Optimally Controllable Perceptual Lossy Compression

Zeyu Yan, Fei Wen, Peilin Liu

Department of Electronic Engineering /
Brain-inspired Application Technology Center,
Shanghai Jiao Tong University, China

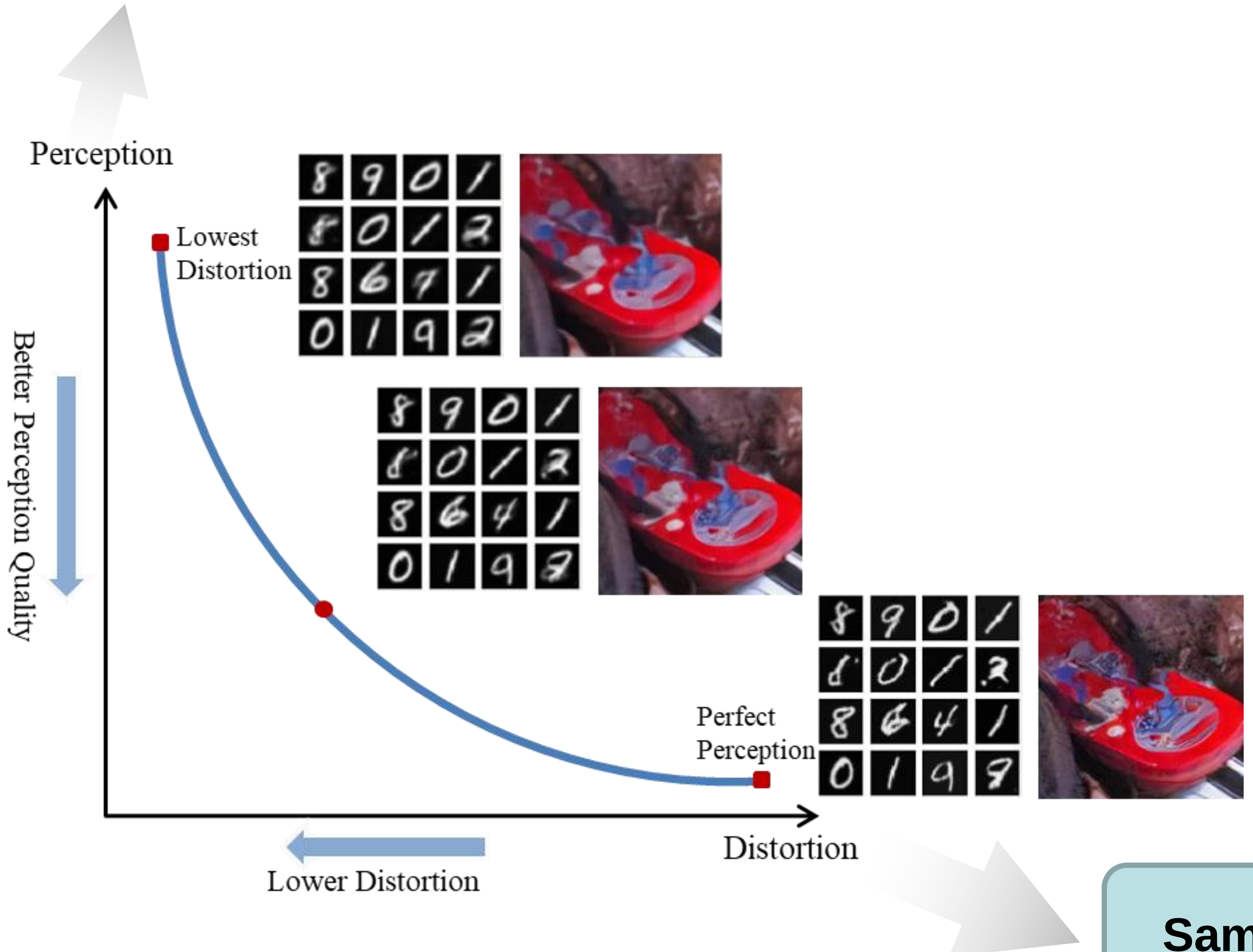
D-P tradeoff in lossy compression

lower distortion
blurred details

higher distortion
clear details

D-P tradeoff in lossy compression

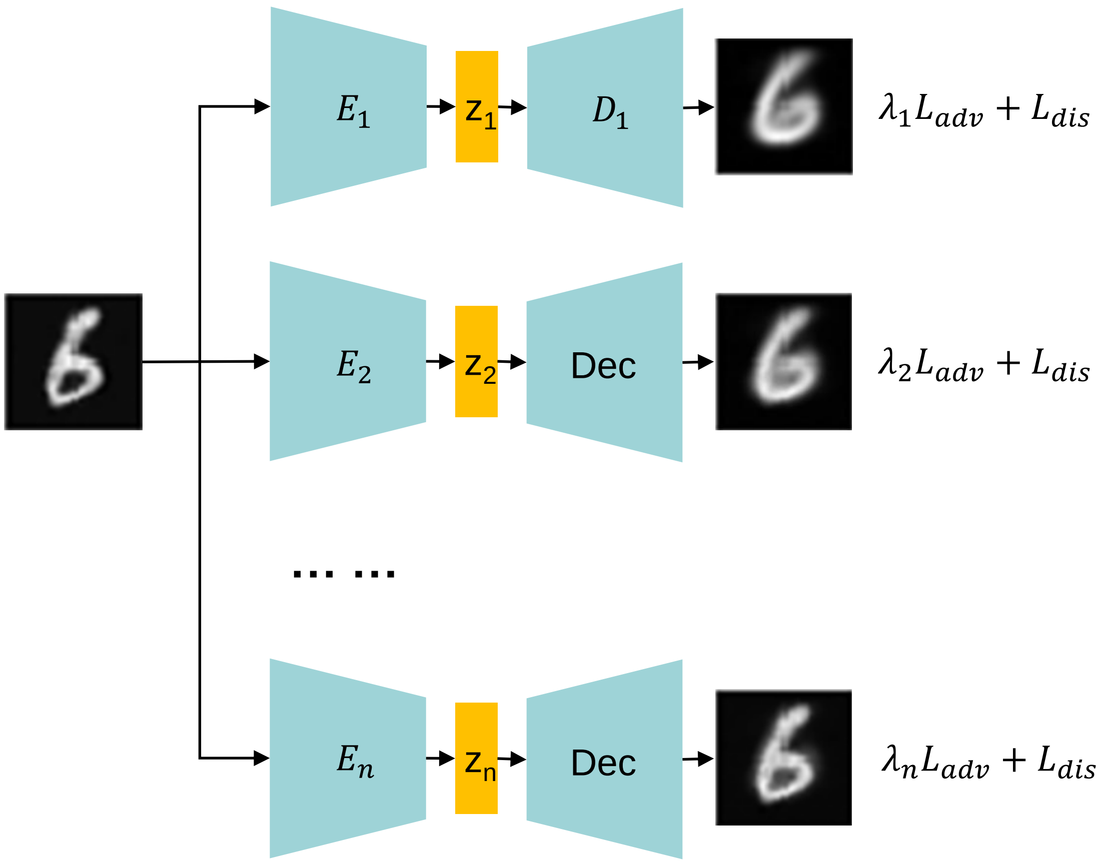
Distribution divergence



Sample distance

How to achieve optimal distortion-perception tradeoff?

Distortion-plus-adversarial loss (DAL) $L = \lambda L_{adv} + L_{dis}$



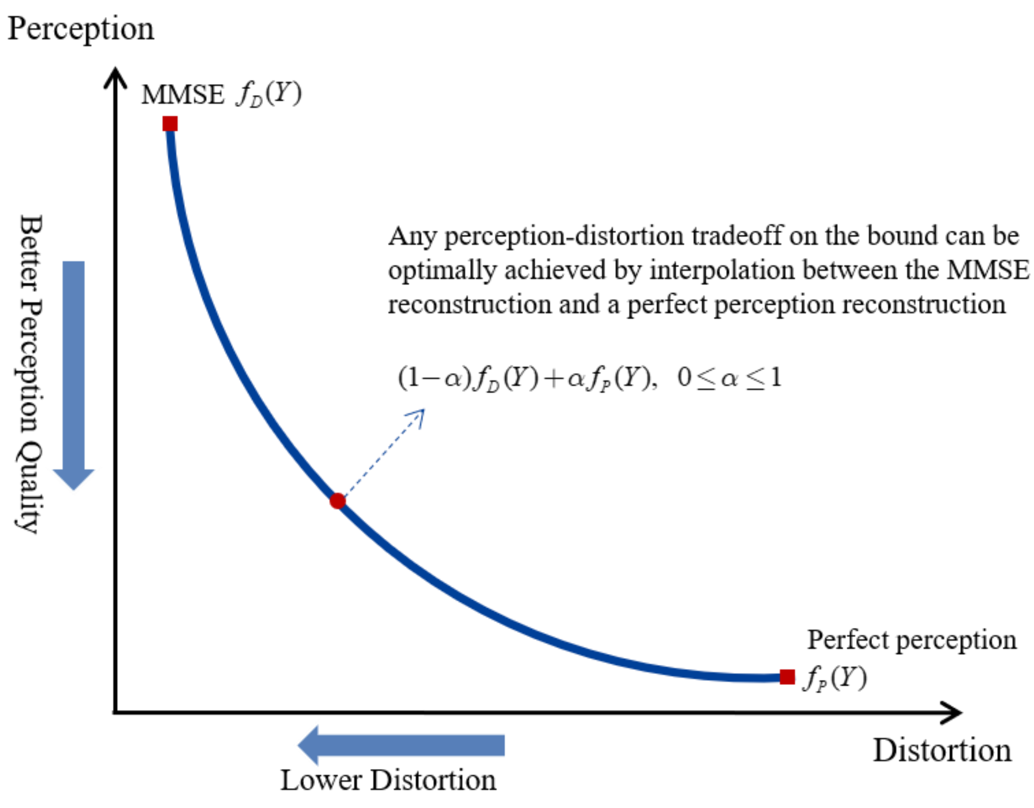
Hard to quantitatively control D-P tradeoff

Infinite number of encoder-decoder pairs are needed to fit D-P tradeoff

Main Contributions

Contribution 1 (a nontrivial theoretical finding):

1) one encoder and two decoders are enough for optimally achieving arbitrary D-P tradeoff in certain condition,
2) the perceptual quality (Wasserstein-2 distance) and distortion (MSE) can be quantitatively controlled by interpolating outputs of two decoders,



Main Contributions

Distortion-perception function can be expressed as

$$D(P) := \min_{E \in \Omega, G} \mathbb{E}\|X - G(E(X))\|^2$$
$$\text{s.t.} \quad W_2^2(p_X, p_{G(E(X))}) \leq P,$$

Ω: the set of encoders with a given bit-rate R

W_2^2: squared Wasserstein-2 distance

Theorem 1. *Let (E_d, G_d) be an optimal encoder-decoder pair to $D(+\infty)$, and G_p be an optimal decoder to $D(0)$ for a fixed encoder E_d. Denote $Z_d := E_d(X)$ and $P_d := W_2^2(p_X, p_{G_d(Z_d)})$. Then, these hold:*

i) *E_d is an optimal encoder for any $P > 0$.*

ii) *Let $\alpha = \min\left(\sqrt{\frac{P}{P_d}}, 1\right) \in [0,1]$, define*

$$G_\alpha^*(Z_d) := \alpha G_d(Z_d) + (1 - \alpha) G_p(Z_d)$$

then (E_d, G_α^*) is an optimal encoder-decoder pair to $D(P)$.

Main Contributions

Contribution 2 (perfect perception decoding):

- We propose a training method for perfect perceptual quality decoder G_p.

An augmented training loss without compromising the optimality

$$\min_{p_{\hat{X},Z_d}} W_1(p_{\hat{X},Z_d}, p_{X,Z_d}) \quad \Rightarrow \quad \min_{p_{\hat{X},X_d}} W_1(p_{\hat{X},X_d}, p_{X,X_d}) + \lambda \mathbb{E}\|\hat{X} - X_d\|$$

Theorem 2. *Let (E_d, G_d) be an MMSE encoder-decoder pair, and $W_1(\cdot,\cdot)$ be the Wasserstein-1 distance. Denote $Z_d = E_d(X)$ and $X_d = G_d(Z_d)$, then these hold:*

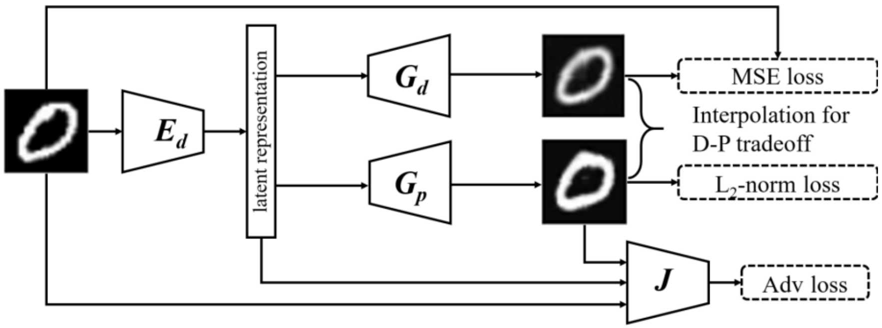
i) When $0 \le \lambda < 1$, the optimal solution satisfies $p_{\hat{X},X_d} = p_{X,X_d}$, or equivalently $p_{\hat{X},Z_d} = p_{X,Z_d}$.

ii) When $\lambda > 1$, the optimal solution satisfies $\hat{X} = X_d$.

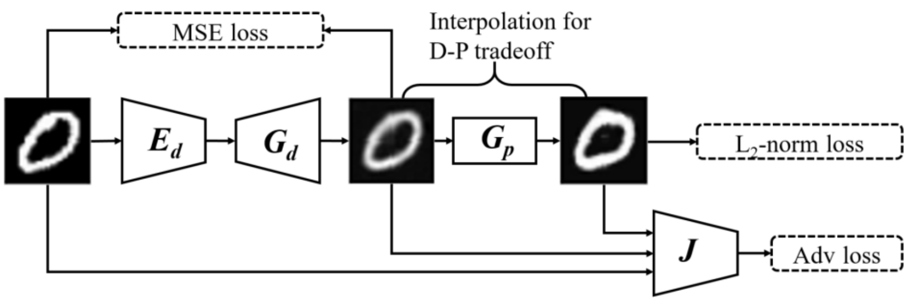
Main Contributions

Contribution 3 (perfect perception decoding):

- We propose two optimal training frameworks for perfect perceptual decoding, which enables the realization of interpolation based optimal D-P tradeoff.



(a) Framework A



(b) Framework B

Experiments

- Results on MNIST

$$\min_{p_{\hat{X},X_d}} W_1(p_{\hat{X},X_d}, p_{X,X_d}) + \lambda \mathbb{E}\|\hat{X} - X_d\|$$

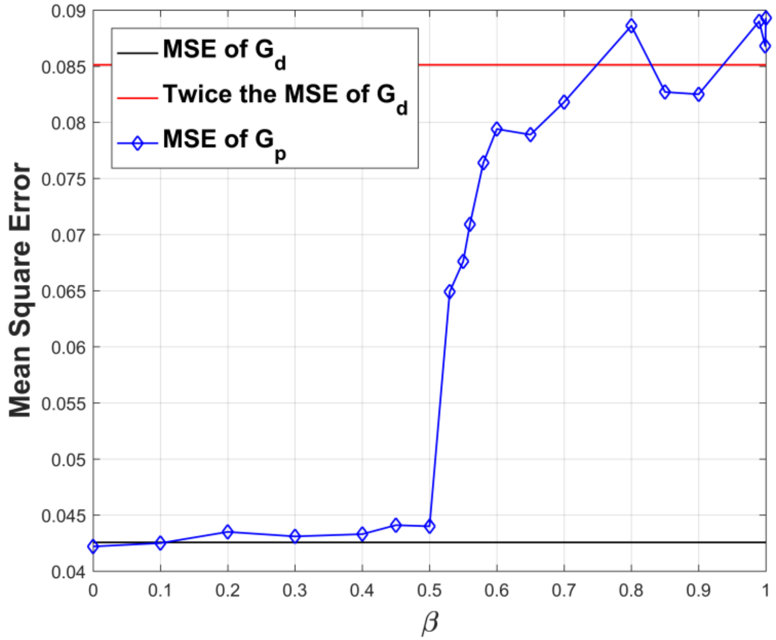
To verify our result in **contribution 2**, we train framework A with loss

$$\max_{\|J\|_L \le 1} \mathbb{E}[J(G_p(E(X)), E(X))] - \mathbb{E}[J(X, E(X))]$$

$$\min_{G_p} (1-\beta)\mathbb{E}\|G_p(X) - X_d\| - \beta \mathbb{E}[J(G_p(X), E(X))]$$

where $\lambda = \frac{1-\beta}{\beta}$

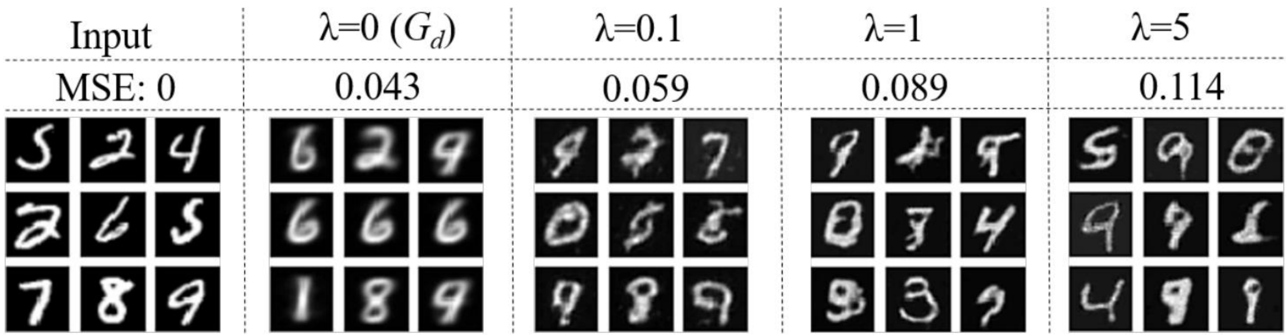
MSE jumps from $D(+\infty)$ (MSE of G_d) to $2D(+\infty)$ at $\beta = 0.5$ ($\lambda = 1$)



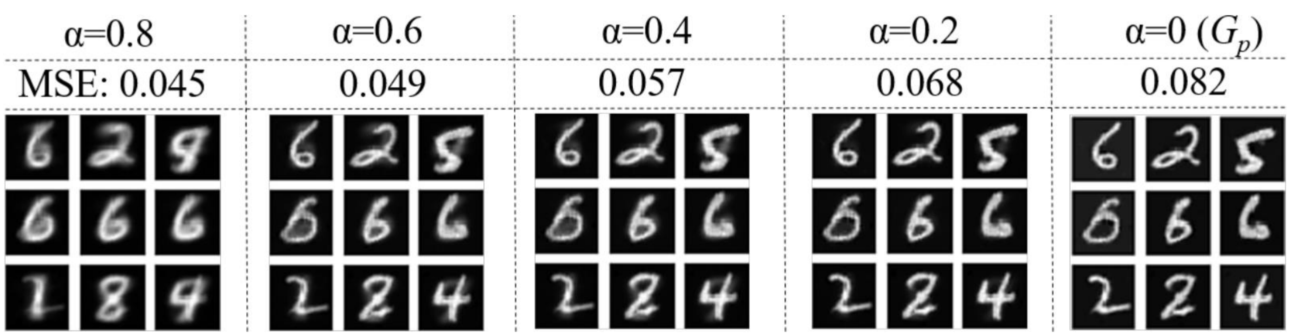
Experiments

- Results on MNIST

Samples decoded by conventional framework $L = \lambda L_{adv} + L_{dis}$

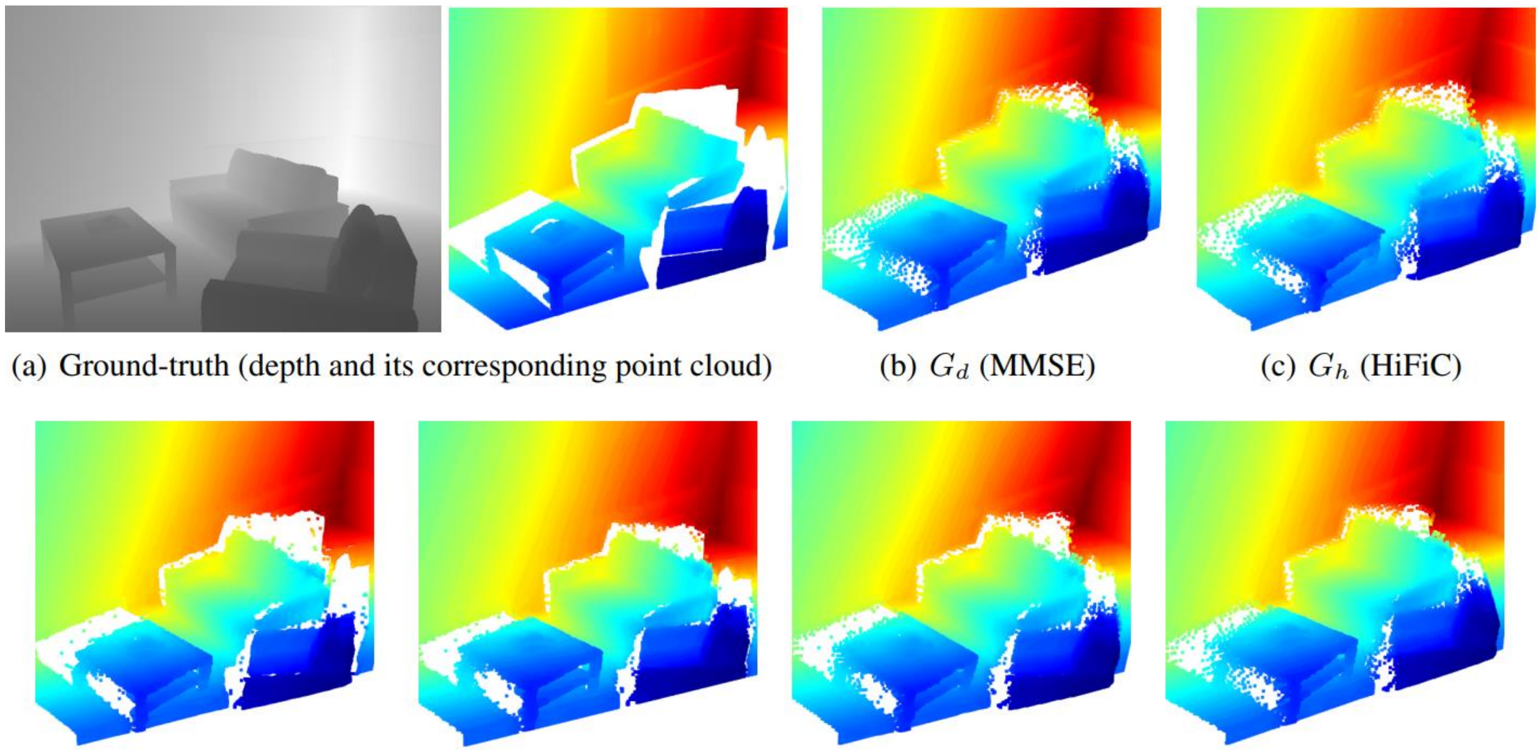


Samples decoded by our framework



Experiments

- Results on SUNCG dataset



(a) Ground-truth (depth and its corresponding point cloud)

(b) G_d (MMSE)

(c) G_h (HiFiC)

(d) G_p (α=0)

(e) α=0.25

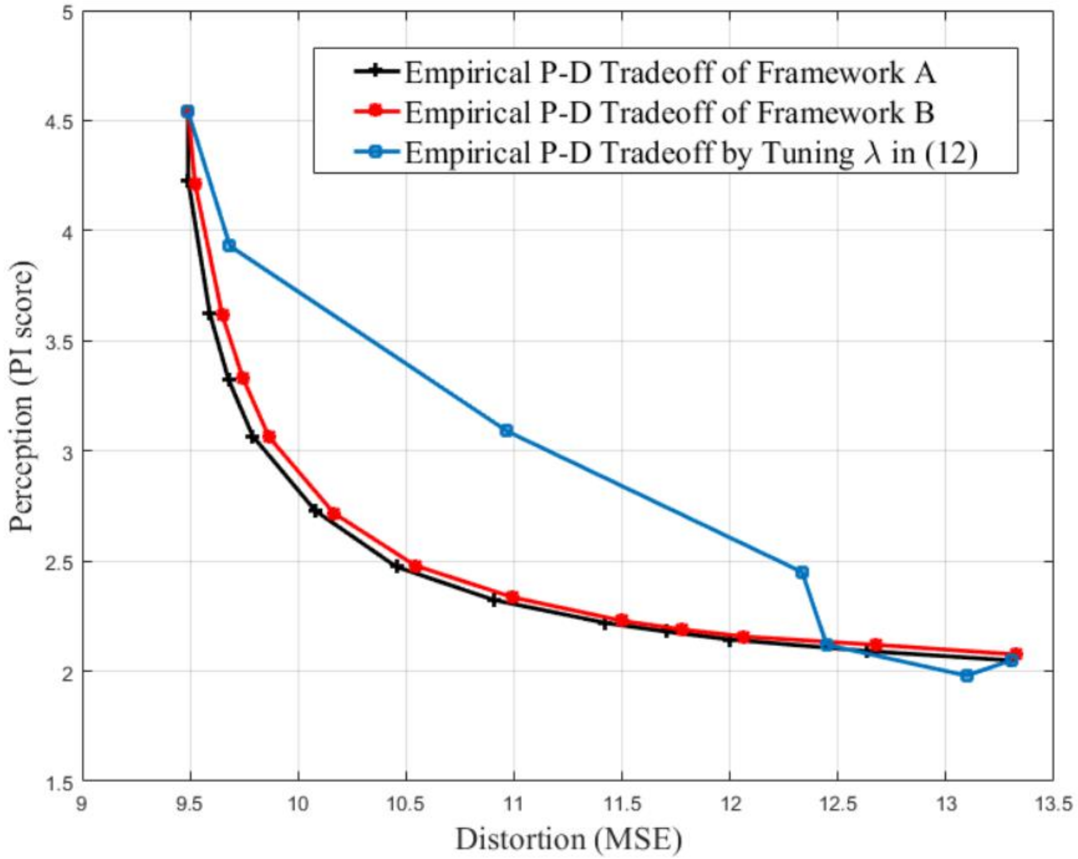
(f) α=0.5

(g) α=0.75

Experiments

- Results on KODAK dataset

Distortion (MSE) vs. perception (PI score)



Experiments

- Results on KODAK dataset

(a) Ground-truth (b) G_d (MMSE) (c) HiFiC

(d) Framework A ($\alpha = 0$ and $\alpha = 0.5$) (e) Framework B ($\alpha = 0$ and $\alpha = 0.5$)

Thank you !