On The Classification-Distortion-Perception Tradeoff

Dong Liu, Haochen Zhang, Zhiwei Xiong

National Engineering Laboratory for Brain-Inspired
Intelligence Technology and Application

Background

Signal degradation is ubiquitous

Computational restoration of degraded signal has been investigated for many years.

Different restoration tasks have various objectives.

Distortion

Signal fidelity metrics that evaluate how similar is the restored signal to the "original" signal.

- Image denoising----------------------recover the noise-free image
- Compression artifact removal--------recover the uncompressed image

All the full-reference quality metrics, such as MSE, SSIM and VGG feature loss

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

Perception

Perceptual naturalness metrics that evaluate how "natural" is the restored signal with respect to human perception.

- Image super-resolution---------------produce image that looks like having high-resolution
- Compression artifact removal--------generate a complete image that looks natural

User study (real-vs.-fake, etc.) and no-reference quality assessment methods.

proportional to

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

Classification

Semantic quality metrics that evaluate how "useful" is the restored signal in the sense that it better serves for the following semantic-related analyses.

Some practical cases:

- Blurred car license plates-------------image deblurring
- Image taken at night ----------------image contrast enhancement

Only a few studies.

We can use a pre-trained classifier to measure this quality.

Contribution

Different restoration tasks have various objectives:

- Signal fidelity --------------------------- Distortion
- Perceptual naturalness ----------------- Perception
- Semantic quality ------------------------ Classification

The Classification-Distortion-Perception (CPD) Tradeoff.

2 | Formulation

$p_X(x_i) \equiv \frac{1}{N}, i = 1,2,\ldots,N$

Consider the process: $X \rightarrow Y \rightarrow \hat{X}$

- 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, there is

$$p_Y(y) = \sum_{x \in \mathcal{X}} p(y|x) p_X(x)$$

$$p_{\hat{X}}(\hat{x}) = \sum_{y \in \mathcal{Y}} p(\hat{x}|y) p_Y(y) = \sum_y \sum_x p(\hat{x}|y) p(y|x) p_X(x)$$

Assume each sample of the original signal X belongs to one of two classes: w_1 or w_2.

- The priori probabilities: P_1, $P_2 = 1 - P_1$
- The probability mass functions: $p_{X1}(x)$ and $p_{X2}(x)$

There are:

$$p_{Yi}(y) = \sum_{x \in \mathcal{X}} p(y|x) p_{Xi}(x), i = 1, 2$$

$$p_{\hat{X}i}(\hat{x}) = \sum_{y \in \mathcal{Y}} p(\hat{x}|y) p_{Yi}(y)$$

$$= \sum_{y} \sum_{x} p(\hat{x}|y) p(y|x) p_{Xi}(x), i = 1, 2$$

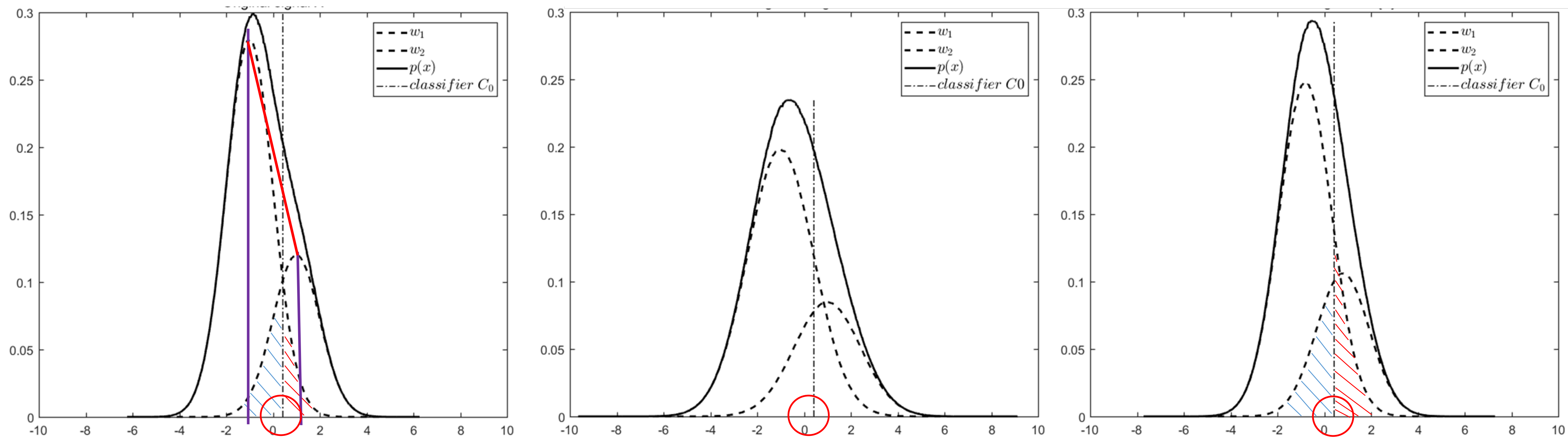
$$p_X(x) = P_1 p_{X1}(x) + P_2 p_{X2}(x)$$

$$p_Y(y) = P_1 p_{Y1}(y) + P_2 p_{Y2}(y)$$

$$p_{\hat{X}}(\hat{x}) = P_1 p_{\hat{X}1}(\hat{x}) + P_2 p_{\hat{X}2}(\hat{x})$$

2 Formulation

$$p_X(x) = P_1 p_{X1}(x) + P_2 p_{X2}(x)$$



(a) original signal X (b) degrades signal Y (c) restored signal $\widehat{X}$

- 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.

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

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

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

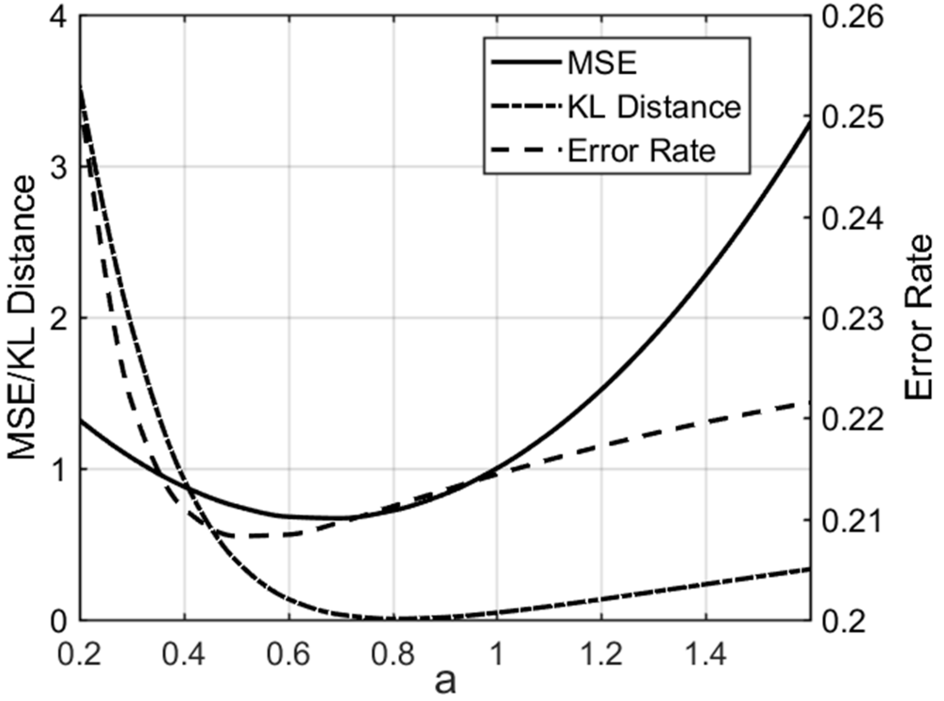
$$\text{Classification Error Rate} := \varepsilon(\hat{X}|c) = \varepsilon(\hat{X}|\mathcal{R}) = P_2 \sum_{\hat{x}\in\mathcal{R}} p_{\hat{X}2}(\hat{x}) + P_1 \sum_{\hat{x}\notin\mathcal{R}} p_{\hat{X}1}(\hat{x})$$

For example,

$$\mathbf{MSE}(a) = \mathbb{E}[(X - \hat{X})^2] = \mathbb{E}(X^2) + \mathbb{E}(\hat{X}^2) - 2\mathbb{E}(X\hat{X})$$

$$Y = X + N$$
$$\hat{X} = aY$$

$$= (2 + \sigma_N^2)a^2 - 4a + 2$$



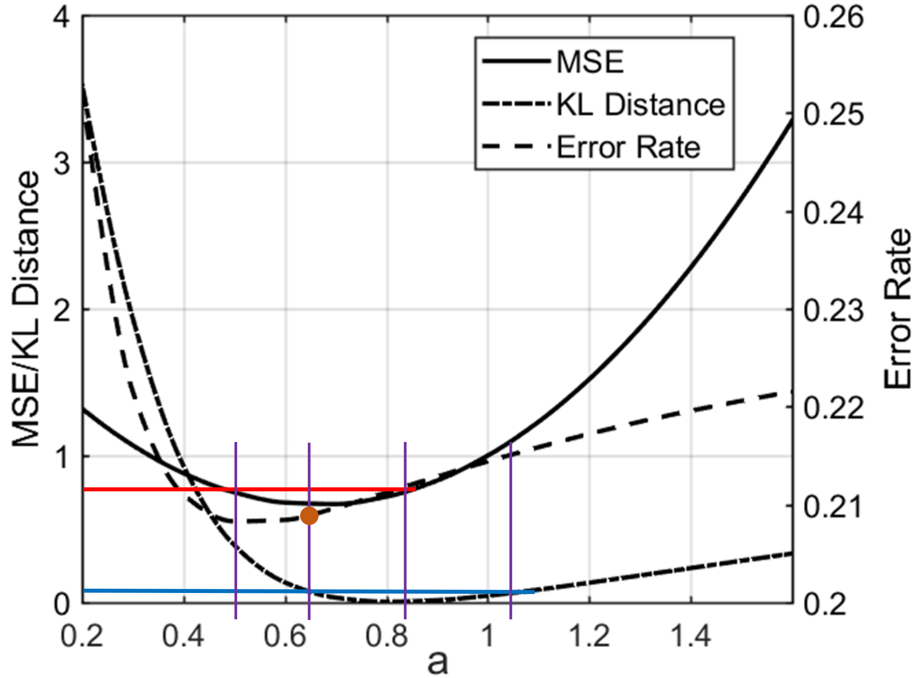
(d) MSE/KL Distance/Error Rate function

2 Formulation

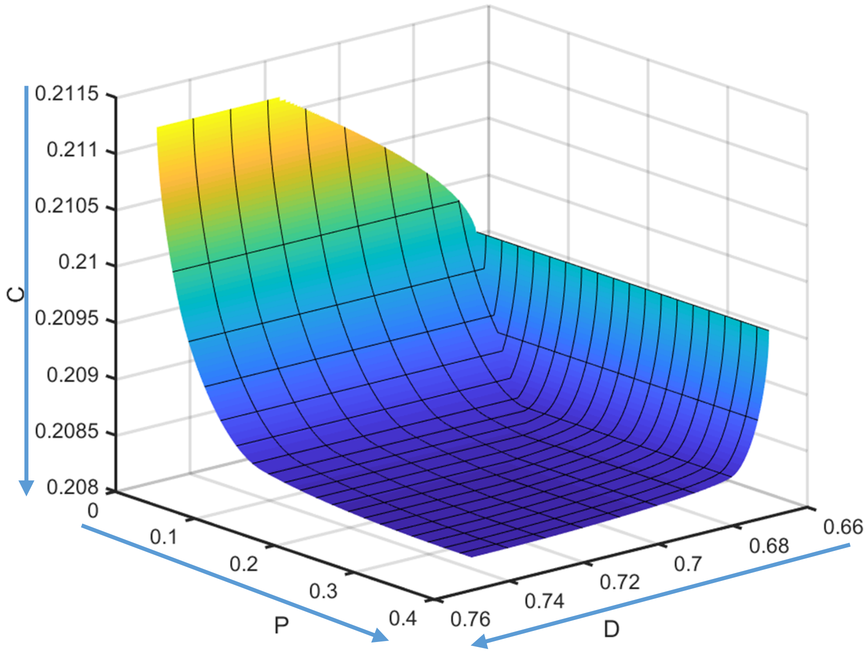
➢ **Definition** The classification-distortion-perception (CDP) function is

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

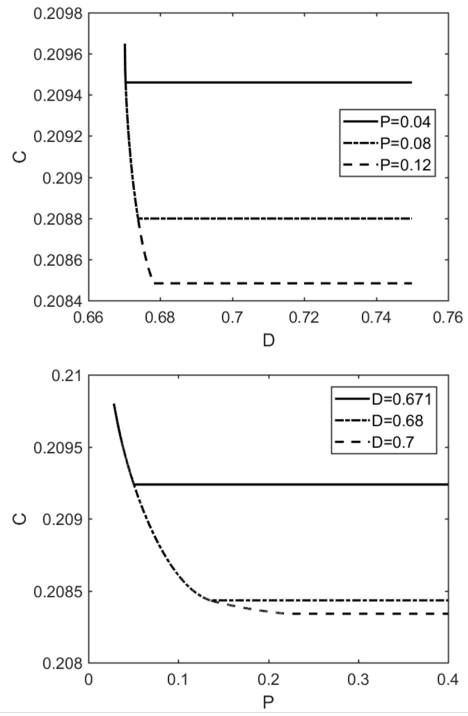
where, $c_0 = c(\cdot|\mathcal{R}_0)$ is a predefined binary classifier.



(d) MSE/KL Distance/Error Rate function



(a)



(b)

- **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.

- **Discussion:**

- The tradeoff indicates that distortion, perceptual difference, and classification error rate cannot be made all minimal simultaneously.

- The convexity of $C(D, P)$ implies the tradeoff is stronger at the low distortion or low perception ranges. In these ranges, any small improvement in distortion/perception achieved by a restoration algorithm, must be accompanied by a large loss of classification accuracy.

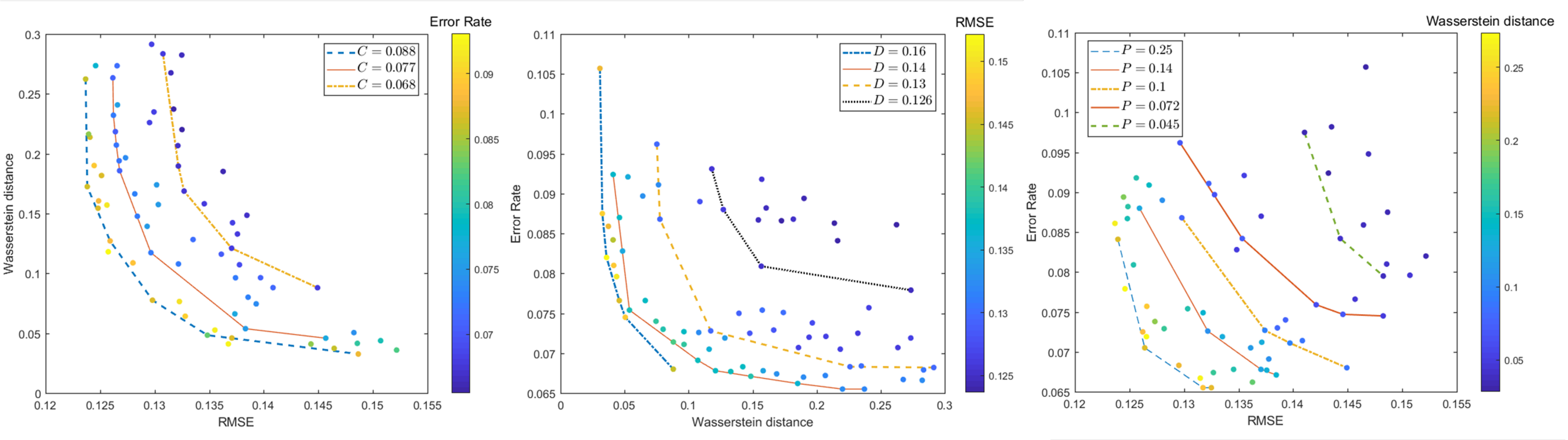
➢ We did five group of experiments using different configurations.

	Dataset	Task	Classifier
Exp-1	MNIST	Denoising	Logistic
Exp-2	MNIST	Denoising	CNN-1
Exp-3	MNIST	Denoising	CNN-2
Exp-4	MNIST	SR	CNN-1
Exp-5	CIFAR-10	SR	CNN-2'

➢ We train the restoration network with the following loss:

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

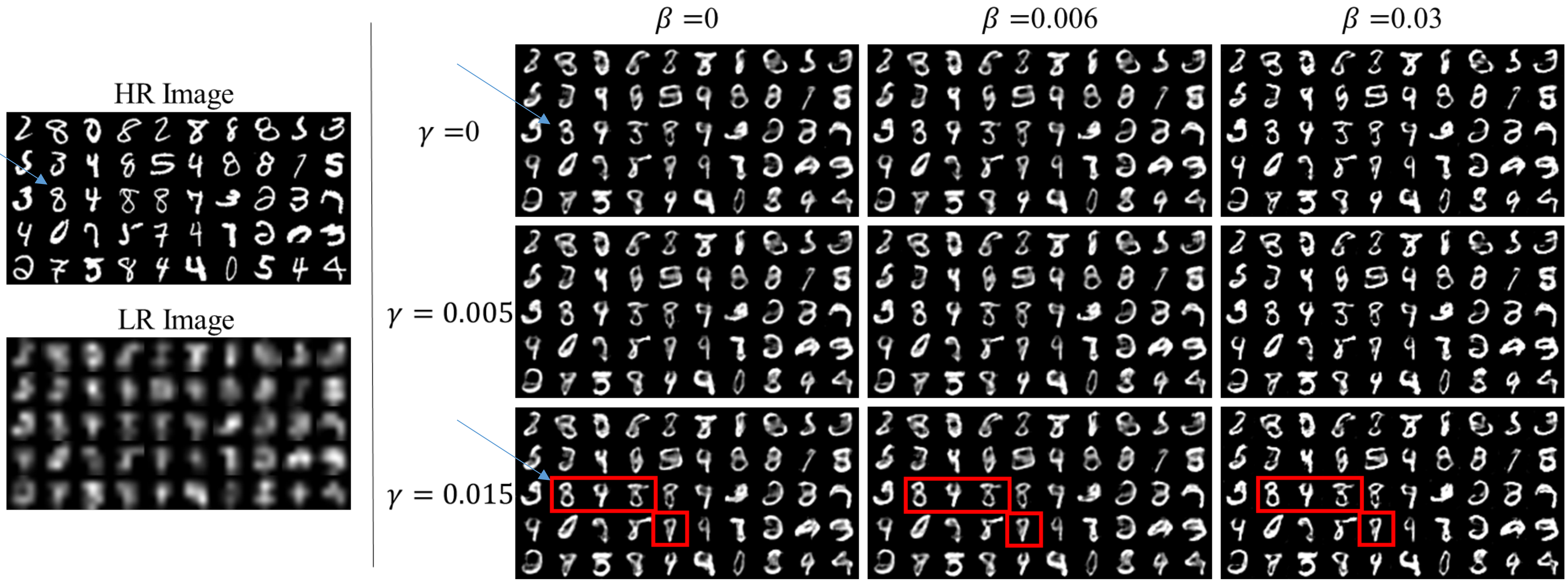
- The first term is MSE loss to represent distortion.
- The second term is an adversarial loss, minimizing which is to ensure perceptual quality.
- The third term is cross entropy, corresponding to classification error rate.



MNIST-SR- CNN based classifier1

- 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.
- We can observe the relations of C-P and C-D and all of them are convex as the theorem forecasts.

3 | Experiments



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

Conclusion

- ✓ 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.

Thank you for your listening

NEL-BITA
类脑智能技术及应用国家工程实验室
National Engineering Laboratory for Brain-inspired Intelligence Technology and Application

How to measure perception in experiment?

Here we adopt the Wasserstein GAN and the adversarial loss ℓ_{adv} is proportional to the Wasserstein distance $d_W(p_X, p_{\hat{X}})$

Note that in the Wasserstein GAN, the discriminator loss is an estimate of the Wasserstein distance.

Thus it can be used to assess the perceptual quality of the restored images quantitatively.

What about retain the classifier?

Theorem2 Let the process of $X \rightarrow Y$ be denoted by $P_{Y|X}$, which is characterized by a conditional mass function $p(y|x)$, then there is $\epsilon_Y \geq \epsilon_X$.

$\epsilon_Y = \epsilon_X$ 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 $\mathcal{R}^+ = \{x|P_1 p_{X1}(x) > P_2 p_{X2}(x)\}$ and $\mathcal{R}^- = \{x|P_1 p_{X1}(x) < P_2 p_{X2}(x)\}$.