Cross-Domain Lossy Compression via Rate- and Classification-Constrained Optimal Transport

Nam Nguyen, Thinh Nguyen, and Bella Bose

Oregon State University, Corvallis, OR, USA

April 7, 2026

Background

Unsupervised Restoration [W. Wang et al. 2023]

- No degradation model; only **unpaired** degraded/clean data.
- Learn a mapping by matching **output distribution** to clean images using GAN.
- Interpreted as an **optimal transport (OT)** problem.

Background

Unsupervised Restoration [W. Wang et al. 2023]

- No degradation model; only **unpaired** degraded/clean data.
- Learn a mapping by matching **output distribution** to clean images using GAN.
- Interpreted as an **optimal transport (OT)** problem.

Cross-domain Lossy Compression [Liu et al. 2022]

- Introduce a **rate constraint** into OT (entropy-constrained OT).
- View as compression between **different source & target distributions**.

---

[1] W. Wang, F. Wen, Z. Yan, and P. Liu, "Optimal Transport for Unsupervised Denoising Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[2] H. Liu, G. Zhang, J. Chen, and A. Khisti, "Cross-Domain Lossy Compression as Entropy Constrained Optimal Transport," *IEEE Journal on Selected Areas in Information Theory*, 2022.

Motivation and Objective

- **Classification is everywhere:** Modern systems rely on classifiers and deep neural networks (DNNs).

Motivation and Objective

- **Classification is everywhere:** Modern systems rely on classifiers and deep neural networks (DNNs).
- **Beyond distortion:** In source compression, we should also preserve *classification performance* (C) [Nguyen et al. 2025; Y. Wang et al. 2025].

$$R(D) \longrightarrow R(D, C), \quad D(R, C)$$

Motivation and Objective

- **Classification is everywhere:** Modern systems rely on classifiers and deep neural networks (DNNs).
- **Beyond distortion:** In source compression, we should also preserve *classification performance* (C) [Nguyen et al. 2025; Y. Wang et al. 2025].

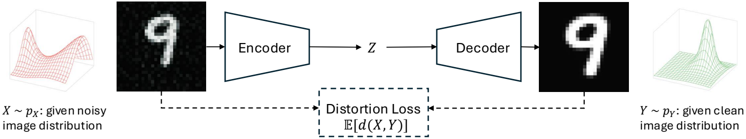
$$R(D) \longrightarrow R(D, C), \quad D(R, C)$$

- **Our goal:**
  - Introduce **rate + classification constraints** to the OT framework.
  - Enable **task-oriented compression** across domains.
  - Jointly address **restoration + compression + classification performance**.

[1] Y. Wang, Y. Wu, S. Ma, Y.-J. Zhang, and A. Zhang, "Task-Oriented Lossy Compression With Data, Perception, and Classification Constraints," *IEEE Journal on Selected Areas in Communications*, 2024.

[2] N. Nguyen, T. Nguyen, T. Nguyen, and B. Bose, "Universal Rate-Distortion-Classification Representations for Lossy Compression," *Information Theory Workshop*, 2025.

Unsupervised Restoration as Optimal Transport



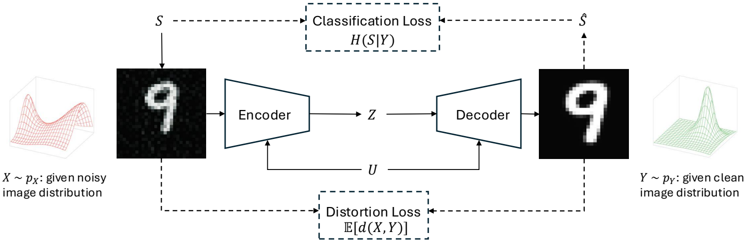
Definition (Optimal Transport)

Let $\Gamma(p_X, p_Y)$ denote the set of all joint distributions $p_{X,Y}$ with marginals p_X and p_Y.

$$D(p_X, p_Y) = \inf_{p_{X,Y} \in \Gamma(p_X, p_Y)} \mathbb{E}[d(X,Y)],$$

where $d(\cdot,\cdot)$: given distortion (cost) function, $p_{X,Y} \in \Gamma(p_X, p_Y)$: a (stochastic) transport plan.

One-Shot Lossy Compression as Constrained Optimal Transport



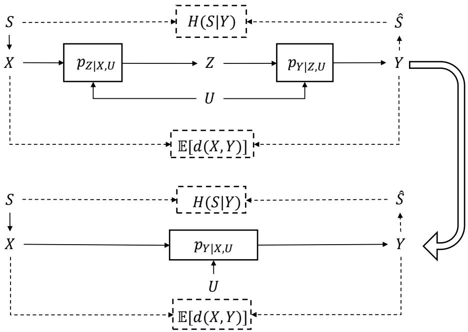
Definition (Constrained Optimal Transport)

Let $X \sim p_X$ $Y \sim p_Y$, $S \sim p_S$, $p_{X,S}$, $p_{U,X,Z,Y} = p_U\, p_X\, p_{Z|X,U}\, p_{Y|Z,U} \in M(p_X, p_Y)$, where $M(p_X, p_Y)$ is the set of $p_{U,X,Z,Y}$ with marginal p_X and p_Y, The constrained optimal transport distortion for given rate R, classification loss C, and shared randomness U is:

$$D(R, C, p_X, p_Y) = \inf_{p_{U,X,Z,Y}} \mathbb{E}[d(X,Y)]$$

$$\text{s.t.} \quad H(Z|U) \le R, \quad H(S|Y) \le C,$$

One-Shot Lossy Compression as Constrained Optimal Transport



Theorem

Let the joint $p_{U,X,Y} = p_U\, p_X\, p_{Y|X,U} \in M(p_X, p_Y)$*, where* $M(p_X, p_Y)$ *is the set of* $p_{U,X,Y}$ *with marginal* p_X *and* p_Y*, then the constrained optimal transport admits the representation:*

$$D(R, C, p_X, p_Y) = \inf_{p_{U,X,Y} \in M(p_X, p_Y)} \mathbb{E}[d(X,Y)]$$

$$\text{s.t.} \quad H(Y|X,U) = 0, \quad I(X;U) = 0,$$

$$H(Y|U) \le R, \quad H(S|Y) \le C.$$

Case Study: Bernoulli

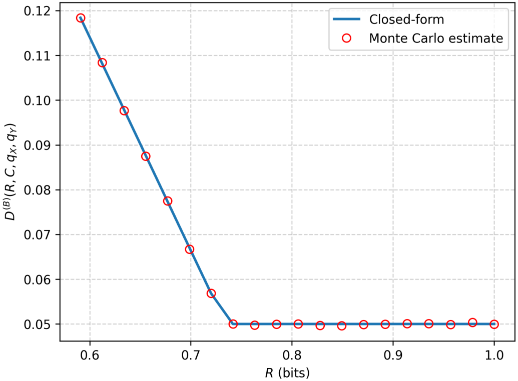
Theorem

Let $X \sim$ *Bern*(p_X), $Y \sim$ *Bern*(p_Y), *and a classification variable* S *with the binary symmetric joint distribution given by* $S = X \oplus S_1$ *where* $S \sim$ *Bern*(p_S) *and* $S_1 \sim$ *Bern*(p_{S_1}) $(0 \leq p_X, p_S, p_{S_1} \leq \frac{1}{2})$*. The constrained optimal transport is feasible if* $C \geq H_b(p_{S_1})$*. Assume the Hamming distortion measure, under common randomness, we have*

$$D^{(B)}(R, C, p_X, p_Y) = \begin{cases} \dfrac{-2(1-p_X)p_X(H_b(m) - C)}{H_b(m) - H_b(p_{S_1})} + D^{(B)}_{\text{ind}}, \\ \qquad\qquad H_b(p_{s_1}) \leq C \leq \frac{R(H_b(p_{S_1}) - H_b(m))}{H_b(p_X)} + H_b(m) \\ \dfrac{-2(1-p_X)p_X R}{H_b(p_X)} + D^{(B)}_{\text{ind}}, \quad C > \dfrac{R(H_b(p_{S_1}) - H_b(m))}{H_b(p_X)} + H_b(m) \\ D^{(B)}_{\min}, \qquad C > H_b(p_S) \text{ and } R > H_b(p_X). \end{cases}$$

where $m = (1-p_X)(1-p_{S_1}) + p_X p_{S_1}$, $p_S = p_X + p_{S_1} - 2p_X p_{S_1}$ *and* $H_b(a) = -a \log a - (1-a)\log(1-a)$ *denotes the binary entropy function.*

Case Study: Bernoulli



$D^{(B)}(R, C, p_X, p_Y)$ versus R with $C = 0.8$, $p_X = 0.3$, $p_Y = 0.25$, $p_{S_1} = 0.2$.

Asymptotic Lossy Compression as Constrained Optimal Transport

Definition (Asymptotic Constrained Optimal Transport)

Consider i.i.d. random variables $X_i \sim p_X$, $Y_i \sim p_Y$, and $S_i \sim p_S$. The asymptotic constrained optimal transport problem with rate constraint R, classification loss C, and shared randomness U in the asymptotic regime ($n \to \infty$) is defined as

$$D^{(\infty)}(R, C, p_X, p_Y) = \inf_{p_{U,X^n,Z,Y^n} \in M(\otimes_{i=1}^n p_X, \otimes_{i=1}^n p_Y)} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[d(X_i, Y_i)]$$

$$\text{s.t.} \quad \frac{1}{n} H(Z|U) \leq R, \quad \frac{1}{n} \sum_{i=1}^n H(S_i|Y_i) \leq C.$$

Asymptotic Lossy Compression as Constrained Optimal Transport

Definition (Asymptotic Constrained Optimal Transport)

Consider i.i.d. random variables $X_i \sim p_X$, $Y_i \sim p_Y$, and $S_i \sim p_S$. The asymptotic constrained optimal transport problem with rate constraint R, classification loss C, and shared randomness U in the asymptotic regime ($n \to \infty$) is defined as

$$D^{(\infty)}(R, C, p_X, p_Y) = \inf_{p_{U,X^n,Z,Y^n} \in M(\otimes_{i=1}^n p_X, \otimes_{i=1}^n p_Y)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(X_i, Y_i)]$$

$$\text{s.t.} \quad \frac{1}{n} H(Z|U) \leq R, \quad \frac{1}{n} \sum_{i=1}^{n} H(S_i|Y_i) \leq C.$$

Theorem

In the asymptotic regime, the DRC function admits the single-letter characterization

$$D^{(\infty)}(R, C, p_X, p_Y) = \inf_{p_{X,Y} \in \Gamma(p_X, p_Y)} \mathbb{E}[d(X, Y)]$$

$$\text{s.t.} \quad I(X;Y) \leq R, \quad H(S|Y) \leq C.$$

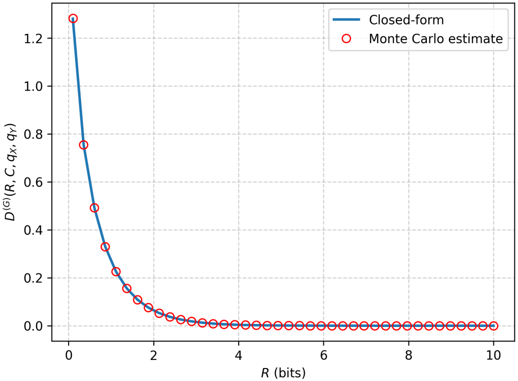
Case Study: Gaussian

Theorem

Consider $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ *and* $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $S \sim \mathcal{N}(\mu_S, \sigma_S^2)$, $\mathrm{Cov}(X, S) = \theta_1$. *The asymptotic constrained optimal transport is feasible if* $C \geq \frac{1}{2}\log\left(1 - \frac{\theta_1^2}{\sigma_S^2\sigma_X^2}\right) + h(S)$. *Under shared randomness, the asymptotic DRC tradeoff is*

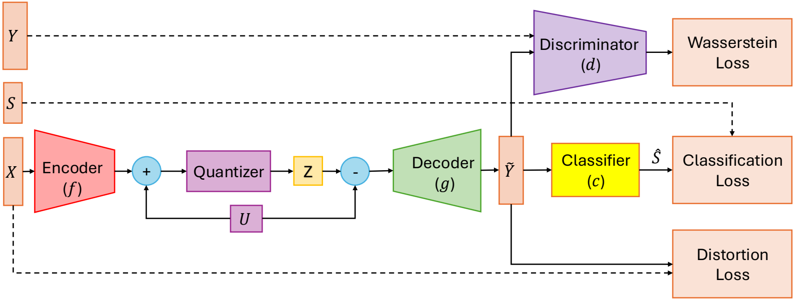
$$D^{(G)}(R, C, q_X, q_Y) = \begin{cases} (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - \frac{2\sigma_S\sigma_X^2\sigma_Y}{\theta_1}\sqrt{1 - e^{-2h(S)+2C}}, \\ \quad \frac{1}{2}\log\left(1 - \frac{\theta_1^2}{\sigma_S^2\sigma_X^2}\right) + h(S) \leq C \leq \frac{1}{2}\log\left(1 - \frac{\theta_1^2(1-2^{-2R})}{\sigma_S^2\sigma_X^2}\right) + h(S) \\ (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\sqrt{1 - 2^{-2R}}, \\ \quad C > \frac{1}{2}\log\left(1 - \frac{\theta_1^2(1-2^{-2R})}{\sigma_S^2\sigma_X^2}\right) + h(S) \\ 0, \quad C > h(S) \text{ and } R > h(X). \end{cases}$$

Case Study: Gaussian



$D^{(G)}(R, C, q_X, q_Y)$ versus R with $C = 2,\ X, Y, S \sim \mathcal{N}(0, 1),\ \theta_1 = 0.6.$

Experimental Results: Schematic



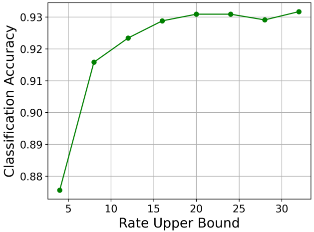
$$\min_{f,g,Q} \quad \mathbb{E}[\|X - g(Q(f(X,U)))\|_2^2]$$

$$\text{s.t.} \quad \phi(g(Q(f(X,U))), p_Y) \leq P, \quad H(Q(f(X,U))) \leq R, \quad H(S|g(Q(f(X,U)))) \leq C.$$

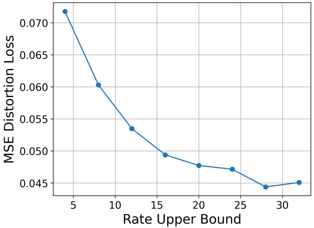
- $\tilde{Y} = g(Q(f(X,U)))$, WGAN discriminator aligns $p_{\tilde{Y}}$ with p_Y via a Wasserstein-1 penalty.
- U: common randomness, $\tilde{Y} = g(Q(f(X) + U) - U)$
- In practice, we optimize the relaxed loss

$$\mathcal{L} = \lambda_d \underbrace{\mathbb{E}\left[\|X - \tilde{Y}\|^2\right]}_{\text{Distortion loss}} + \lambda_p \underbrace{W_1(p_Y, p_{\tilde{Y}})}_{\text{Wasserstein loss}} + \lambda_c \underbrace{\mathrm{CE}(S, \hat{S})}_{\text{Cross-entropy loss}} . \tag{1}$$

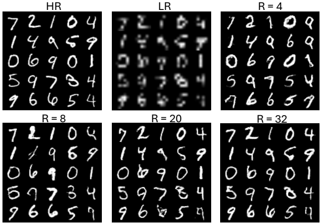
Experimental Results: Super-Resolution



(a) Accuracy vs. Rate



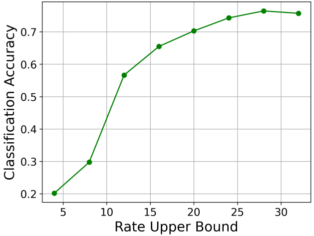
(b) MSE vs. Rate



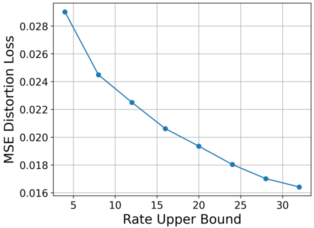
(c) Reconstruction

Application to Super-Resolution for MNIST Dataset

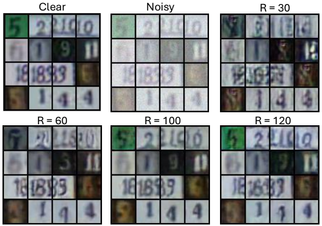
Experimental Results: De-noising



(a) Accuracy vs. Rate



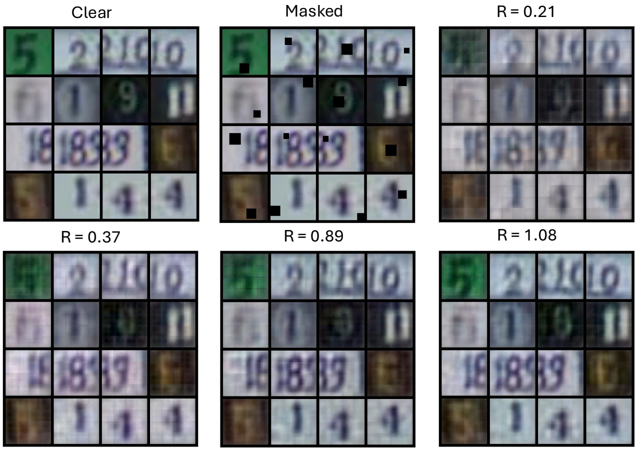
(b) MSE vs. Rate



(c) Reconstruction

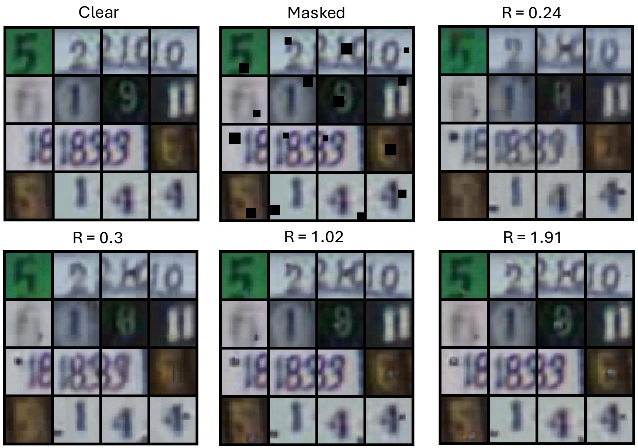
De-noising for SVHN Dataset

Experimental Results: Supervised In-painting



Supervised learning: have access to pairs of corrupted and clean images during training

Experimental Results: Unsupervised In-painting



Unsupervising learning: No access to pairs of corrupted and clean images during training

Thank you for listening!

Question?

Reference

[1] Huan Liu et al. "Cross-Domain Lossy Compression as Entropy Constrained Optimal Transport". In: *IEEE Journal on Selected Areas in Information Theory* 3.3 (2022), pp. 513–527.

[2] Nam Nguyen et al. "Universal Rate-Distortion-Classification Representations for Lossy Compression". In: *2025 IEEE Information Theory Workshop (ITW)*. 2025, pp. 1–6.

[3] Wei Wang et al. "Optimal Transport for Unsupervised Denoising Learning". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 45.2 (2023), pp. 2104–2118.

[4] Yuhan Wang et al. "Task-Oriented Lossy Compression With Data, Perception, and Classification Constraints". In: *IEEE Journal on Selected Areas in Communications* 43.7 (2025), pp. 2635–2650.