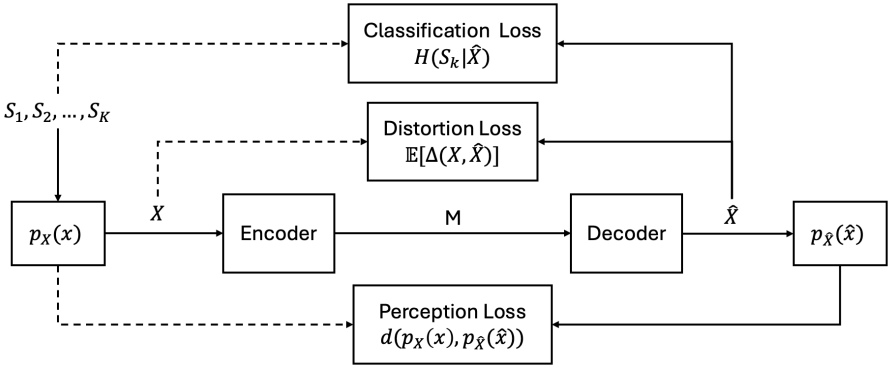
Universal Representations for Classification-enhanced Lossy Compression

Nam Nguyen

School of Electrical Engineering and Computer Science
Oregon State University

April 14, 2025

Task-oriented Lossy Compression



Source & Target labels:

- Source: $X \sim p_X(x)$
- Target labels: $S_1, \cdots, S_K$

Lossy Compression:

- Encoder: source $X \mapsto$ message M.
- Decoder: message $M \mapsto$ data $\hat{X}$, *for task-oriented downstream applications.*

Rate-Distortion-Perception/Classification Functions

Definition

The rate-distortion-perception function:

$$R(D,P) = \min_{p_{\hat{X}|X}} \quad I(X;\hat{X}) \tag{1a}$$

$$\text{s.t.} \quad \mathbb{E}[\Delta(X,\hat{X})] \leq D, \tag{1b}$$

$$d(p_X, p_{\hat{X}}) \leq P. \tag{1c}$$

Definition

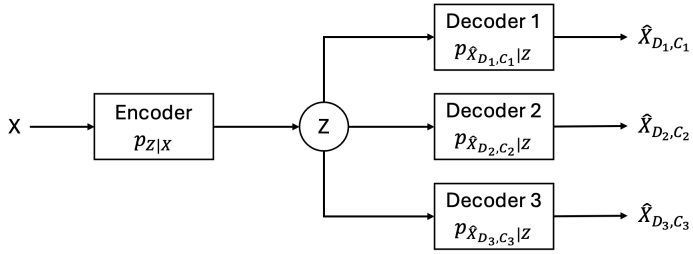
The rate-distortion-classification function:

$$R(D,C) = \min_{p_{\hat{X}|X}} \quad I(X;\hat{X}) \tag{2a}$$

$$\text{s.t.} \quad \mathbb{E}[\Delta(X,\hat{X})] \leq D, \tag{2b}$$

$$H(S|\hat{X}) \leq C. \tag{2c}$$

Universal Rate-Distortion-Classification Function



Definition

Let Z represent X through a transform $p_{Z|X}$. Define:

$$\mathcal{P}_{Z|X}(\Theta) = \{p_{Z|X} : \exists p_{\hat{X}_{D,C}|Z},\ (D,C) \in \Theta,\ \mathbb{E}[\Delta(X, \hat{X}_{D,C})] \leq D,\ H(S|\hat{X}_{D,C}) \leq C\},$$

The universal rate-distortion-classification function:

$$R(\Theta) = \inf_{p_{Z|X} \in \mathcal{P}_{Z|X}(\Theta)} I(X;Z). \tag{3}$$

Rate Penalty

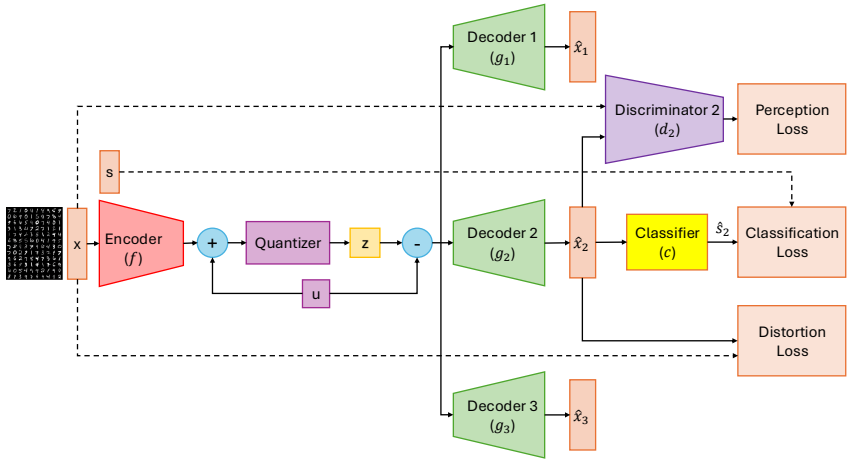
Definition

The rate penalty incurred by meeting all constraints in Θ with **fixed encoder**:

$$A(\Theta) = R(\Theta) - \sup_{(D,C)\in\Theta} R(D,C). \tag{4}$$

- Ideally, $A(\Theta) = 0$, i.e., a **single encoder** suffices for the entire (D, C) tradeoffs.

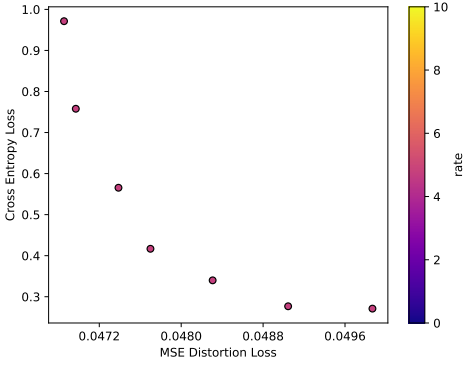
Universal Representation for Image Lossy Compression



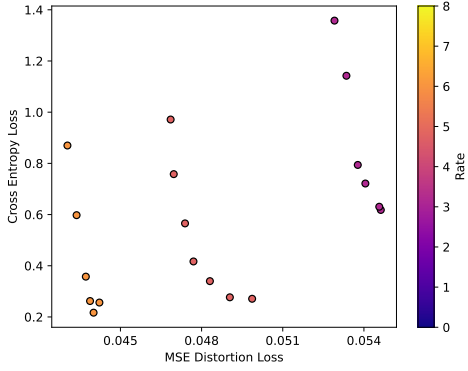
- Dataset: MNIST
- Pre-trained classifier (c)
- Train the encoder (f), decoder (g), and GAN-based discriminator (d):

$$\mathcal{L}_{RDC} = \mathbb{E}[\|X - \hat{X}\|^2] + \lambda_c \mathsf{CE}(s, \hat{\mathsf{s}}), \qquad \mathcal{L}_{RDP} = \mathbb{E}[\|X - \hat{X}\|^2] + \lambda_p W_1(p_X, p_{\hat{X}}).$$

Experimental Results



(a) Classification-distortion-rate function with $R = 4.75$.



(b) Classification-distortion-rate functions with vary rates.

Experimental Results

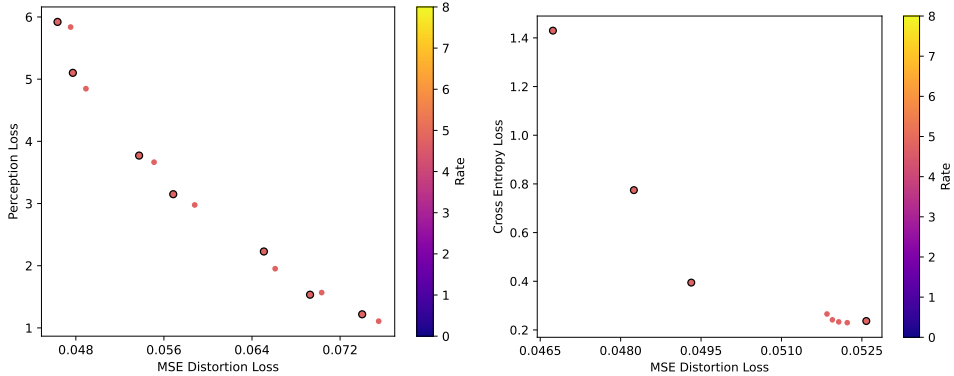


Figure: Perception (classification)-distortion-rate functions with $R = 4.75$. Points with black outlines indicate results from end-to-end trained encoder-decoder models for a particular perception-distortion target. Other points denote losses for universal models, in which decoders are trained separately using a fixed encoder.