COMPASS: High-Efficiency Deep Image Compression with Arbitrary-scale Spatial Scalability

Jongmin Park[1] Jooyoung Lee[2] Munchurl Kim*,[1]

[1]Korea Advanced Institute of Science and Technology (KAIST), Korea
[2]Electronics and Telecommunications Research Institute (ETRI), Korea
*Corresponding author

: jm.park@kaist.ac.kr
: https://www.viclab.kaist.ac.kr
: https://github.com/ImJongminPark/COMPASS

ICCV23 PARIS

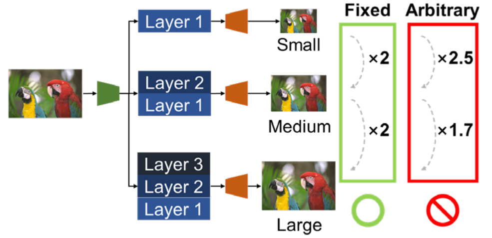
INTRODUCTION

★ Spatially Scalable Image Compression

- Various resolutions of an image are encoded into a single bitstream in a hierarchical manner with multiple layers

★ Limitations of the existing methods

- Tool-based coding : The coding efficiency is insufficient compared to the single-layer coding
- NN-based coding : Only one fixed scale factor of 2 is used between adjacent layers

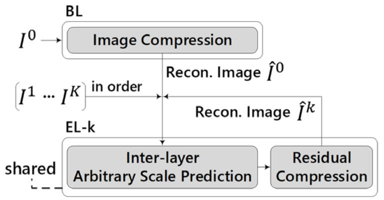


< Limitation of the existing NN-based methods >

★ COMPASS (our contributions)

- The first NN-based spatially scalable image compression method that supports arbitrary scale factors
- Significantly better coding efficiency than the existing methods (comparable or even better coding efficiency compared to the single-layer coding)
- New adoption of INR (implicit neural representation)-based inter-layer arbitrary scale prediction

CONCEPT

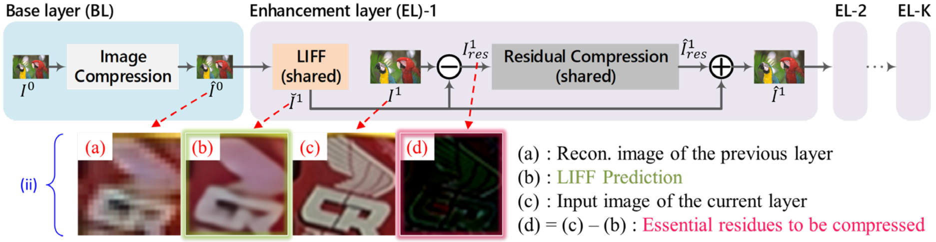


- Base Layer (BL) encodes the lowest resolution image
- Enhancement Layers (ELs) sequentially encode higher resolution images (Parameters of all the ELs are shared)

PROPOSED METHOD

★ Overall Network of COMPASS

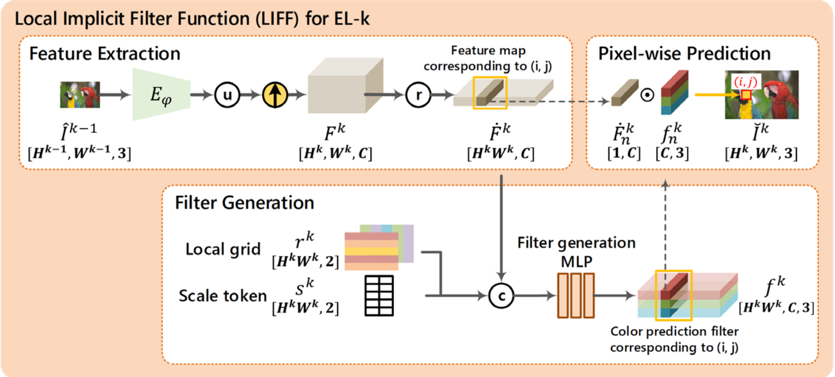
i) One or more ELs operate in an iterative manner
ii) LIFF module effectively predicts/reduces the redundancy between adjacent layers for arbitrary scales
iii) Residual Compression module only encodes the essential residues with high coding efficiency



(a) : Recon. image of the previous layer
(b) : LIFF Prediction
(c) : Input image of the current layer
(d) = (c) – (b) : Essential residues to be compressed

★ LIFF: Local Implicit Filter Function

- Inter-layer arbitrary scale prediction based on implicit neural representation (inspired by LIIF and Meta-SR)
- 3 stages : ① Feature Extraction, ② Filter Generation, ③ Pixel-wise Prediction



- Local grid : a normalized relative coord.

$$r^k(i,j) = p^k(i,j) - p^{k-1}(i',j')$$

*pixel coord. for the current layer; *corresponding pixel coord. for the prev. layer

- Scale token : a height/width ratio

$$s^k(i,j) = (2 \cdot \frac{H^{k-1}}{H^k}, 2 \cdot \frac{W^{k-1}}{W^k})$$

OPTIMIZATION

★ Combined RD loss function (CRD loss)

- Summation of RD losses for all ELs
- The same λ value to maintain the R-D balance over layers

$$L = \sum_{k=1}^{K} R^k + \lambda \cdot D^k \qquad R^k = H^k(\tilde{y}^k|\tilde{z}^k) + H^k(\tilde{z}^k) \qquad D^k = MSE(\hat{I}^k, I^k)$$

QUANTITATIVE RESULTS

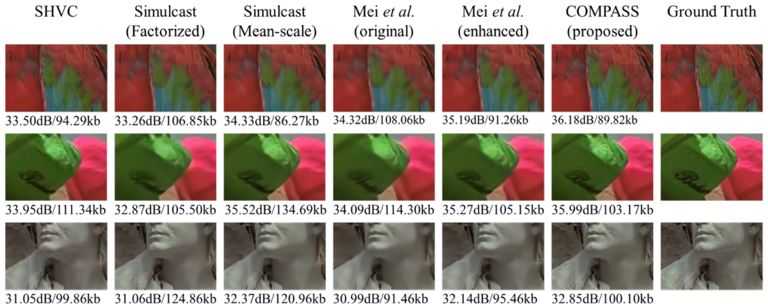
★ BD-rate (↓) of our COMPASS over the various method (Negative value : BD-rate gain of ours)

Methods	BD-rate↓ for Two-layer scalable coding					BD-rate↓ for Three-layer scalable coding				
	Scale Factors of final EL (vs. BL)					Scale Factors of final EL (vs. BL)				
	1.2×	1.6×	2.0×	2.4×	2.8×	2.4×	2.8×	3.2×	3.6×	4.0×
SHVC	-53.88%	-42.58%	-35.87%	-26.24%	-22.34%	-58.33%	-51.51%	-46.72%	-43.65%	-33.34%
Simulcast (Mean-Scale)	-44.04%	-31.40%	-16.29%	-16.35%	-12.38%	-49.85%	-42.20%	-35.33%	-30.81%	-22.90%
Mei et al. (original)	-36.26%	-29.44%	-28.20%	-33.19%	-36.31%	-47.17%	-38.73%	-34.34%	-33.56%	-32.12%
Mei et al. (enhanced)	-29.09%	-17.45%	-13.52%	-20.63%	-23.94%	-38.23%	-26.46%	-19.70%	-17.08%	-14.23%
Single-layer (Mean-Scale)	-8.19%	-3.70%	8.80%	0.31%	0.94%	-6.60%	-4.23%	-1.25%	-0.74%	4.74%

*Kodak Dataset

VISUAL RESULTS

★ Visual Comparison w/ fixed scale of 2



★ Visual Comparison w/ arbitrary scale factors