SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

CONDITIONAL AND RESIDUAL METHODS IN SCALABLE CODING FOR HUMANS AND MACHINES

Anderson de Andrade, Alon Harell, Yalda Foroutan, and Ivan V. Bajić

School of Engineering Science
Simon Fraser University
Burnaby, BC, Canada

July 10, 2023

GOAL

We would like to ***compose*** *the information required for tasks in a* ***scalable*** *fashion, in which base representations are* ***shared*** *among multiple tasks and only* ***incremental*** *amounts of information are required for more specific tasks.*

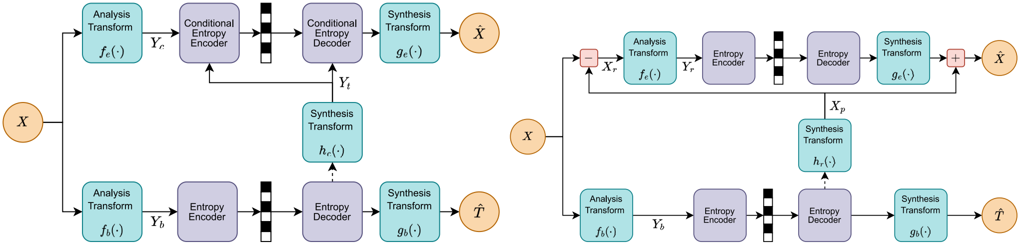
CONTRIBUTIONS

1. Conditional and residual approaches in *scalable image coding for humans and machines*[1]
2. Upper and lower baselines
3. Entropy model for conditional coding
4. Empirical results on semantic image segmentation and object detection

[1] H. Choi and I. V. Bajić. "Scalable Image Coding for Humans and Machines". In: *IEEE TIP* 31 (2022), pp. 2739–2754.

- Dedicated and shared representations are concatenated and used as input for a target task
- Results show considerable redundancy[2]

[2] E. Ozyilkan et al. "Learned Disentangled Latent Representations for Scalable Image Coding for Humans and Machines". In: *arXiv 2301.04183* (2023); Choi and Bajić, "Scalable Image Coding for Humans and Machines".



(a) Conditional

(b) Residual

Figure 1: Overall architecture of the conditional and residual methods.

LOSSLESS BOUNDS FOR CONDITIONAL CODING

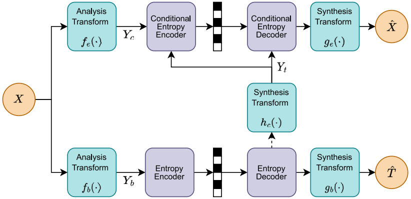
- We model $H(Y_b) + H(Y_c|Y_t)$
- Lower bound:

$$H(Y_c) \leq H(Y_b) + H(Y_c|Y_t). \qquad (1)$$

- Tight when $H(Y_b) = I(Y_c; Y_t)$
- Upper bound:

$$H(Y_b) + H(Y_c|Y_t) \leq H(Y_b) + H(Y_c). \qquad (2)$$

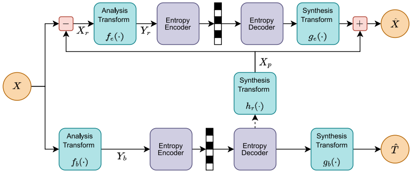
- Tight when $I(Y_c; Y_t) = 0$



Conditional coding is an upper bound of residual coding[3]:

$$H(X|X_p) \leq H(X_r) = H(X|X_p) + I(X_p; X_r). \quad (3)$$

$I(X_p; X_r)$ acts as a penalty term.

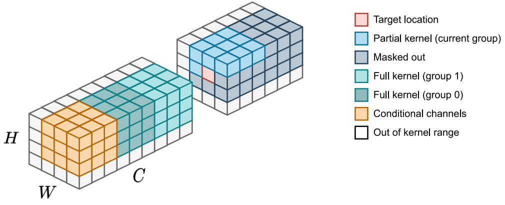


[3] F. Brand, J. Seiler, and A. Kaup. "On Benefits and Challenges of Conditional Interframe Video Coding in Light of Information Theory". In: *PCS*. 2022, pp. 289–293.

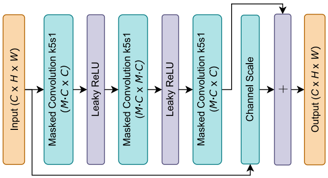
BASELINES FOR SCALABLE CONDITIONAL AND RESIDUAL CODING

- Rate-distortion bounds motivated by lossless bounds
- Upper baseline - Y_e generated without side information: $\hat{H}(Y_e) = R(D_{Y_e})$
- Lower baseline as $\hat{H}(Y_b) + \hat{H}(Y_e)$
- Identical baselines for both approaches since $H(Y_c|Y_t) = H(Y_r)$ is easier to achieve

ENTROPY MODEL FOR CONDITIONAL CODING



(a) Convolution mask

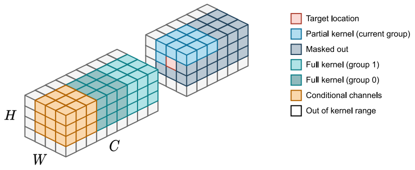


(b) CNN block

Figure 2: Entropy model overview.

ENTROPY MODEL FOR CONDITIONAL CODING

- Model spatial-dimensional dependencies using an auto-regressive CNN
- Scaled residual connections and kernels sizes larger than 1 throughout
- Group channels with fixed size K and process them in parallel
- All reachable locations in the previous groups as context [4]



[4] D. Minnen and S. Singh. "Channel-Wise Autoregressive Entropy Models for Learned Image Compression". In: *IEEE ICIP*. 2020, pp. 3339–3343.

- Small reconstruction penalty added to base layer RD objective:
  - Allows reconstruction task to use more available information

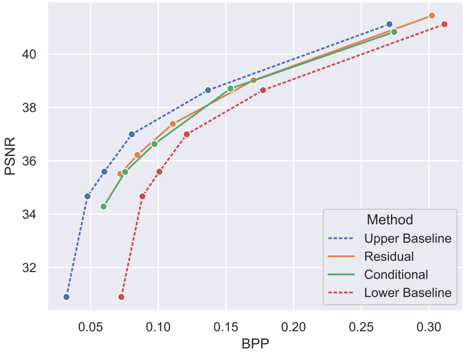
$$\mathcal{L}_b = D_b + \lambda_b \hat{H}(Y_b) + \beta\, \mathbb{E}[d_e(\hat{h}_r(Y_b), X)] \tag{4}$$

- Standard rate-distortion loss function for enhancement:

$$\mathcal{L}_c = D_c + \lambda_e \hat{H}(Y_c|Y_t), \qquad \mathcal{L}_r = D_r + \lambda_r \hat{H}(Y_r) \tag{5}$$

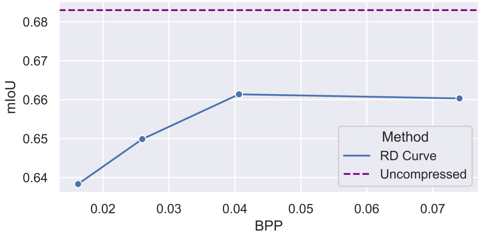
EXPERIMENTS

- Semantic segmentation: Cityscapes; object detection: COCO 2017
- Train the base independently: $f_b(\cdot)$ and $g_b(\cdot)$ under various λ_b
- Pick single (good) base representation Y_b for conditional/residual setting
- Upper baseline: reconstruction task with no side information
- Lower baseline: upper baseline + base rate $\hat{H}(Y_b) + \hat{H}(Y_e)$



(a) Scalable reconstruction

	Conditional	Residual
BD-Rate	-16.56%	-14.6%
Utilization	43.01%	37.91%

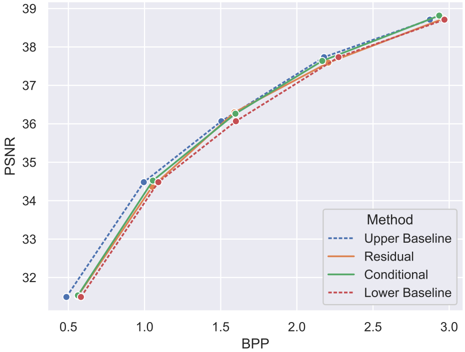


(b) Semantic segmentation

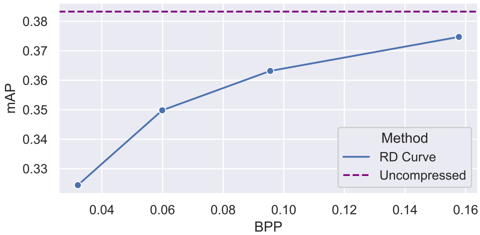
Figure 3: Rate-distortion curves for Cityscapes.

EXPERIMENTS

	Conditional	Residual
BD-Rate	-4.14%	-2.47%
Utilization	49.24%	29.32%



(a) Scalable reconstruction



(b) Object detection

Figure 4: Rate-distortion curves for COCO.

CONCLUSION

- Conditional and residual coding perform similarly using proposed design
- RD curves between presented baselines (operational bounds)
- Proposed conditional entropy model and small reconstruction penalty leads to improved utilization of side information

DERIVATIONS

Lossless lower bound:

$$\begin{aligned} H(Y_c) &\leq H(Y_c) + H(Y_t|Y_c) = H(Y_c, Y_t) \\ &= H(Y_t) + H(Y_c|Y_t) \leq H(Y_b) + H(Y_c|Y_t). \end{aligned} \tag{6}$$

Lossless upper bound:

$$\begin{aligned} H(Y_b) + H(Y_c|Y_t) &= H(Y_b) + H(Y_c) - I(Y_c; Y_t) \\ &\leq H(Y_b) + H(Y_c). \end{aligned} \tag{7}$$

Residual and conditional coding:

$$H(X|X_p) = H(X_r + X_p|X_p) = H(X_r|X_p) \tag{8}$$

$$= H(X_r) - I(X_p; X_r) \leq H(X_r) = H(X|X_p) + I(X_p; X_r). \tag{9}$$