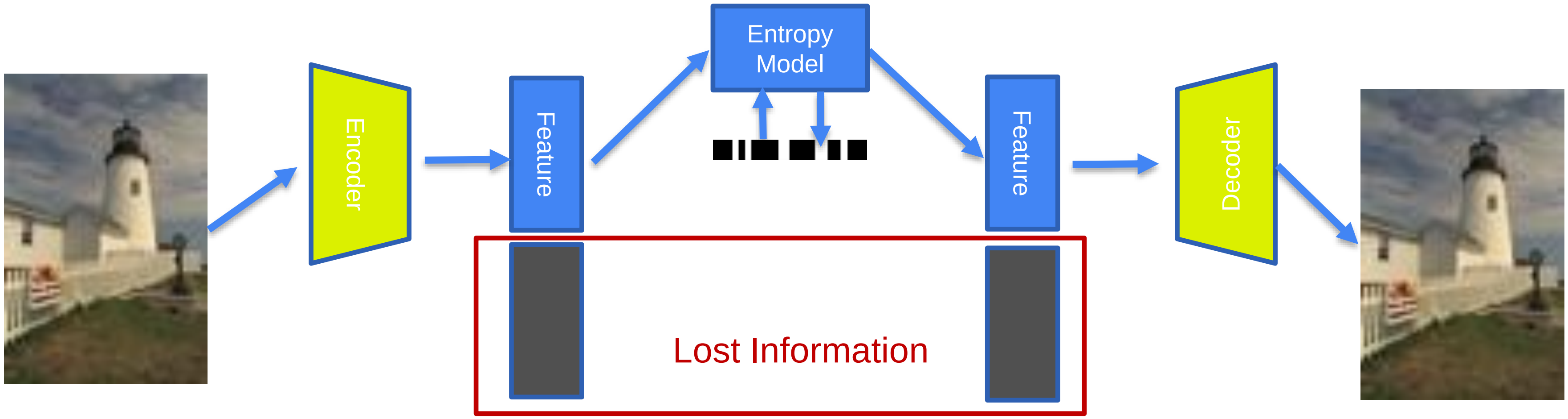
Enhanced Invertible Encoding for Learned Image Compression

Yueqi Xie*, Ka Leong Cheng*, Qifeng Chen
HKUST

Background
Autoencoder-based Image Compression



Motivation
Information Loss Problem

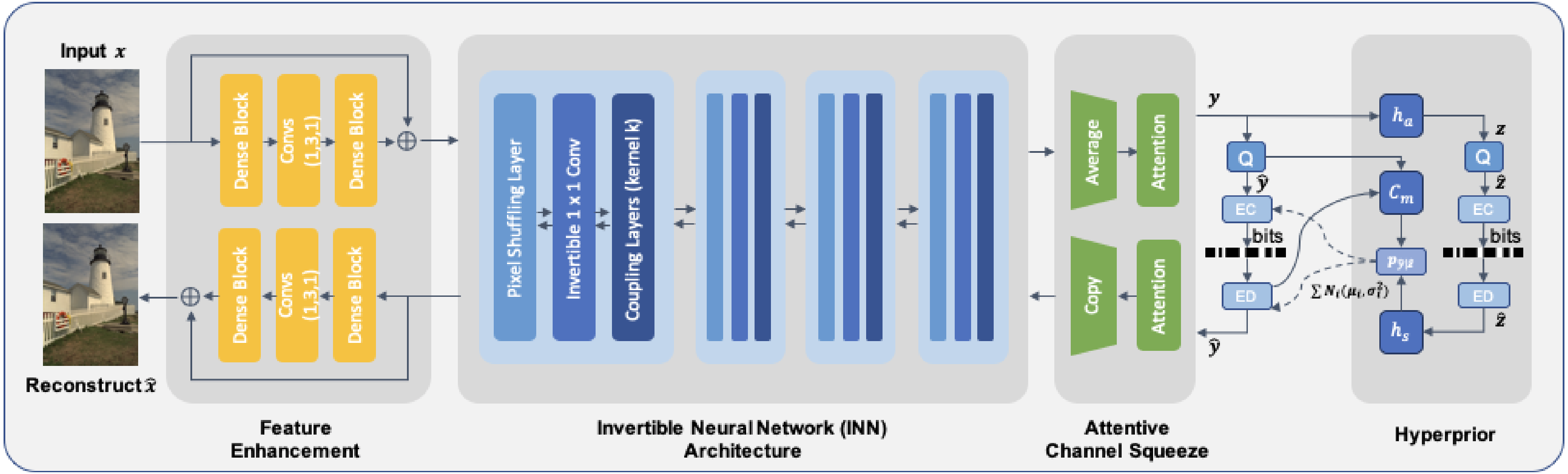


Challenges

- The reduction of total dimension for low bitrate compression
  - Previous solution: model lost information with a given distribution [1]
  - Problem:unwanted errors and unstable training
- Limited nonlinear transformation capacity of INN [2]

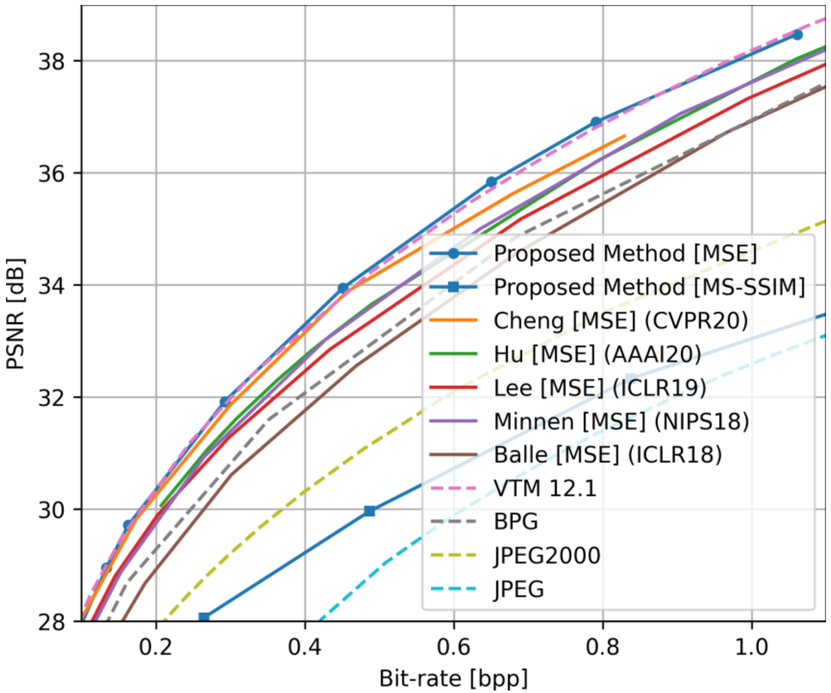
[1] Yaolong Wang, Mingqing Xiao, Chang Liu, Shuxin Zheng, and Tie-Yan Liu.2020. Modeling Lost Information in Lossy Image Compression
[2] Laurent Dinh, David Krueger, and Yoshua Bengio. 2015. NICE: Non-linear Inde-pendent Components Estimation.

Methods

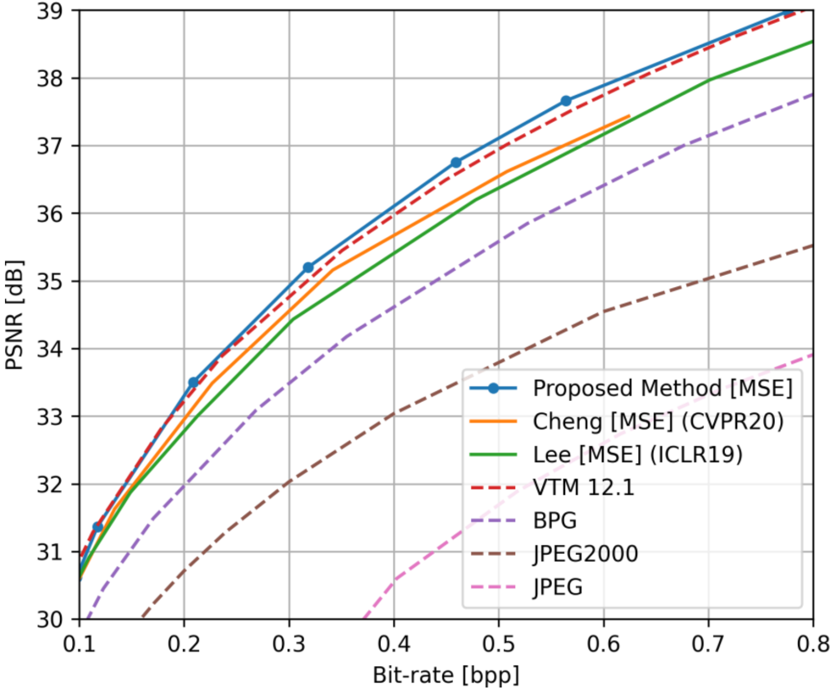


Input x
Reconstruct $\hat{x}$
Dense Block
Convs (1,3,1)
Dense Block
Pixel Shuffling Layer
Invertible 1 x 1 Conv
Coupling Layers (kernel k)
Average
Attention
Copy
Attention
y
$\hat{y}$
Q
EC
ED
bits
h_a
z
$\hat{z}$
C_m
$p_{\hat{y}|\hat{z}}$
$\sum N_i(\mu_i, \sigma_i^2)$
h_s
Feature Enhancement
Invertible Neural Network (INN) Architecture
Attentive Channel Squeeze
Hyperprior

Experiments

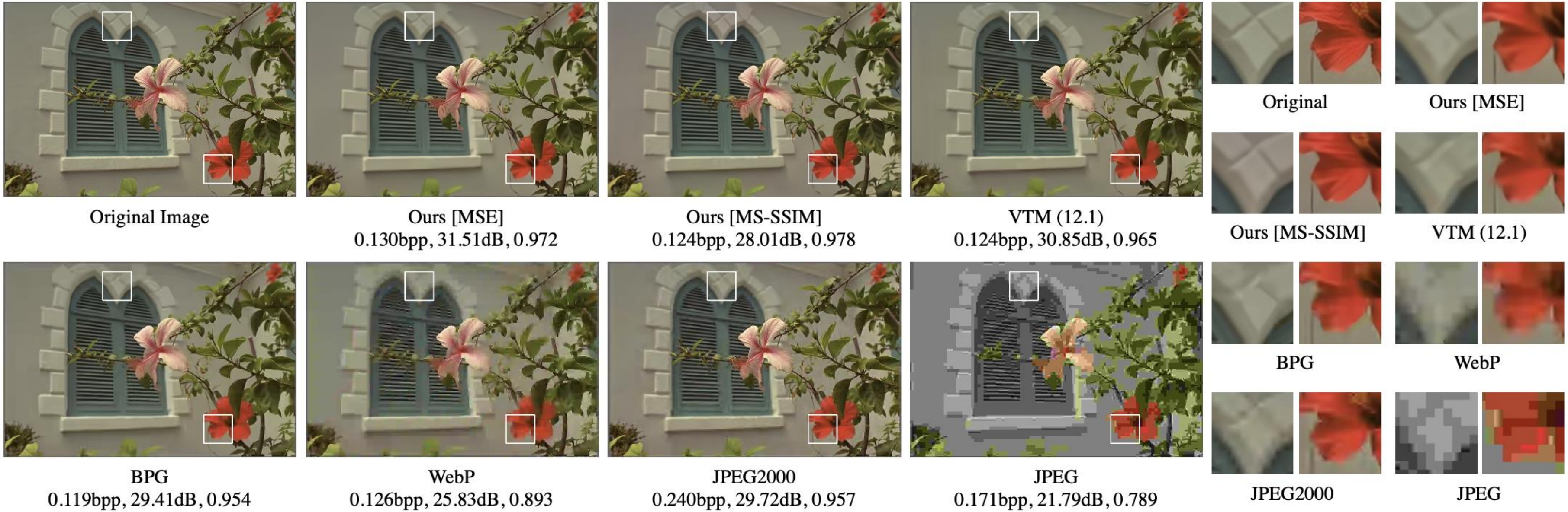


Kodak



CLIC

Experiment


Original Image
Ours [MSE]
0.130bpp, 31.51dB, 0.972
Ours [MS-SSIM]
0.124bpp, 28.01dB, 0.978
VTM (12.1)
0.124bpp, 30.85dB, 0.965
BPG
0.119bpp, 29.41dB, 0.954
WebP
0.126bpp, 25.83dB, 0.893
JPEG2000
0.240bpp, 29.72dB, 0.957
JPEG
0.171bpp, 21.79dB, 0.789
Original
Ours [MSE]
Ours [MS-SSIM]
VTM (12.1)
BPG
WebP
JPEG2000
JPEG

IICNet: A Generic Framework for Reversible Image Conversion

Ka Leong Cheng*, Yueqi Xie*, Qifeng Chen
HKUST

Background: Reversible image conversion

- Reversible image conversion (RIC) aims to build a reversible transformation between specific **visual content (e.g., short videos)** and an **embedding image**, where the original content can be restored from the embedding when necessary.
- For example,
    - GIF frames <-> single still image
    - Color image <-> grayscale image
    - High resolution image <-> low resolution image
    - Steganography: hiding multiple images into one cover image
    - Etc.

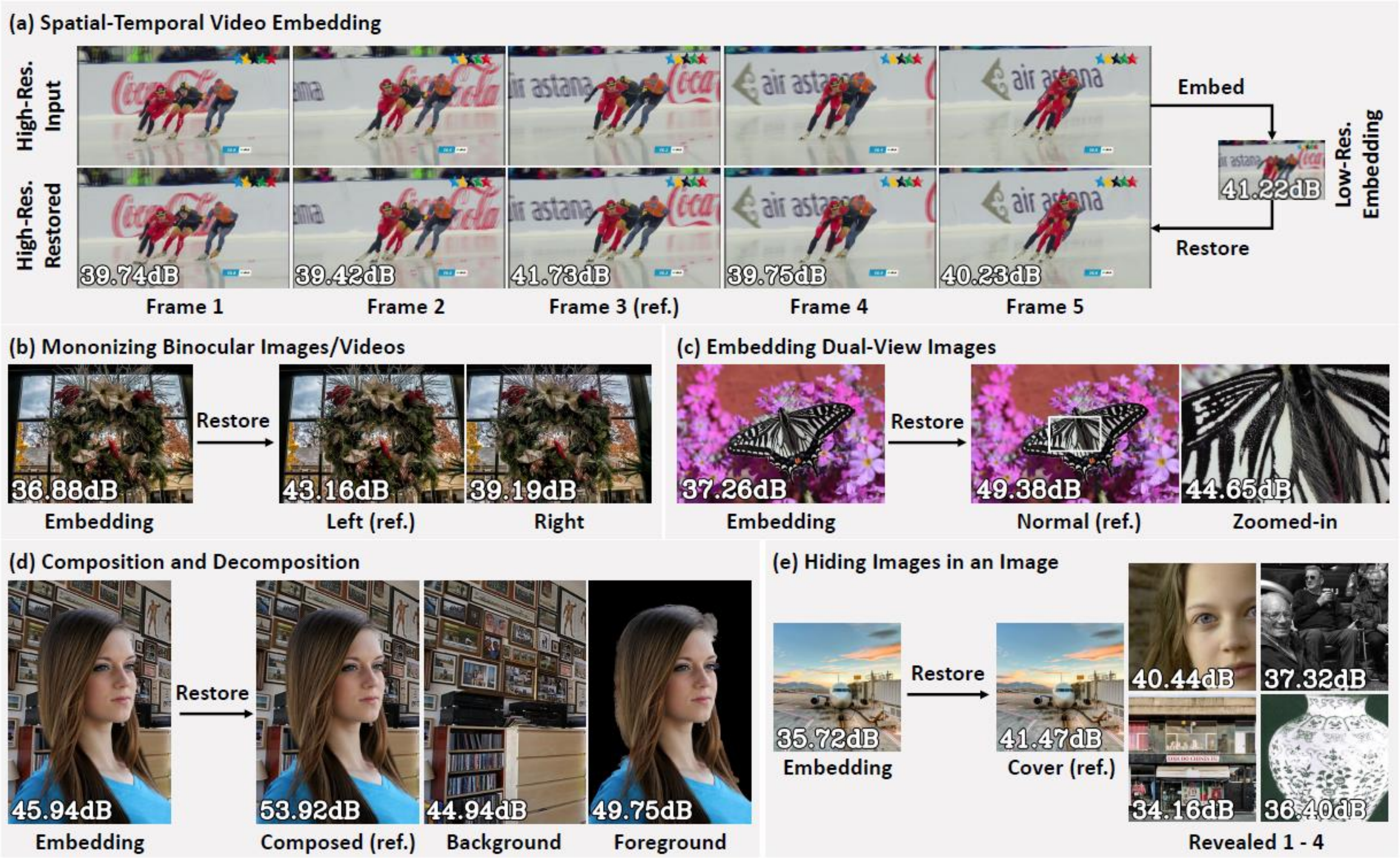


Figure 1: (a) Our IICNet can embed a sequence of high-resolution frames into one low-resolution embedding image and restore the original content nearly perfectly. (b)-(e) Our IICNet is the first approach that can generalize among various Reversible Image Conversion (RIC) tasks and yield state-of-the-art performance. Here, we show the restoration process.

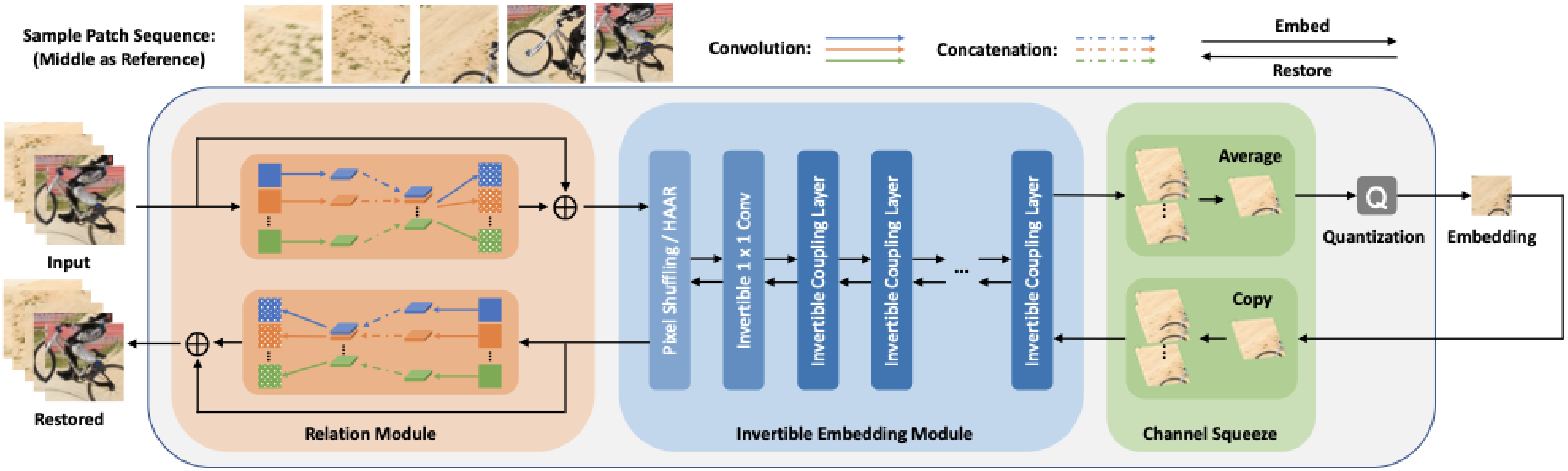
Challenges

- RIC tasks are challenging since we often need to embed much richer information implicitly in one single image, which may lead to unavoidable information loss.
  - Previous methods are generally encoder-decoder based methods, which learns the informative bottleneck representation but has limited ability to capture the lost information.
- All the existing methods usually have task-specific designs, so they can only solve for one specific RIC task.
  - For example, to embed a gif frame seqeuence into one single image, optical flow is usually considered in the methods.

Motivations

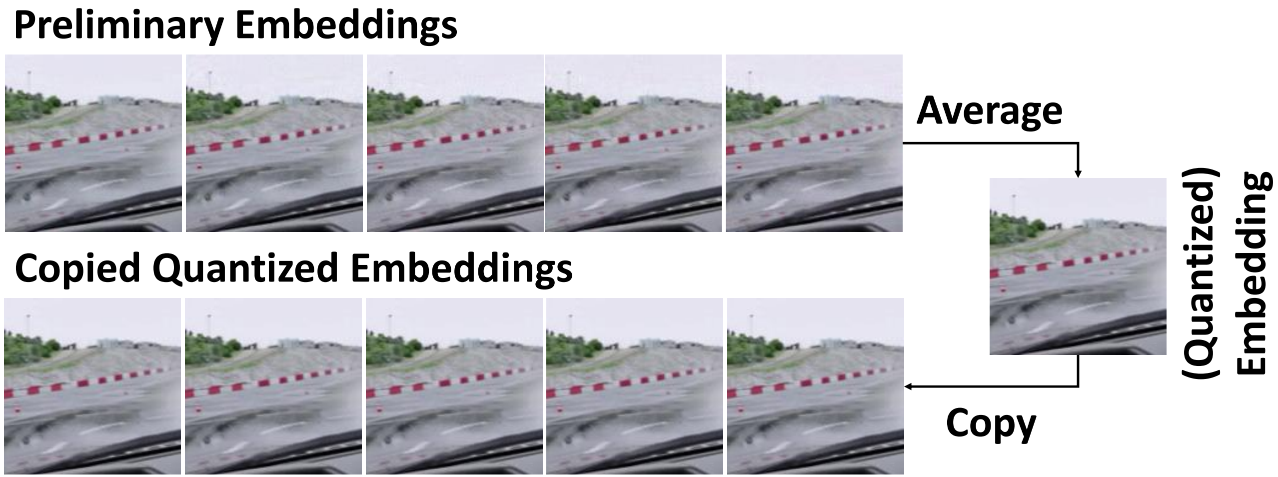
- Considering these aspects, we propose **Invertible Image Conversion Net (IICNet)** as a generic framework for RIC tasks.
- To alleviate the information loss problem, we utilize invertible neural networks (INNs) as a strictly invertible embedding module.
- With strong embedding capacity and generic modules, IICNet does not rely on any task-specific technique as prior work, which makes it possible to deal with different content types.
- Also, we no longer need to hand-engineer task-specific embedding networks when dealing with rapidly occurring visual content.

Our Framework



Channel Squeeze Layer

- Preliminary embedding images are like the target embedding image.



Strong Information Embeding Capacity

Detailed Restored Images

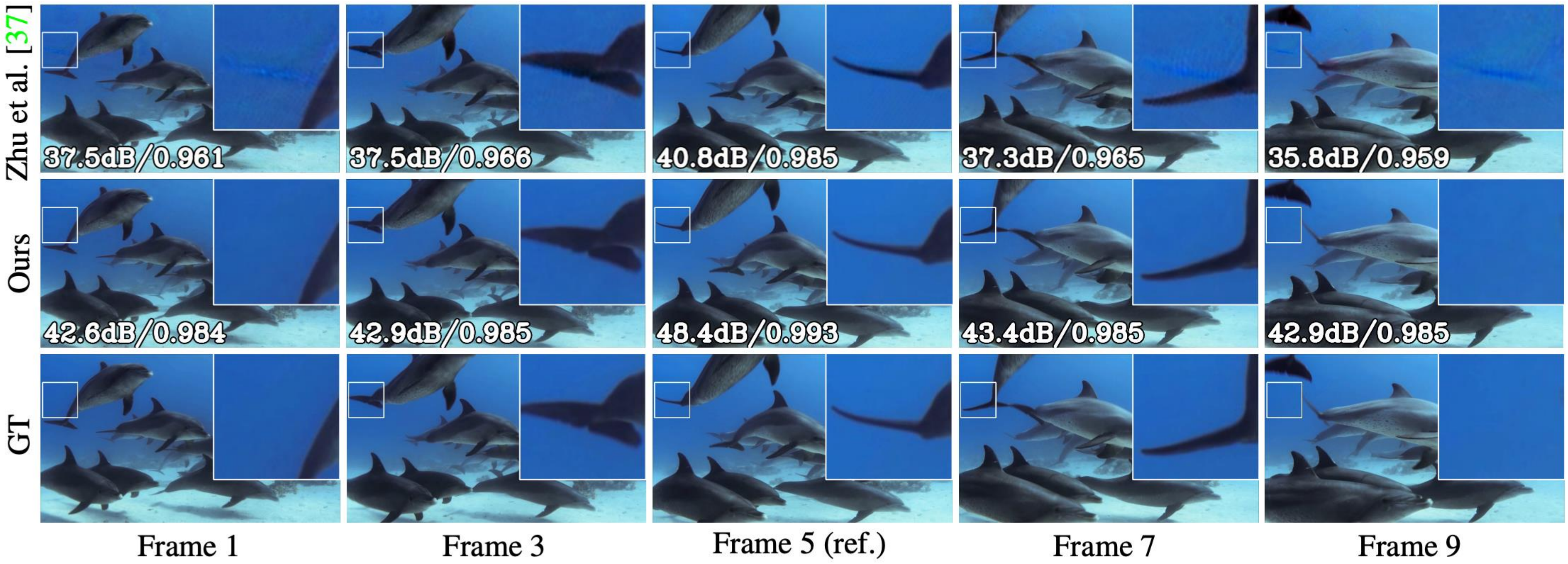


Figure 5: Visual result comparisons on restored frames.

Detailed Restored Images

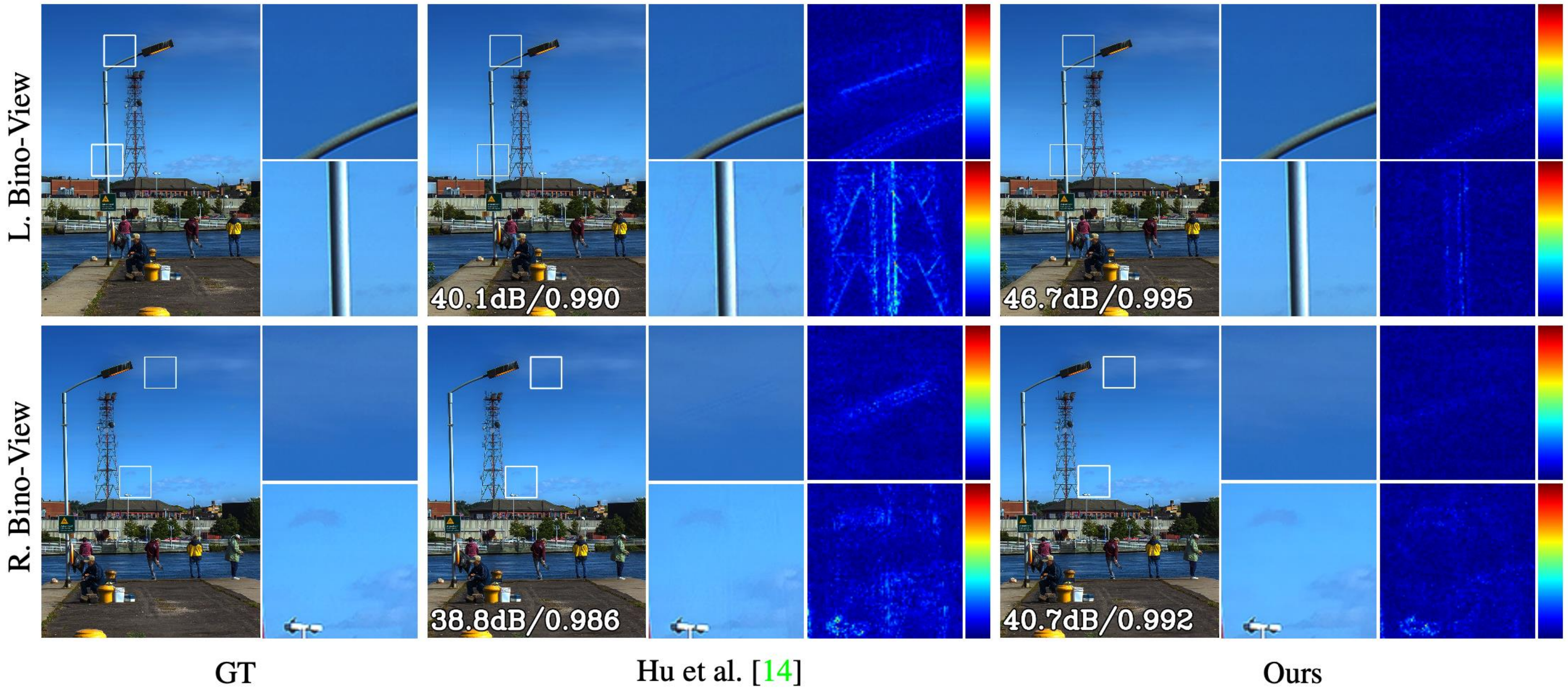
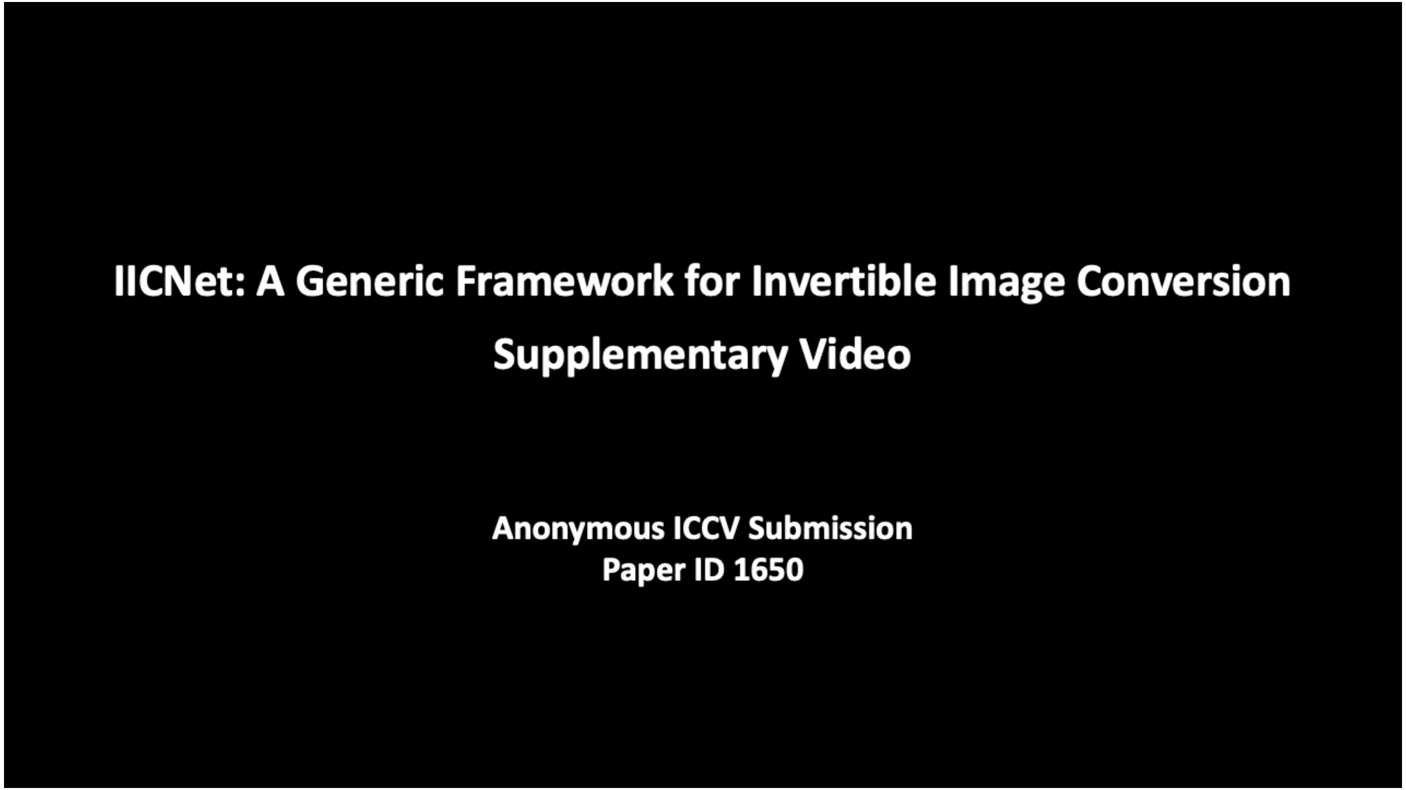


Figure 6: Visual comparison results of the restored Binocular Images. We show the zoomed-in patches with the corresponding error map aside. Note that we amplified the error maps by 10 times for better visualization.

Demo: Generic Framework

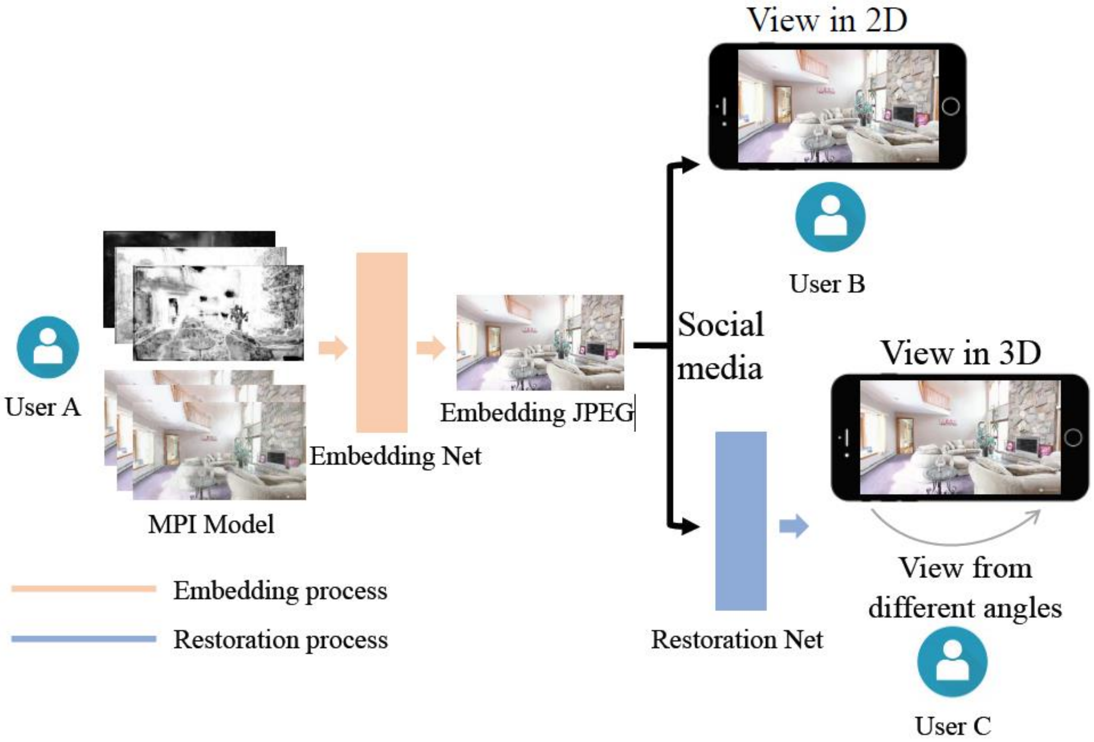


Embedding Novel Views in a Single JPEG Image

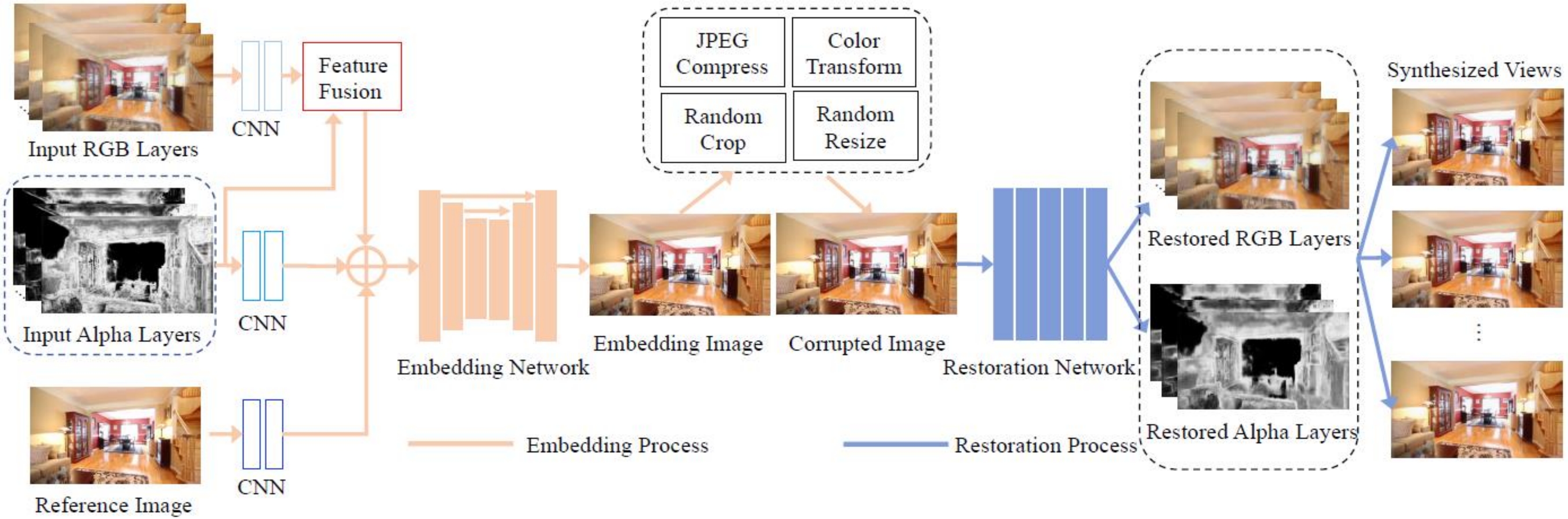
Yue Wu, Guotao, Meng, Qifeng Chen

ICCV 2021

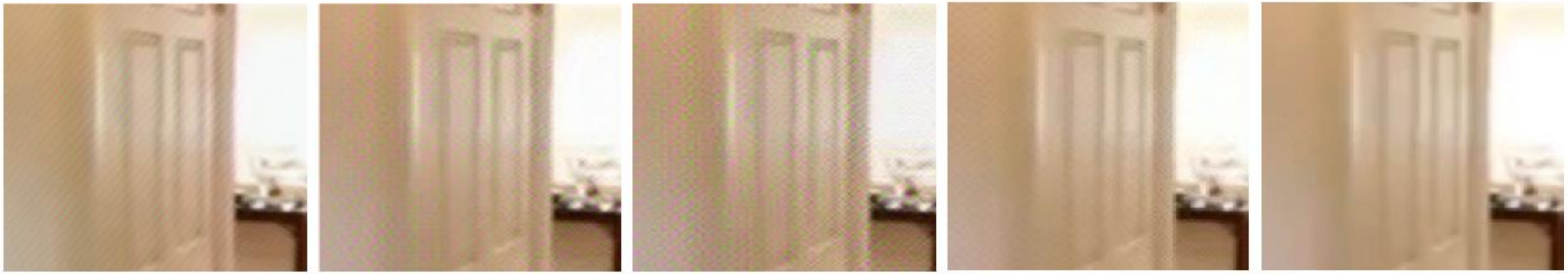
Can we share Novel Views as a JPEG Image?



Method


Input RGB Layers
CNN
Feature Fusion
Input Alpha Layers
CNN
Reference Image
CNN
Embedding Network
JPEG Compress
Color Transform
Random Crop
Random Resize
Embedding Image
Corrupted Image
Restoration Network
Restored RGB Layers
Restored Alpha Layers
Synthesized Views
Embedding Process
Restoration Process

Frequency domain loss



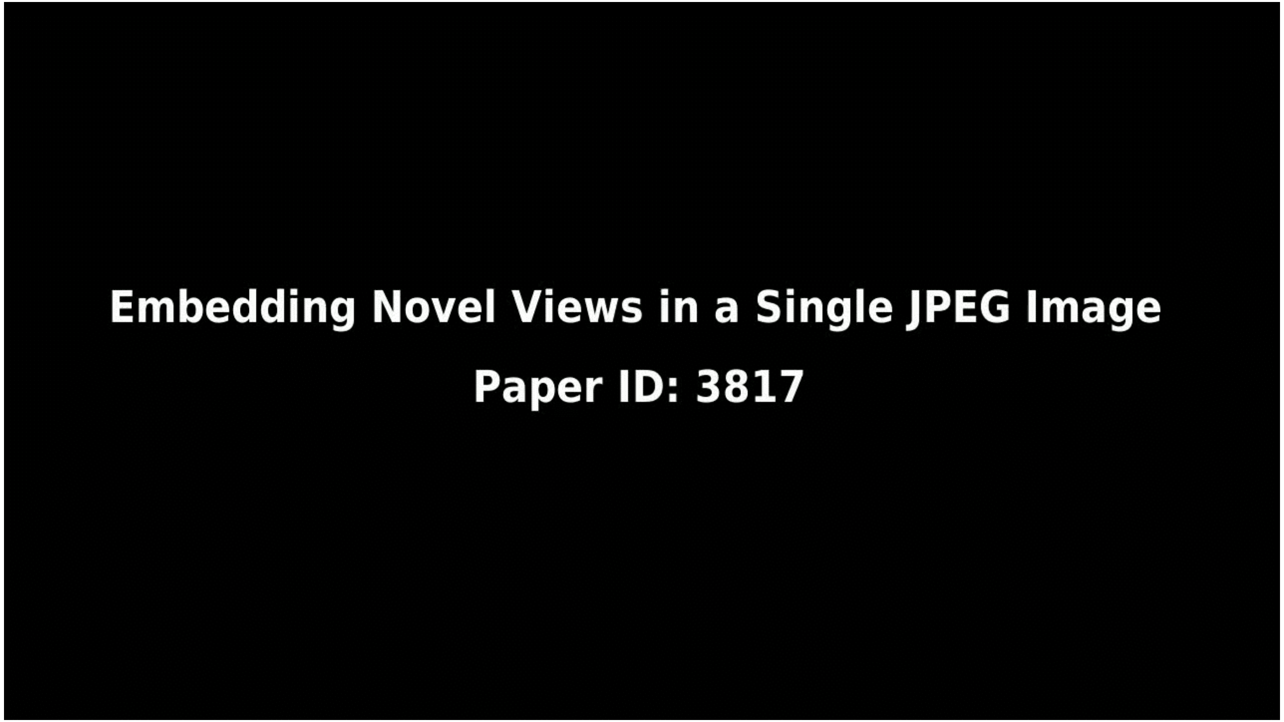
Video Snapshot [33] W/O frequency W/O adversarial Full model Ground truth

Figure 5. The comparison shows the effect of loss functions. Without the frequency loss and the adversarial loss, the embedding image has evident color strips and high-frequency artifacts. We recommend readers zoom in for the details.

$$\mathcal{L}_{freq} = ||FFT(\tilde{I}_e) - FFT(I_{ref})||^2$$

Qualitative Results

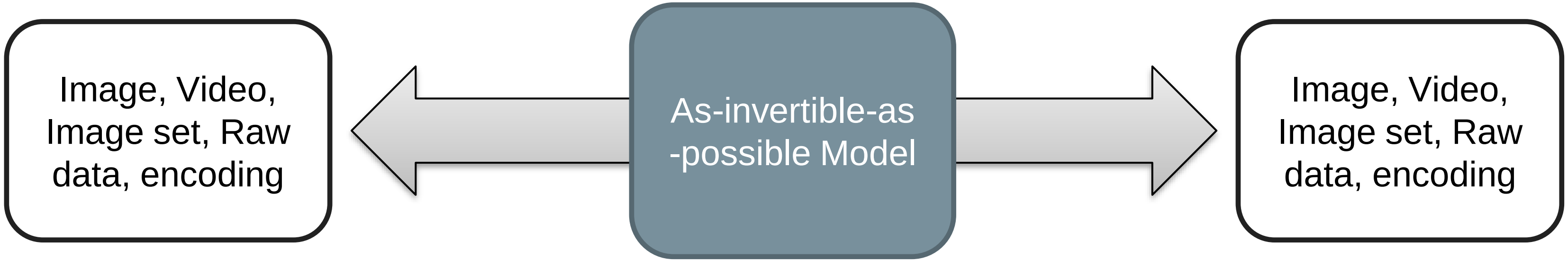
Results



Conclusions

As-invertible-as-possible models can enable

- Invertible Image Signal Processing
- State-of-the-art Image Compression
- General Reversible Image Conversion
- Embedding Novel Views in a JPEG image



Thanks

https://cqf.io