Google

Learned Image Compression Guided Adaptive Quantization For Perceptual Quality

Authors: Cheng Chen, Ruiqi Geng, Bohan Li, Maryla Ustarroz-Calonge, Frank Galligan, Jingning Han, Yaowu Xu

Presenter: Cheng Chen

ICIP 2023

Oct 2023

Overview

Perceptual quality is a complex problem

Perceptual quality

- Human visual system (HVS) is complex
- No single metric could model HVS precisely
- PSNR correlates poorly with HVS
- Conventional image compression algorithms are developed and optimized for PSNR
- Various metrics have been proposed to quantify perceptual quality

Learned image codecs

- Recent advances in learned image coding have shown that deep features learned from the latent space significantly outperform conventional perceptual metrics.

- By incorporating the Learned Perceptual Image Patch Similarity (LPIPS) metric into the loss function and optimizing it end-to-end, learned image codec achieves significant improvements of perceptual quality at low bit rates*.

* "High-Fidelity Generative Image Compression", Mentzer, Fabian and Toderici, George D and Tschannen, Michael and Agustsson, Eirikur, Advances in Neural Information Processing Systems, volume 33, 2020.

Google

Can conventional image codecs learn from learned image codec?

Comparisons

Similarities:

- Both conventional and learned image codecs are transform based
- Both have some forms (approximation) of quantization
- Both use context-adaptive probability models (CABAC vs hyper priors)

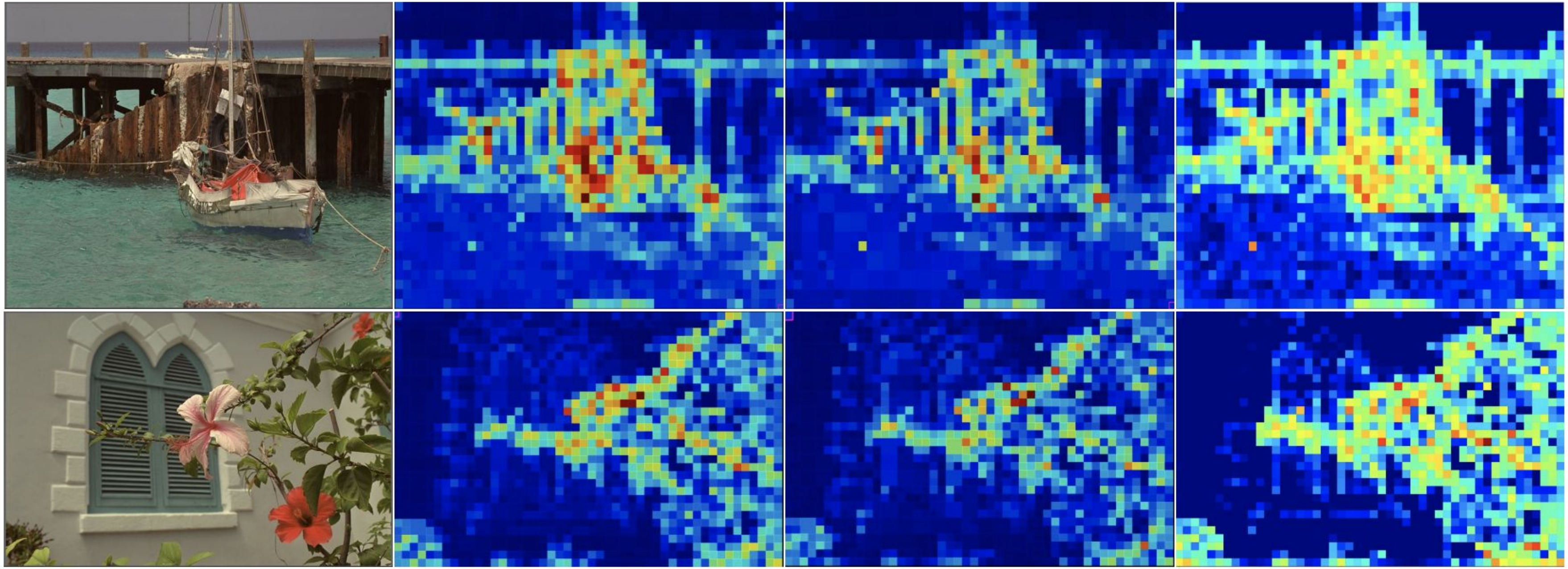
Differences:

- Linear (DCT, ADST) ***vs*** non-linear (CNN) transforms
- Local (conventional block based) ***vs*** global (convolution)
- Pixel domain ***vs*** latent space prediction

Given: Distance in the latent space representation is potentially highly correlated with HVS

Thoughts: What does the rate distribution of the latent space look like?

Rate distribution



Source | Rate distribution of AVIF | Rate distribution of HEVC | Rate distribution of HiFiC latent space

Observations: The rate distribution of the latent representation looks similar to those of conventional codecs!

Claims: Researchers found the latent representation correlates better with HVS than other metrics

Hypothesis

If we let conventional image codecs produce similar rate distribution map, we could get a better perceptual quality

Methods

Step 01	• Obtain rate distribution map of the latent representation from HiFiC
Step 02	• Map rate distribution to quantizers adaptively
Step 03	• Estimate the rate distribution
Step 04	• Dynamic range adjustment

Methods

Step 01

- **Obtain rate distribution map of the latent representation from HiFiC**

- Input image ***X*** of size ***W*** * ***H***
- Obtain the latent representation ***Y*** of size (***W***/16) * (***H***/16) * ***C*** by analysis transform
- Obtain the rate estimate for the ***k***-th channel of ***Y*** from the hyper prior model
- Obtain the rate distribution map ***r*** by aggregating across all channels ***C***

Methods

Step 02

- **Map rate distribution to quantizers adaptively**

- Image codec minimizes: $\min_{\pi} \; D + \lambda R.$
- The distortion ***D*** could be modeled as a function of rate, as well as the uniform quantizer ***Δ*** :

$$D \approx \frac{\Delta^2}{12} \approx A\sigma^2 2^{-2R},$$

- Suppose we encode with a uniform quantizer ***Δ***0, and obtain rate ***R_j*** for block ***j*** :

$$\frac{\Delta_0^2}{12} = A\sigma_j^2 2^{-2R_j}.$$

- Given the rate distribution map ***r*** obtained from HiFiC, with ***r_j*** representing rate for block ***j*** , the desired quantizer ***Δ_j***:

$$\frac{\Delta_j^2}{12} = A\sigma_j^2 2^{-2r_j}$$

- We obtain the adaptive quantizer as a function of rate:

$$\Delta_j = 2^{R_j - r_j} \Delta_0.$$

Methods

Step 03

- **Estimate the rate distribution**

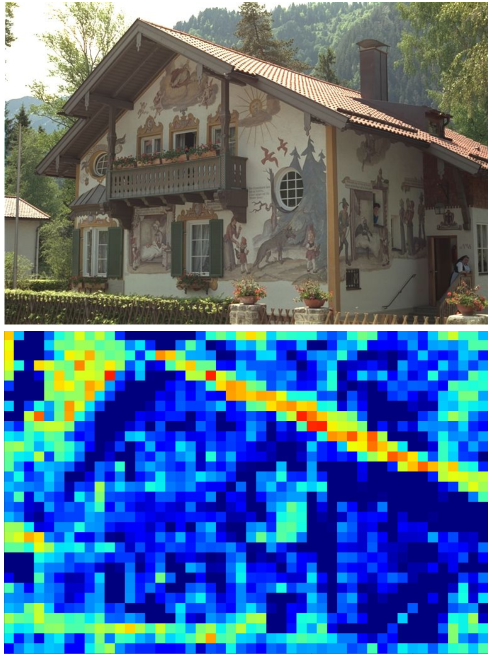
- Apply a fast first pass to get AV1's rate distribution with a uniform quantizer Δ_0
  - DCT only
  - 8x8 transform size
  - Use entropy to estimate rates, not full arithmetic coding

Methods

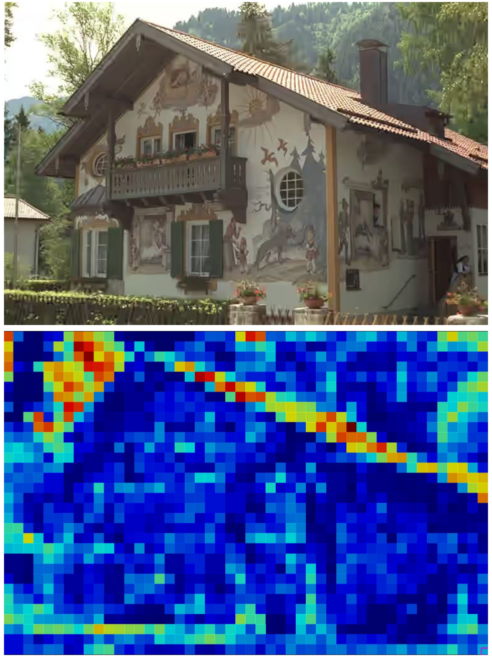
Step 04

- **Dynamic range adjustment**
- Avoid a large quantizer that quantizes all transform coefficients to zero
- The max quantizer is: $\Delta_j^{max} = \lfloor \frac{\Delta_0}{Q_j} \rfloor.$
- The final form of quantizer: $\Delta_j = min(2^{R_j - r_j} \Delta_0, \Delta_j^{max}).$

A visual example

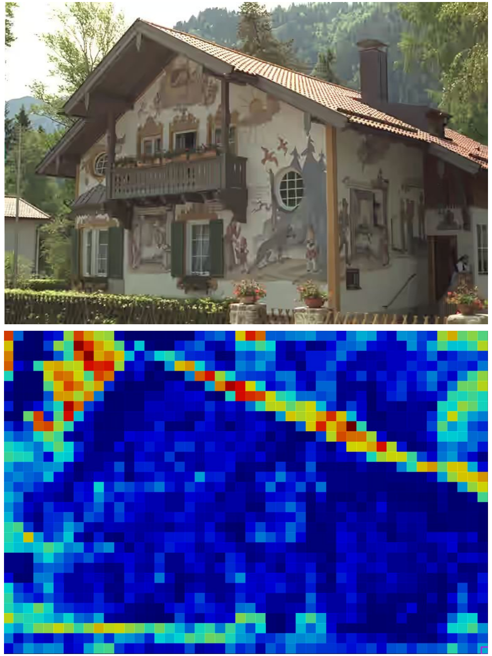


Source (top) and HiFiC rate map (bottom)



Proposed method

0.3bpp



AV1 baseline

0.3bpp

Detail comparisons

Source

Proposed method

0.3bpp

AV1 baseline

0.3bpp

Google

Detail comparisons

Source

Proposed method
0.3bpp

AV1 baseline
0.3bpp

Human evaluation

- We sent random questions to human viewers to compare the reconstructed images coded by two methods, against the original image.
- For each question, human viewers have to pick which one is closer to the original.
- Images are cropped, if the resolution is larger than 1024×1024, so that images can fit on a screen.
- ELO scores are computed to rank different methods.

Human evaluation

- For 0.14 bpp, the proposed method shows the similar performance as the DCM approach [1].
- For 0.3 bpp, human viewers favor the proposed method on both Kodak (68%) and the photographic (55%) image datasets.
- For 0.45 bpp, the proposed method is slightly worse on Kodak dataset (48%), but shows a clear lead on the photographic dataset (59%).

Table 1. ELO scores of human evaluations for perceptual quality.

ELO score	0.14 bpp			0.3 bpp			0.45 bpp		
dataset	DCM	Proposed	Δ / prob	DCM	Proposed	Δ / prob	DCM	Proposed	Δ / prob
Kodak	1997	2011	+14 / 52%	1927	2058	+131 / 68%	2011	1997	-14 / 48%
Photographic [22]	2009	1994	-15 / 48%	1985	2016	+31 / 55%	1965	2031	+66 / 59%

[1]. "Differential contrast based adaptive quantization for perceptual quality optimization in image coding", Jingning Han, Cheng Chen, Frank Galligan, Pascal Massimino, Paul Wilkins, Wan-Teh Chang, Yannis Guyon, Yaowu Xu, and James Bankoski, in 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022, pp. 3026–3030.

Summary

01 Conventional and learned image codecs have similar rate distribution.

02 We provided a general approach to map rate distribution to adaptive quantization.

03 Learned image codec guided adaptive quantization achieves better perceptual quality.