Learning True Rate-Distortion-Optimization for End-To-End Image Compression

Friedrich-Alexander-Universität
Technische Fakultät

Fabian Brand, Kristian Fischer, and André Kaup

Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Cauerstr. 7, 91058 Erlangen, Germany

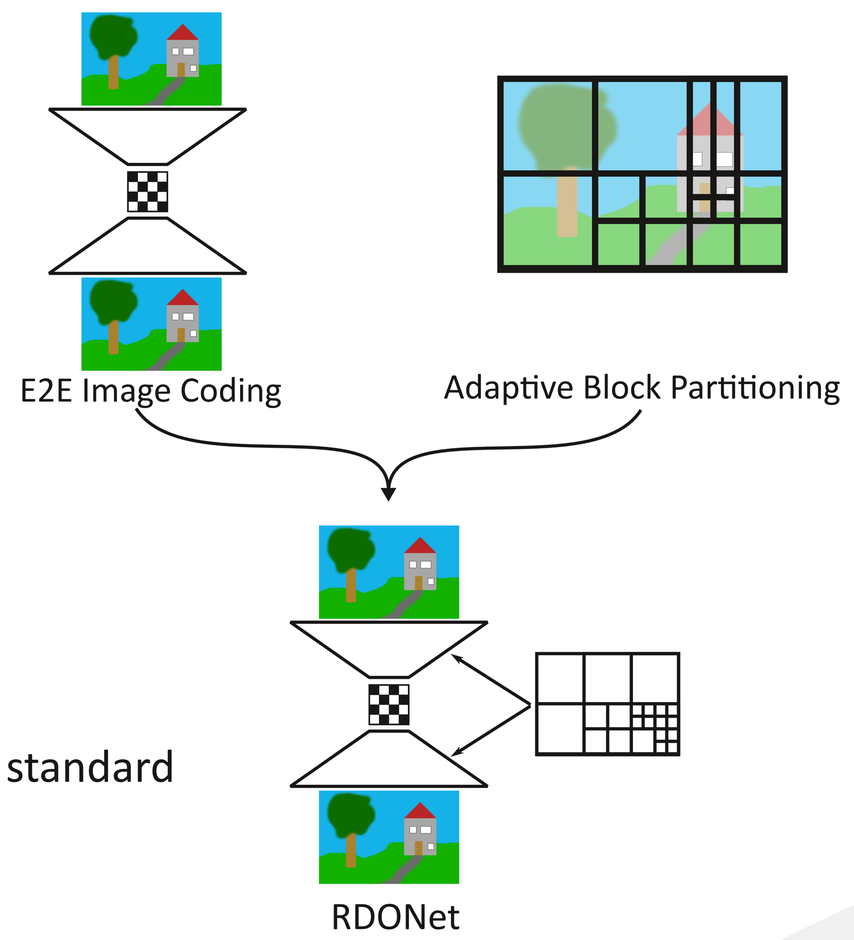
1. Motivation

End-To-End Trained Image Coders [1,2]

- Compression into latent space with encoder network e : $\boldsymbol{f} = e(\boldsymbol{x})$
- Decompression to image with decoder network d : $\hat{\boldsymbol{x}} = d(\hat{\boldsymbol{f}})$
- Typically, one fixed function for compression and decompression
- RDONet [3]: Network that allows dynamic rate-distortion-opmization

Hybrid coders [4]

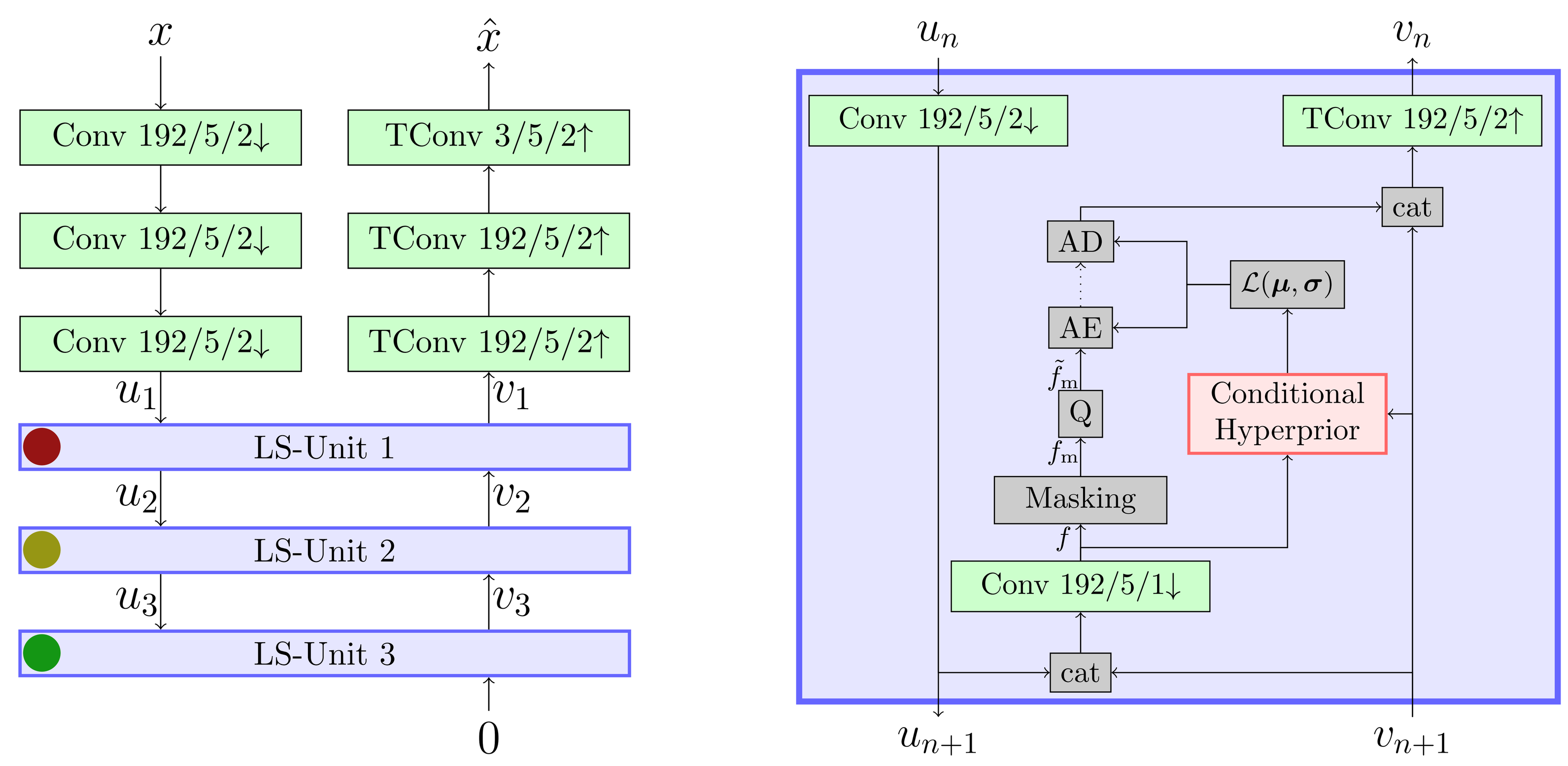
- Adaptive block partitioning
- Granularity is chosen adaptively
- Stationary areas are transmitted in large chunks
- Rate distortion optimized encoding



Proposal

- Traning procedure which approximates RDO during training
- Low-complexity encoding mode by zero-pass RDO
- On average, 23% bit savings compared to standard autoencoder

2. RDONet [3]



RDONet

- Hierarchical structure: Compression in three different granularities possible
- Each Latent Space Unit (LS-Unit) transmits one level
- Levels can be controlled externally block-wise
- Redundancy reduction between levels with conditional hyperprior

Training:

- No rate-distortion-optimization possible during training
- Choose levels randomly
- Misalignment between training and inference

3. Proposed Method

Training content adaptive masking difficult

- Masking operator non-differentiable
- RDO at training time computationally infeasible

Inference Complexity

- RDO Search requires multiple coding runs
- 2-pass RDO: 12 coding runs per 64x64 block

Variance Based Mask Estimation

- Based on variance of pixels in block
- Split block if variance exceeds threshold
- Generate three levels with different thresholds

Training procedure

- Training with random masks for 2000 epochs
    - All levels can compress general image content
- Training with variance-based masks for 600 epochs
    - Levels can specialize

Fast encoding

- Initialize RDO with variance-based mask
- Faster convergence
- "Zero-pass" RDO: Compress with estimated mask

Example image "wojciech-szaturski-3611" and estimated mask. Red: finest latent space resolution; Green: coarsest resolution. Structures with fine details are compressed with finest latent space resolution.

5. Conclusion

RDONet became feasible compression network

- Large improvement by specializing layers on content and mask
- Increased rate savings from 7.7% to 27.3%
- Fast rate-distortion-optimization possible
- Half the number of coding passes obtains almost same results
- Zero-pass RDO available which saves 23.6% rate
- Successfully transferred great strength of block-based coding to deep-learning-based methods

RDONet [3]

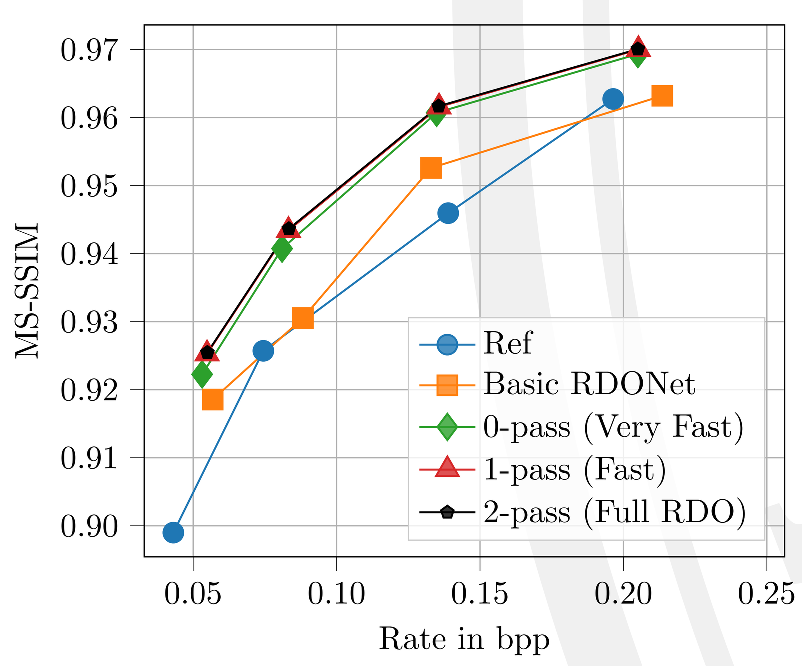
25.86 dB/0.16 bpp

Proposed RDONet

25.95 dB/0.14 bpp

[1] J. Ballé, et al. "End-to-end optimized image compression," in ICLR 2017
[2] D. Minnen, et al. "Joint autoregressive and hierarchical priors for learned image compression," in NeurIPS, 2018
[3] F. Brand, et al. "Rate-distortion optimized learning-based image compression using an adaptive hierachical autoencoder with conditional hyperprior," in Proc. CVPRW, 2021.
[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," TCSVT, 2012.

4. Results



Rate-distortion curve comparing proposed RDONet with basic RDONet and conventional autoencoder (Ref)

Test Conditions

- Train networks on CLIC21 training set, DIV2K and TECNICK
- Evaluate network on CLIC21 test set
- Compare against RDONet trained with random masks [3] and conventional autoencoder with hyperprior and context model [2]

Extended training method

- Proposed training method superior
- Performance about 20% better than previous method

Fast RDO

- RDO with initialization outperforms RDO with static initialization
- 1-pass RDO sufficient
- Very fast mode (zero-pass) saves 23.6% rate

Network	Basic RDONet [3]		RDONet-Var (Ours)				
RDO-Init	Static		Static		Variance Adaptive		
RDO-Passes	1	2	1	2	0	1	2
Best Case	-18.9%	-22.5%	-43.7%	-45.3%	-36.6%	-44.4%	-45.2%
Worst Case	+7.5%	+3.5%	-11.9%	-12.5%	-6.67%	-11.9%	-12.5%
Average	-4.1%	-7.7%	-23.3%	-25.0%	-23.6%	-26.8%	- 27.3%
RDO-Complexity	$6 \cdot N_{64}$	$12 \cdot N_{64}$	$6 \cdot N_{64}$	$12 \cdot N_{64}$	0	$6 \cdot N_{64}$	$12 \cdot N_{64}$

Bjøntegaard Delta Rate savings compared to classical compressive autoencoder. RDO Complexity is given as network runs per image. N_{64} is the number of 64x64 blocks in that image.