On The Classification-Distortion-Perception Tradeoff

Dong Liu, Haochen Zhang, Zhiwei Xiong

University of Science and Technology of China, Hefei, Anhui 230027, China

Email: dongeliu@ustc.edu.cn

National Engineering Laboratory for Brain-inspired Intelligence Technology and Application

Introduction

- **Motivation:** Different restoration tasks have various objectives.
- *Signal fidelity metrics* that evaluate how similar is the restored signal to the "original" signal. This metric is important for *image denoising* which wants to recover the noise-free image, and *compression artifact removal* which want to recover the uncompressed image.
- *Perceptual naturalness metrics* that evaluate how "natural" is the restored signal with respect to human perception. Some tasks may concern more about this metric, for example, *image super-resolution* is to produce image details to make the enhanced image look like having "high-resolution," *image inpainting* is to generate a complete image that looks "natural."
- *Semantic quality metrics* that evaluate how "useful" is the restored signal in the sense that it better serves for the following semantic-related analyses. For one example, an image containing a car license plate may have blur, and *image deblurring* can achieve a less blurred image so as to recognize the license plate; for another example, an image taken at night is difficult to identify, and *image contrast enhancement* can produce a more naturally looking image that is better understood.
- **Contribution**: This work considers these three groups of metrics jointly. When semantic quality is defined as the classification error rate achieved on the restored signal using a predefined classifier, we provide a rigorous proof of the existence of the classification-distortion-perception (CDP) tradeoff, i.e. the distortion, perceptual difference, and classification error rate cannot be made all minimal simultaneously.

Formulation

- Consider the process: $X \to Y \to \hat{X}$

X denotes the ideal "original" signal with the probability mass function $p_X(x)$, Y denotes the degraded signal, and $\hat{X}$ denotes the restored signal.
The degradation model and the restoration method can be denoted by conditional mass function $p(y|x)$ and $p(\hat{x}|y)$, respectively.
Thus, $p_Y(y) = \sum_{x\in\mathcal{X}} p(y|x)p_X(x)$ and $p_{\hat{X}}(\hat{x}) = \sum_{y\in\mathcal{Y}} p(\hat{x}|y)p_Y(y) = \sum_y\sum_x p(\hat{x}|y)p(y|x)p_X(x)$

Assume each sample of the original signal X belongs to one of two classes: w_1 or w_2. The a priori probabilities and the conditional mass functions are assumed to be known as P_1, $P_2 = 1 - P_1$ and $p_{X1}(x)$, $p_{X2}(x)$. There are:

$$p_{Yi}(y) = \sum_{x\in\mathcal{X}} p(y|x)p_{Xi}(x), i = 1,2 \qquad p_X(x) = P_1p_{X1}(x) + P_2p_{X2}(x)$$

$$p_{\hat{X}i}(\hat{x}) = \sum_{y\in\mathcal{Y}} p(\hat{x}|y)p_{Yi}(y) \qquad p_Y(y) = P_1p_{Y1}(y) + P_2p_{Y2}(y)$$

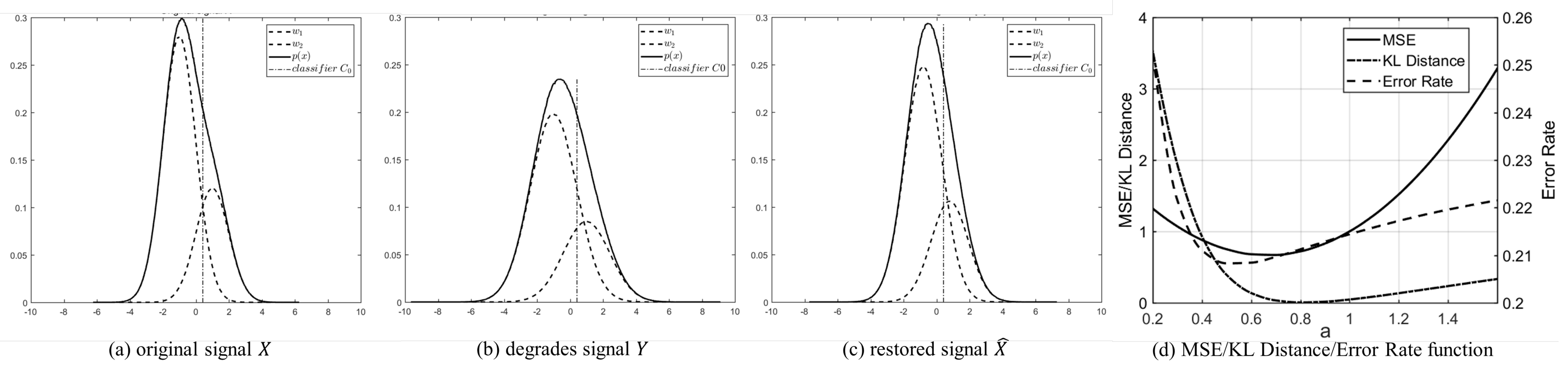
$$= \sum_y\sum_x p(\hat{x}|y)p(y|x)p_{Xi}(x), i = 1,2 \qquad p_{\hat{X}}(\hat{x}) = P_1p_{\hat{X}1}(\hat{x}) + P_2p_{\hat{X}2}(\hat{x})$$

Given a classifier $c(t) = c(t|\mathcal{R}) = \begin{cases}\omega_1, & \text{if } t\in\mathcal{R}\\ \omega_2, & \text{otherwise}\end{cases}$, there are

$$\text{Distortion} := \mathbb{E}[\Delta(X,\hat{X})]$$

$$\text{Perceptual Difference} := d(p_X, p_{\hat{X}})$$

$$\text{Classification Error Rate} := \varepsilon(\hat{X}|c) = \varepsilon(\hat{X}|\mathcal{R}) = P_2\sum_{\hat{x}\in\mathcal{R}} p_{\hat{X}2}(\hat{x}) + P_1\sum_{\hat{x}\notin\mathcal{R}} p_{\hat{X}1}(\hat{x})$$



(a) original signal X (b) degrades signal Y (c) restored signal $\hat{X}$ (d) MSE/KL Distance/Error Rate function

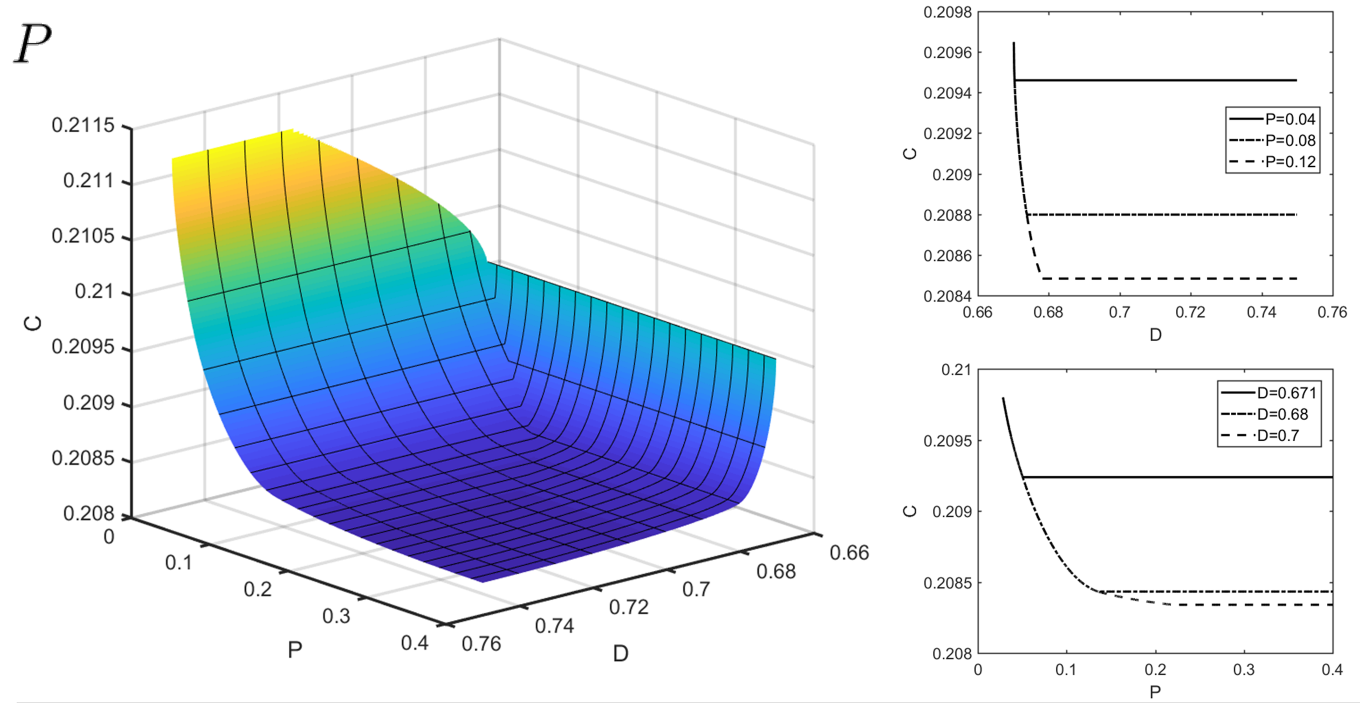
Figure This figure shows a simulation where X follows $P_1 = 0.7$, $P_2 = 0.3$, $p_{X1}(x) = \mathcal{N}(-1,1)$, $p_{X2}(x) = \mathcal{N}(1,1)$. This signal is corrupted by additive white Gaussian noise $Y = X + N$, where $N \sim \mathcal{N}(0,1)$. The denoising method is linear: $\hat{X} = aY$ where a is an adjustable parameter. C_0 is the optimal classifier for X.

- **Definition** The classification-distortion-perception (CDP) function is

$$C(D,P) = \min_{P_{\hat{X}|Y}} \varepsilon(\hat{X}|c_0), \text{subject to } \mathbb{E}[\Delta(X,\hat{X})] \le D, d(p_X,p_{\hat{X}}) \le P$$

where, $c_0 = c(\cdot|\mathcal{R}_0)$ is a predefined binary classifier.

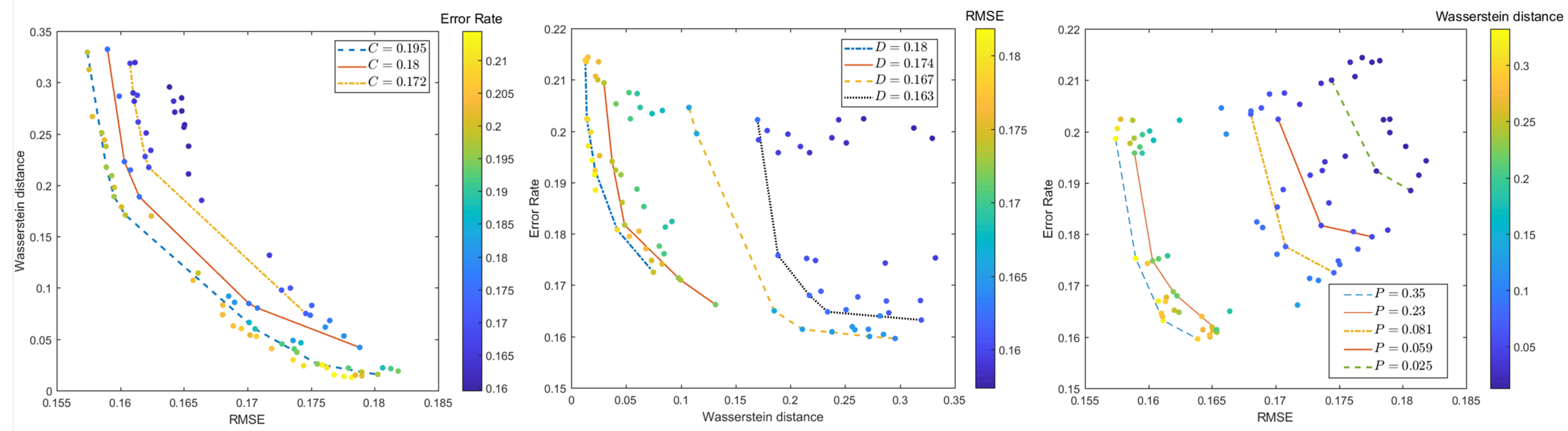
- **Theorem1** Considering the CDP function, if $d(\cdot,q)$ is convex in q, then $C(D,P)$ is:
  1. monotonically non-increasing
  2. convex in D and P.
- **Theorem2** Let the process of $X \to Y$ be denoted by $P_{Y|X}$, which is characterized by a conditional mass function $p(y|x)$, then $\epsilon_Y \ge \epsilon_X$. $\epsilon_Y = \epsilon_X$ if and only if $p(y|x)$ satisfies:
$\forall x_1 \in \mathcal{R}^+, \forall x_2 \in \mathcal{R}^-, \forall y, p(y|x_1)p(y|x_2) = 0,$
where $\mathcal{R}^+ = \{x|P_1p_{X1}(x) > P_2p_{X2}(x)\}$ and $\mathcal{R}^- = \{x|P_1p_{X1}(x) < P_2p_{X2}(x)\}$



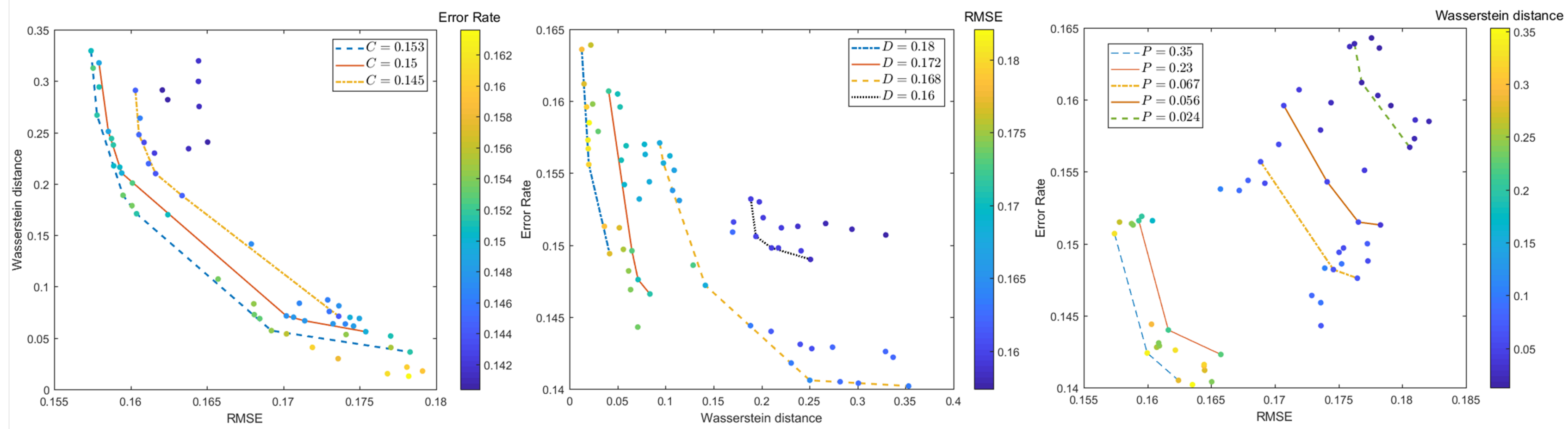
(a) (b)

Experiments

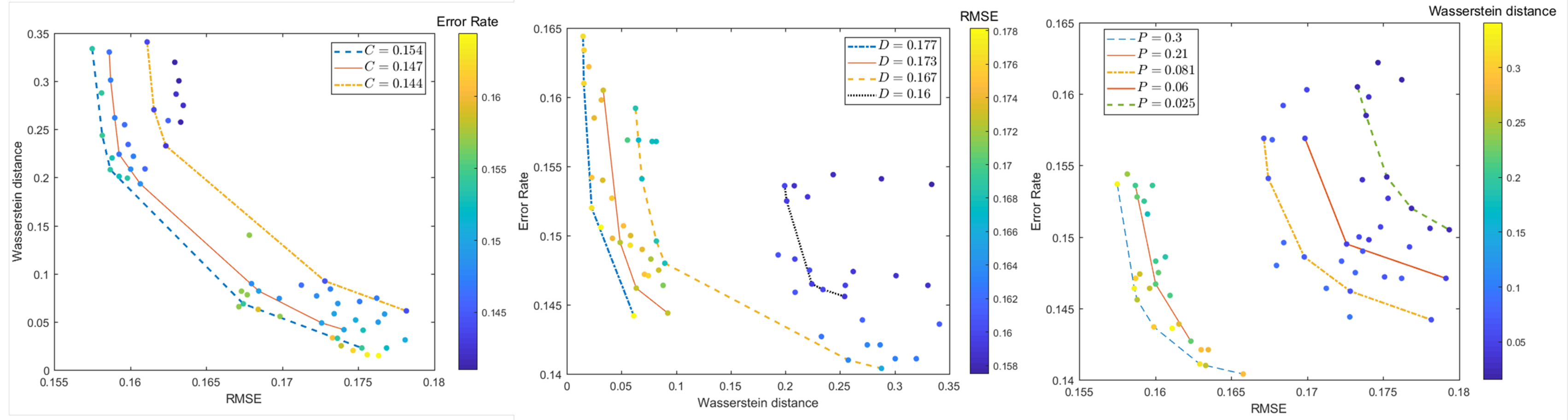
Profiles of the CDP Function



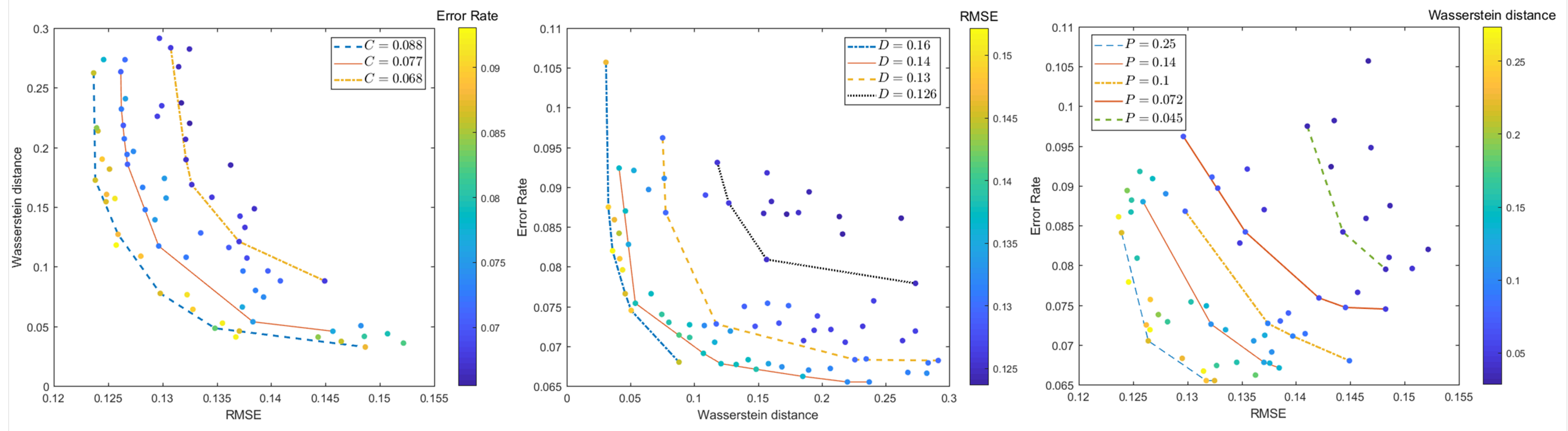
(a) MNIST-Denoising- Logistic Regression



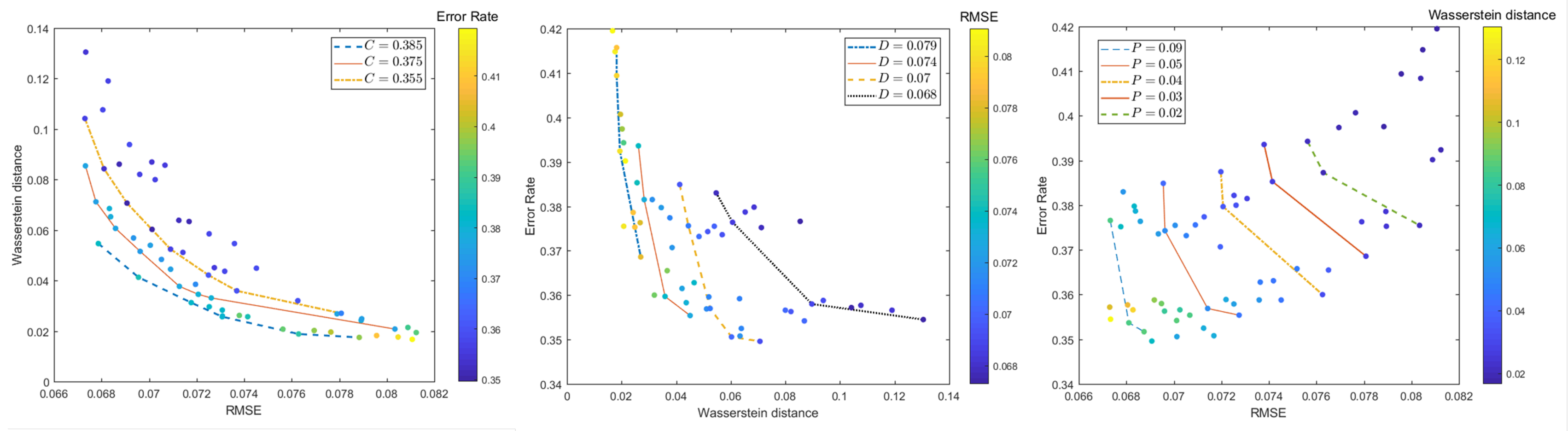
(b) MNIST-Denoising- CNN based classifier1



(c) MNIST-Denoising- CNN based classifier2



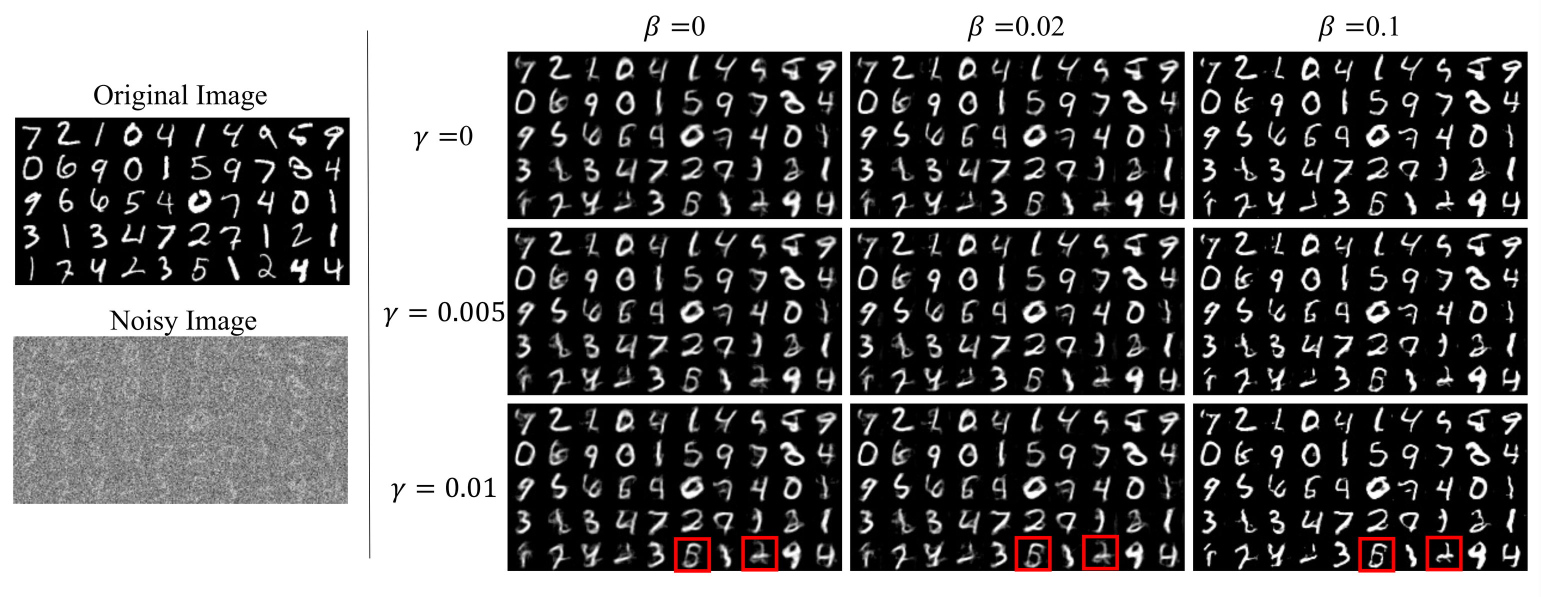
(d) MNIST-SR- CNN based classifier1



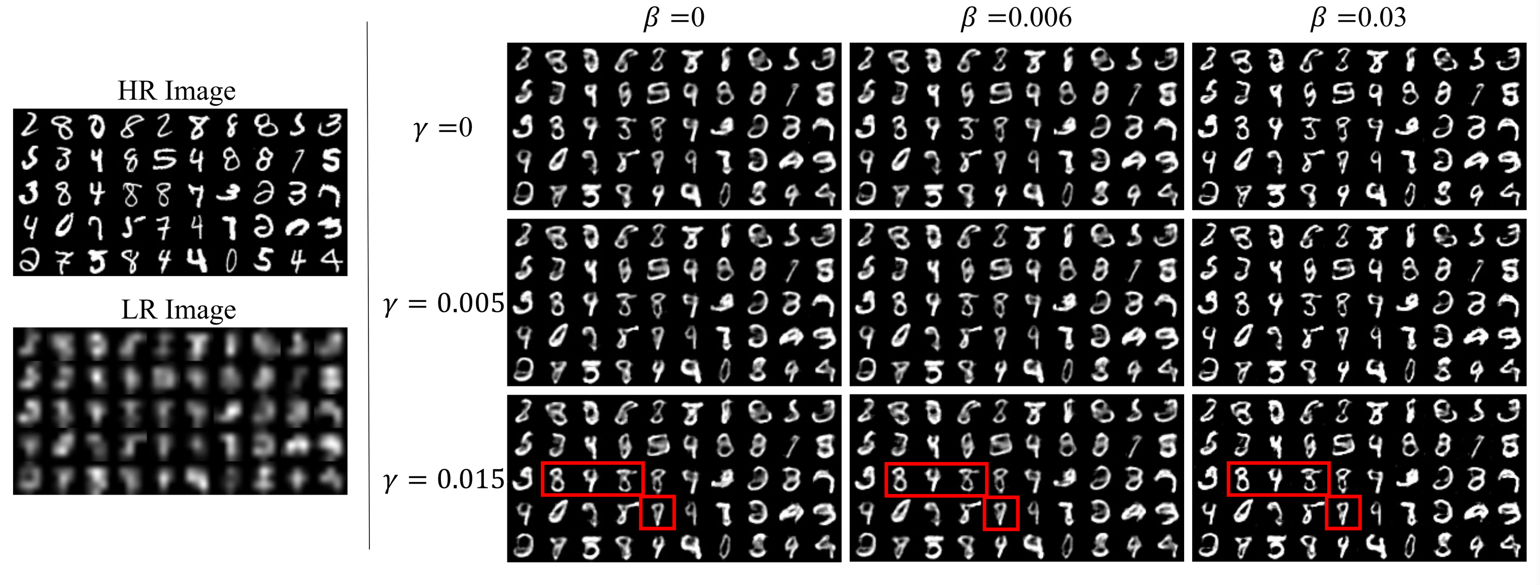
(e) CIFAR10-SR- CNN based classifier2

$$\ell_{restoration} = \alpha\ell_{MSE} + \beta\ell_{adv} + \gamma\ell_{CE}$$

Visual Results



(a) MNIST-Denoising- CNN based classifier1



(b) MNIST-SR- CNN based classifier1

Discussion

- **Profiles of CDP function**
- In the first column, when C is sufficiently large, there is a tradeoff between P and D. Once C is smaller, the P-D curve elevates, indicating that better classification performance comes at the cost of higher distortion and/or worse perceptual quality.
- Similarly in the other two columns, and the relations of C-P and C-D are convex as the theorem forecasts.
- Comparing (a), (b) and (c), although the error rates differ much in number, the trends of the CDP tradeoff are similar.
- **Visual result**
- The visual quality of restored images in general increases along with the weight β.
- Given the same β, when increasing γ, the visual quality decreases.
- There seems a positive correlation between classification and human recognition

Conclusion

- We have investigated the classification-distortion-perception tradeoff theoretically and experimentally.
- Regardless of the restoration algorithm, the classification error rate on the restored signal evaluated by a predefined classifier cannot be made minimal along with the distortion and perceptual difference.
- The CDP function is convex, indicating that when the error rate is already low, any improvement of classification performance comes at the cost of higher distortion and worse perceptual quality.