NFLX-LIVE Research Project

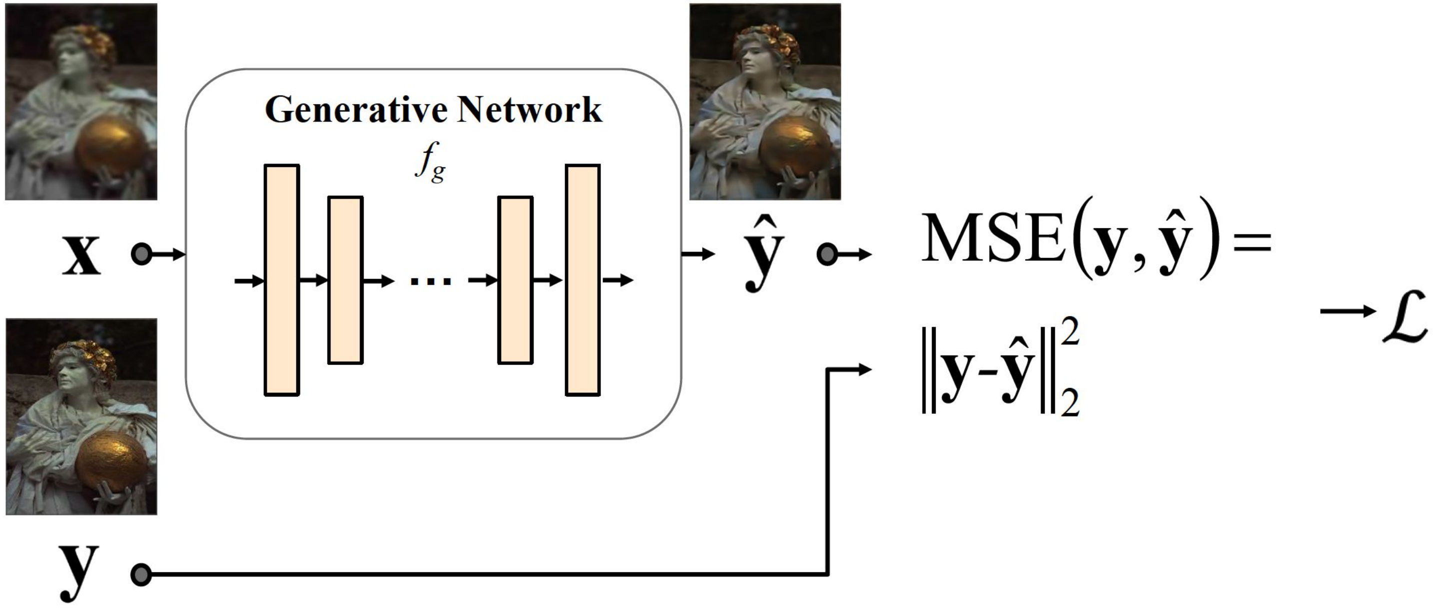
Perceptual Optimization using Deep Compression Model

Li-Heng Chen, Christos G. Bampis, Zhi Li, Andrey Norkin and Alan C. Bovik

The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

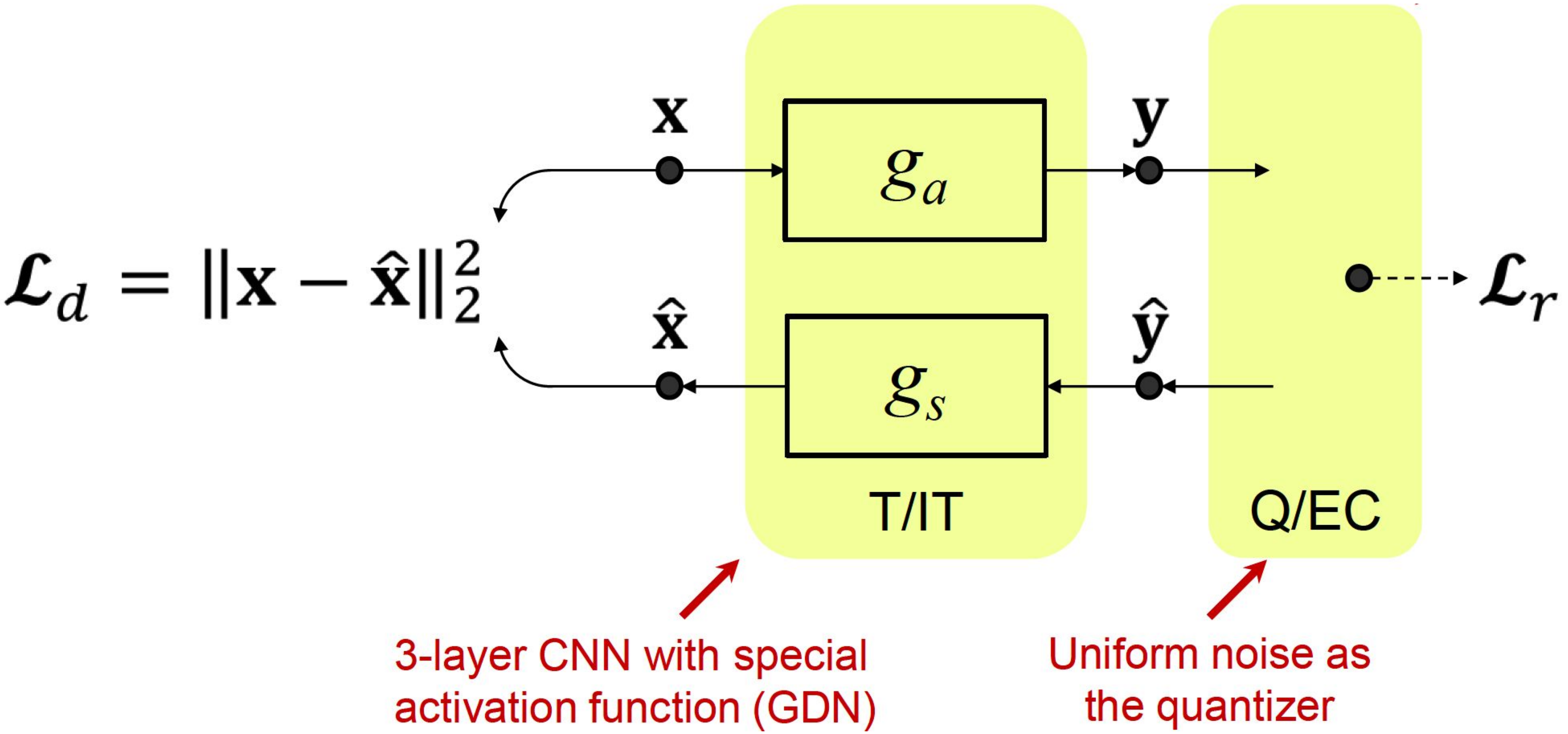
goal: deploy *any* perceptual quality model to optimize “deep” image and video compression networks.

Quick background on generative modeling



Applications: super-resolution, denoising, frame-rate conversion, compression (recently!)

Deep image compression (Balle *et al.*)



Intuitions behind deep image compression modeling

"pixel/distortion loss"

$$\mathcal{L}_d = ||\mathbf{x} - \hat{\mathbf{x}}||_2^2$$

"rate loss"

$$\mathcal{L}_r$$

total loss: $\mathcal{L}_{total} = \lambda \mathcal{L}_d + \mathcal{L}_r$

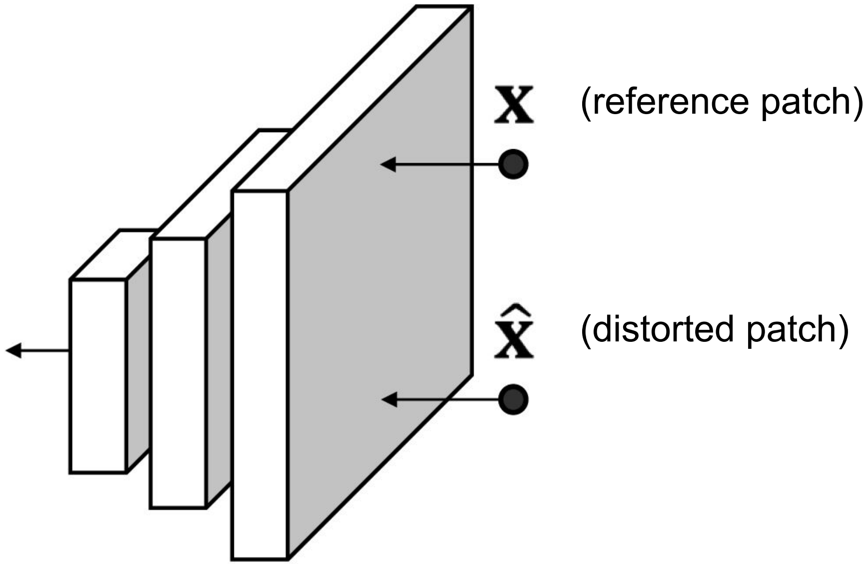
- ➔ tradeoff between rate and distortion (RDO)
- ➔ what about other pixel losses?

Perceptual optimization of deep image encoders

- ➔ so far, mostly MSE, SSIM and MS-SSIM optimization
- ➔ analytically tractable, differentiable, etc.
- ➔ cannot use more complicated models, like VMAF
- ➔ given the specifics of each metric, can we generalize the approach to any desired metric?

Simple idea: use a pre-trained network

$\text{VMAF}_{\text{proxy}}(\mathbf{x}, \hat{\mathbf{x}})$

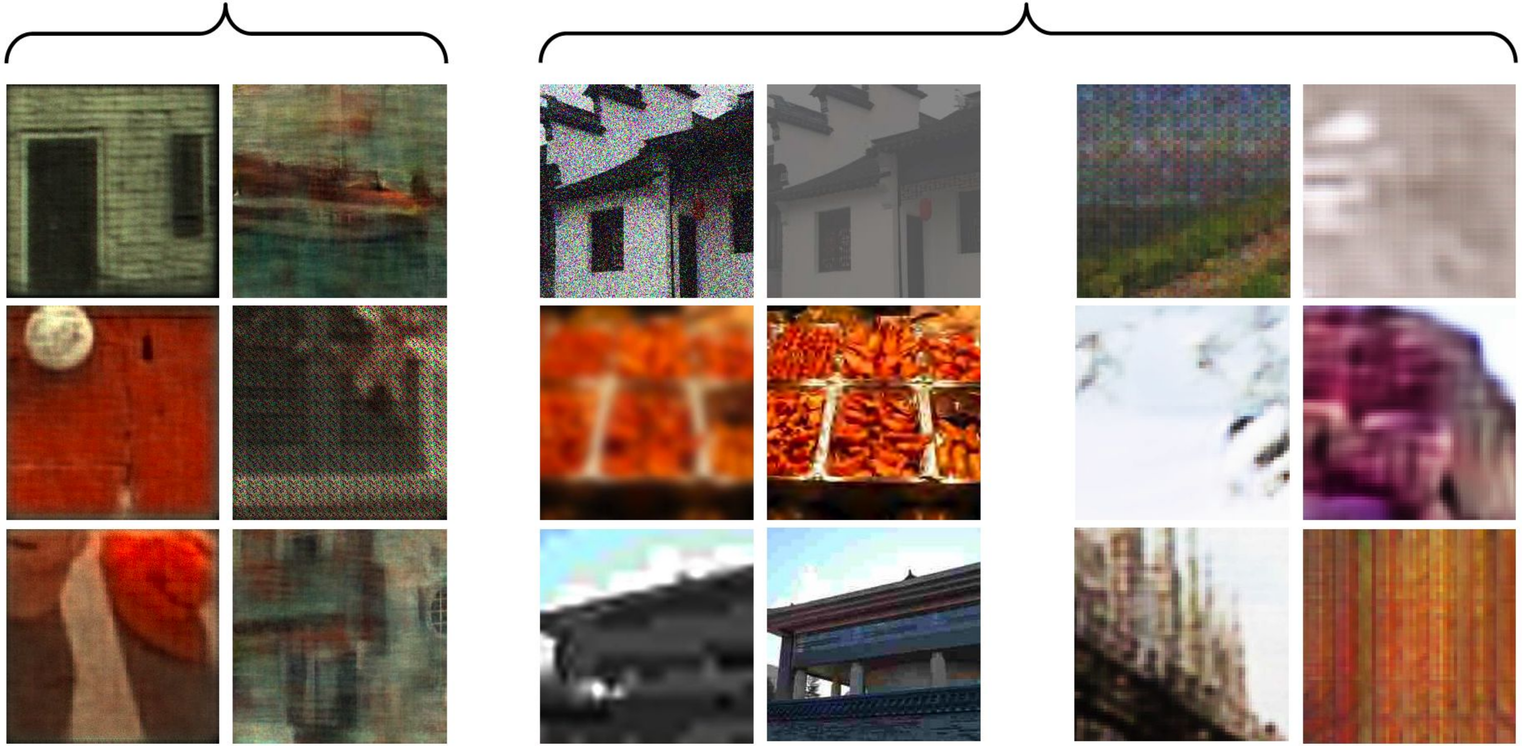


- ➔ train network so that proxy VMAF score matches VMAF
- ➔ 3 or 4 layers are enough

Conceptual problem with a pre-trained network

What we're dealing with

What we have for training (existing datasets)



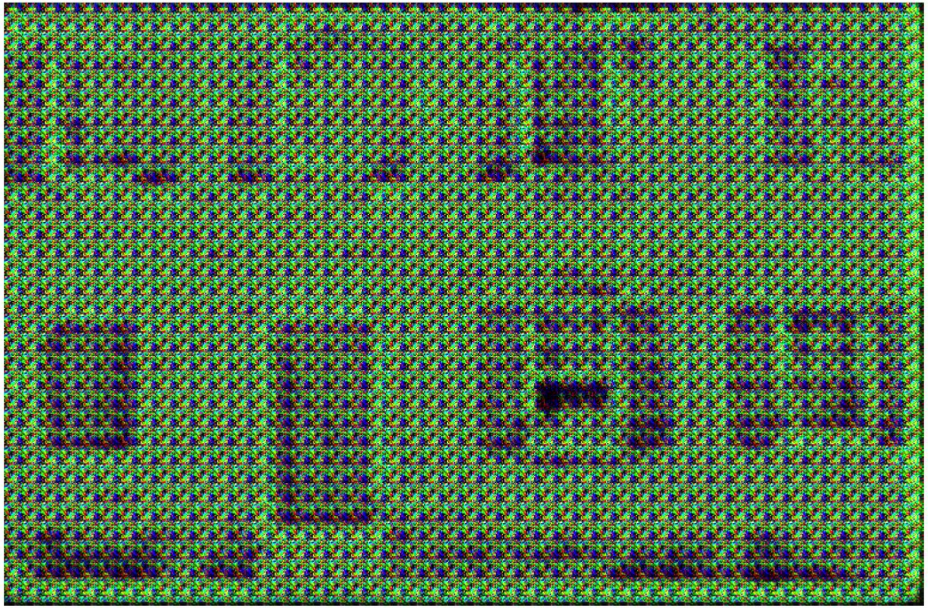
(a) Reconstructed Patches

(b) Waterloo Dataset

(c) BAPPS Dataset

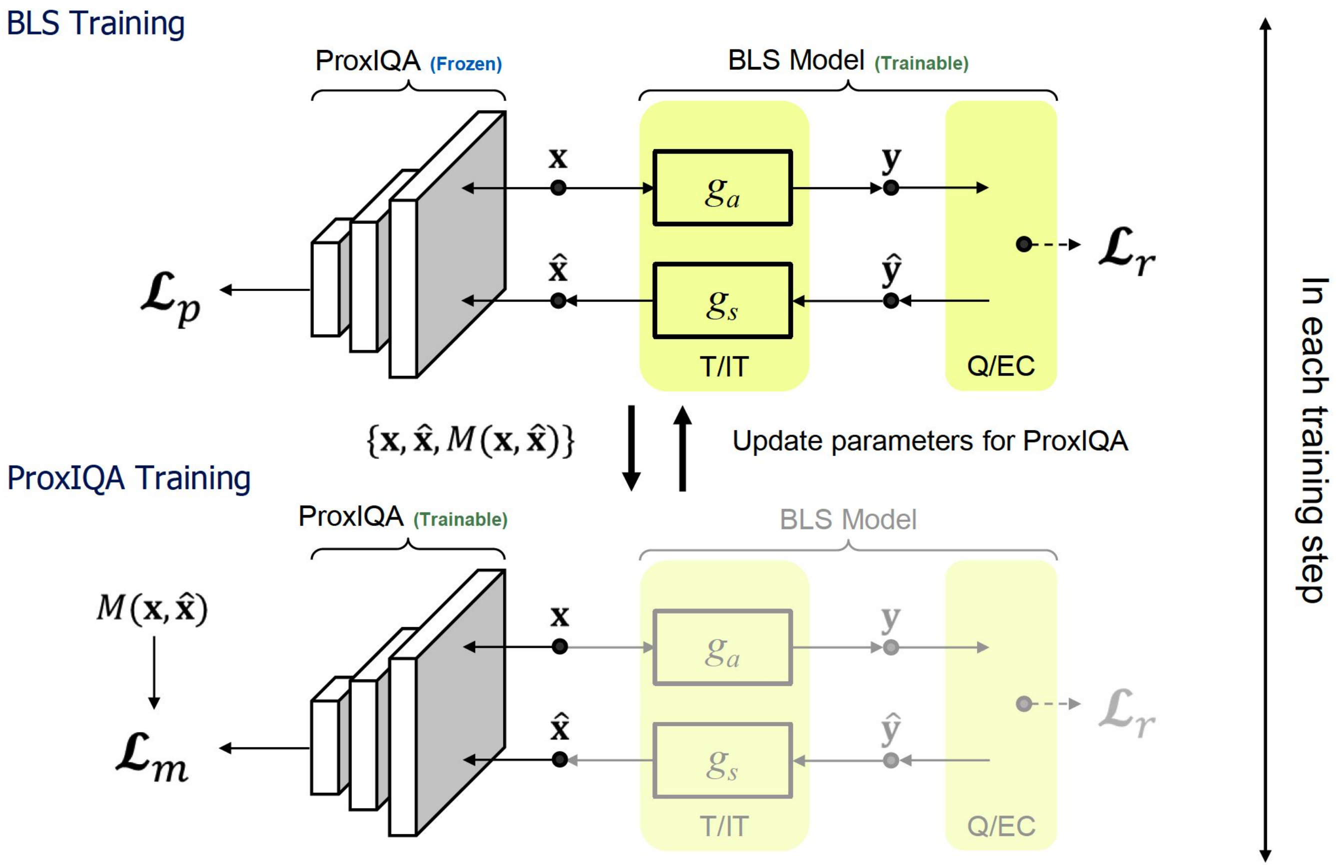
“Adversarial examples”

(a) Source Image



(b) Decoded Image
($VMAF_{Prox}$ = 97.74 and VMAF = 5.35)

Proposed alternating training



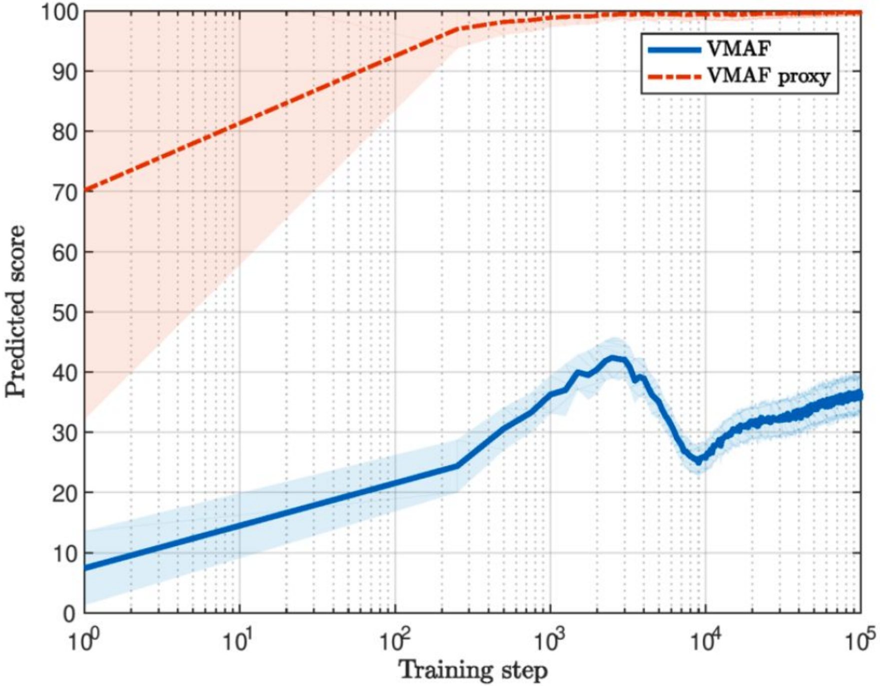
Training loss intuition

$$\mathcal{L}_{total} = \lambda \mathcal{L}_d + \mathcal{L}_r$$

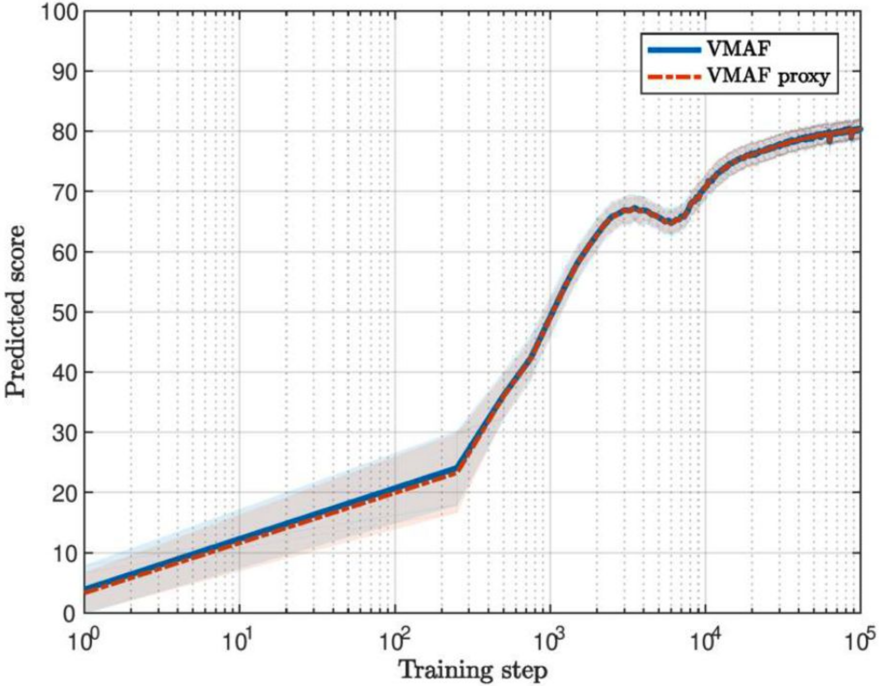
$$\downarrow$$

$$\mathcal{L}_{total} = \lambda \left[\alpha \mathcal{L}_p + (1 - \alpha) \mathcal{L}_d \right] + \mathcal{L}_r$$

Fixing the adversarial examples



(a) Pre-trained Model

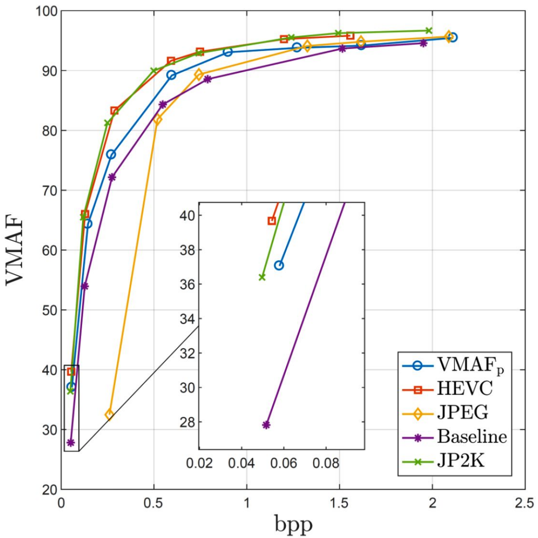


(b) Proposed alternating learning

Experimental results

Summary: **VMAF** BD-rate (%); ($\alpha = 0.00154$); Baseline: BLS model optimized for MSE;

Test image	JPEG	$BLS_{Baseline}$	$BLS_{VMAF\ Proxy}$	JP2K	HEVC Intra
$Mean_{Arithmetic}$	78.36%	0%	-23.35%	-33.39%	-28.23%
std	20.33%	0%	3.92%	8.77%	12.15%



Future work

- ➔ still need to beat state-of-the-art codecs, such as HEVC intra
- ➔ try on other generative modeling applications, e.g. de-noising, super-resolution, etc.
- ➔ gain better understanding of the distortions generated by these deep models
- ➔ video is a natural next step