SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

TOWARDS TASK-COMPATIBLE COMPRESSIBLE REPRESENTATIONS

Anderson de Andrade, Ivan Bajić

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

August 14, 2025

MOTIVATION

- Representations induced by choices: task, architecture, co-parameters
- Relevant information cannot be fully extracted by different processes
- Critical in multi-task learning and scalable coding (side-information)

Referred as co-adaptation in transfer learning:

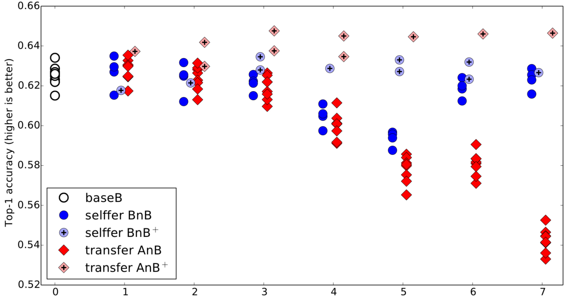


Figure 1: "Features interact with each other in a complex or fragile way such that this co-adaptation could not be relearned" (credit: Yosinski et al.[1])

[1] J. Yosinski et al. "How transferable are features in deep neural networks?" In: *NIPS*. 2014.

- Predictive $\mathcal{V}$-information[2] considers limited synthesis transforms $\mathcal{G}$
- Choose analysis transform that optimizes $\mathcal{V}$-information between representation and output:

$$\sup_{\mathbf{f} \in \mathcal{F}} I_{\mathcal{V}}(\mathbf{f}(X) \to Z)$$

- Predictive set $\mathcal{V}$ is parameterized by $\mathcal{G}$
- For unrestricted $\mathcal{F}$, attaining $I(X;Z)$ limited by expressiveness of $\mathcal{G}$ and $\mathcal{V}$

[2] Yilun Xu et al. “A Theory of Usable Information under Computational Constraints”. In: *ICLR*. 2020.

INSIGHT

- Several analysis transforms $\mathcal{F}_b^*$ can achieve desired RD for task Z_b
- Selected analysis transform $\mathbf{f}_b^*$ can underperform in secondary task Z_e:

$$I_{\mathcal{V}_e}(\mathbf{f}_b^*(X) \rightarrow Z_e) \leq \arg\max_{\mathbf{f} \in \mathcal{F}_b^*} I_{\mathcal{V}_e}(\mathbf{f}(X) \rightarrow Z_e)$$

- Achieve similar RD on base while optimizing for secondary tasks

PROPOSAL

- New objective for base:

$$\sup_{\mathbf{f}\in\mathcal{F}} \left\{ I_{\mathcal{V}_b}(\mathbf{f}(X) \to Z_b) + \beta I_{\mathcal{V}_a}(\mathbf{f}(X) \to Z_a) \right\}$$

- $\mathcal{V}_a$ is a predictive set given by synthesis transforms $\mathcal{G}_a$
- Propose reconstruction as to generalize to other tasks
- Small reward β on capacity of $\mathbf{f}$ to reconstruct input X

PROPOSAL

Choice of reconstruction task $Z_a = X$ and simple $\mathcal{G}_a$:

- Partial reconstruction can be subsumed by secondary tasks
- Task models often designed to work on original input space
- Trained in an unsupervised manner

- Set $I_{\mathcal{V}_b}(Y_b \to Z_b)$ as information bottleneck[3] for rate-constrained Y_b
- Define $\mathcal{V}_b$ according to task loss Z_b
- Assume Normal distribution centered around $\hat{X} = \mathbf{g}_a(\hat{Y}_b)$ for $\mathcal{V}_a$
- Base loss:

$$\mathcal{L}_b(\psi_b, \phi_b, \theta_b, \theta_a) = \lambda_b \mathbb{E}[d_b(\hat{Z}_b, Z_b)] + \hat{h}(\hat{Y}_b; \phi_b) + \beta\, \mathbb{E}[\|\hat{X} - X\|]$$

- $\beta \ll \lambda_b$

[3] N. Tishby, F. C. N. Pereira, and W. Bialek. “The information bottleneck method”. In: *CoRR* physics/0004057 (2000).

- Information bottleneck for enhancement representation
- Entropy estimates conditional on base
- Enhancement loss:

$$\mathcal{L}_e(\psi_e, \phi_e) = \lambda_e \mathbb{E}[d_e(\hat{Z}_e, Z_e)] + \hat{h}(\hat{Y}_e|\hat{Y}_b; \phi_e)$$

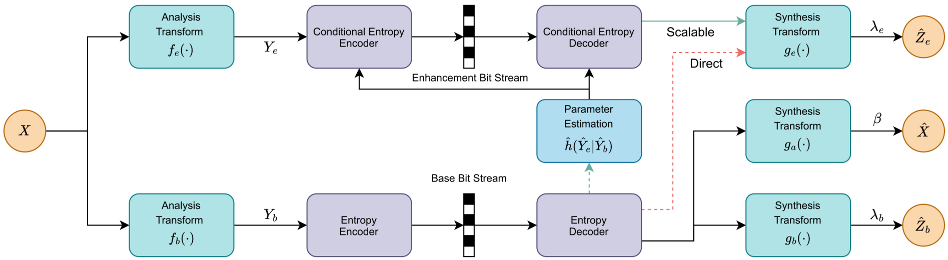


Figure 2: Architecture overview.

- Analysis and synthesis transform in ELIC[4]
- Append model to synthesis transform according to task
- Gaussian entropy model[5]
- Quantization modelled by STE instead of uniform noise

[4] D. He et al. "ELIC: Efficient Learned Image Compression with Unevenly Grouped Space-Channel Contextual Adaptive Coding". In: *IEEE CVPR*. 2022.
[5] J. Ballé et al. "Variational image compression with a scale hyperprior". In: *ICLR*. 2018.

EXPERIMENTS

- Reconstruction with object detection base
- Reconstruction with depth estimation base
- Semantic segmentation with depth estimation base
- Baseline with $\beta = 0$, proposal with $\beta = 0.01$
- Rate of scalable approach is sum of base and enhancement rates
- Standalone method with no side-information as lower bound
- Base task trained separately and frozen

EXPERIMENTS

- Object detection on COCO 2017
- Depth estimation and semantic segmentation on Cityscapes
- Faster R-CNN[6] with ResNet-50[7] for object detection
- LRASPP with MobileNetV3[8] for depth estimation
- DeepLabV3[9] with ResNet-50 for segmentation

---

[6] S. Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE TPAMI* 39 (2017).
[7] K. He et al. "Deep Residual Learning for Image Recognition". In: *IEEE CVPR*. 2016.
[8] A. Howard et al. "Searching for MobileNetV3". In: *IEEE ICCV*. 2019.
[9] L.-C. Chen et al. "Rethinking Atrous Convolution for Semantic Image Segmentation". In: *CoRR* abs/1706.05587 (2017).

RESULTS

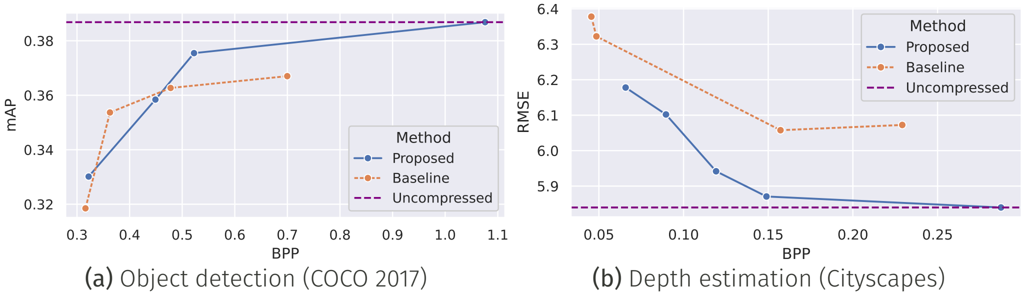
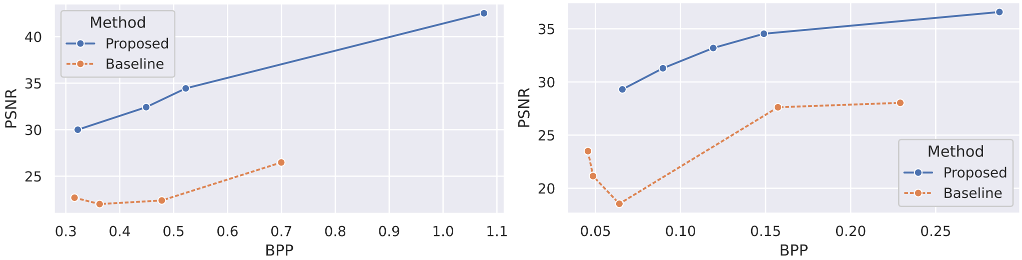
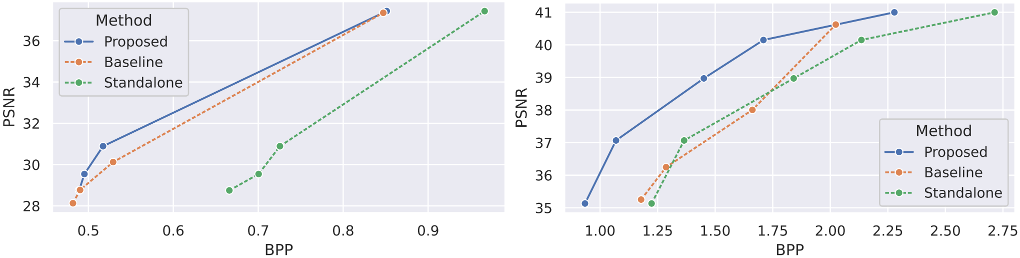


Figure 3: Rate-distortion of base tasks. 1.3% and 75% BD-rate.



(a) Direct with object detection (COCO 2017) (b) Direct with depth estimation (Cityscapes)

Figure 4: Rate-distortion of input reconstruction from the base representation directly. No-overlap and 85% BD-rate.



(a) Scalable with detection side-information (COCO 2017)

(b) Scalable with depth side-information (Cityscapes)

Figure 5: Rate-distortion of input reconstruction conditional on base representation. 10% and 22% BD-rate w/r baseline.

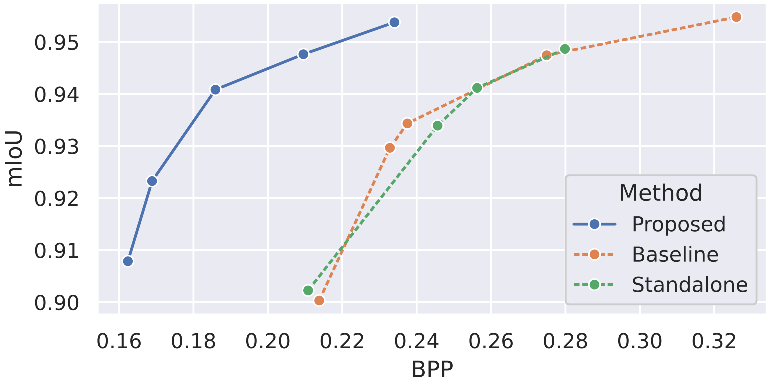


Figure 6: Rate-distortion with depth estimation as base and semantic segmentation as enhancement on Cityscapes. 26% BD-rate w/r baseline. Training set due to dataset size.

RESULTS

(a) Preview

(b) Original

Figure 7: Preview of object detection base at 17.23 PSNR.

(a) Preview

(b) Original

Figure 8: Preview of depth estimation base at 30.45 PSNR.

CONCLUSIONS

- Baseline unable to extract useful information in some experiments
- Base RD lift suggests co-adaptation between Y_b and entropy/task models

CONCLUSIONS

- $\mathcal{V}$-information framework provides possible interpretation
- Auxiliary task is sufficiently non-specific and $\mathcal{V}_a$ is sufficiently simple to reduce co-adaptation and generalize to other tasks