Deep Generative Models for Distribution-Preserving Lossy Compression

Michael Tschannen (michaelt@nari.ee.ethz.ch)[1], Eirikur Agustsson[2], Mario Lucic[3]

[1]ETH Zürich [2]Google AI Perception [3]Google Brain

Abstract

We propose and study the problem of distribution-preserving lossy compression (DPLC). Motivated by the recent advances in extreme image compression which allow to maintain artifact-free reconstructions even at very low bitrates [1], we propose to optimize the rate-distortion tradeoff under the constraint that the reconstructed samples follow the distribution of the training data. Such a compression system recovers both ends of the spectrum: On one hand, at zero bitrate it learns a generative model of the data, and at high enough bitrates it achieves perfect reconstruction. Furthermore, for intermediate bitrates it smoothly interpolates between matching the distribution of the training data and perfectly reconstructing the training samples. We study several methods to approximately solve the proposed optimization problem, including a novel combination of Wasserstein GAN (WGAN) and Wasserstein Autoencoder (WAE), and present strong theoretical and empirical results for the proposed compression system.

The distribution-preserving lossy compression (DPLC) problem

Optimize the rate-distortion tradeoff under a distribution constraint

$$\min_{E,D} \quad \mathbb{E}_{X,D}[d(X, D(E(X)))] \quad \text{s.t.} \quad D(E(X)) \sim X, \qquad (1)$$

where

- $X \in \mathcal{X}$ models the data ,
- $E\colon \mathcal{X} \to \mathcal{W} := \{1, \dots, 2^R\}$ is a rate-constrained encoder,
- $D\colon \mathcal{W} \to \mathcal{X}$ is a (stochastic) decoder, and
- $d\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is a distortion measure.

Motivation: Artifact-free reconstructions at all bitrates thanks to the distribution constraint in (1), see [1] and Fig. 1.

Challenges

Known optimization approaches (relying on a Lagrangian formulation of (1)) fail to produce stochastic decoders even when feeding noise to D, see [2].

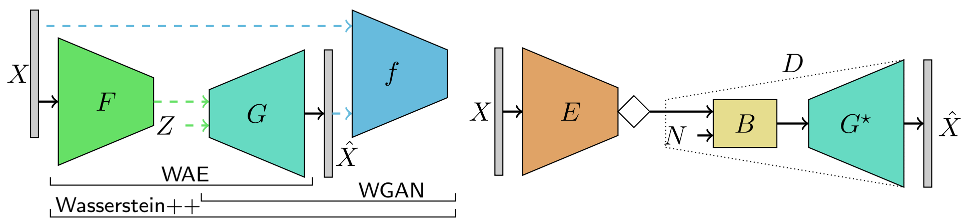

Figure 1: Learning G (left) and B, E (right). Stochasticity is provided by the noise N.

Proposed method to solve (1)

1. Learn a generative model G with a fixed prior $Z \in \mathcal{Z}$, $Z \sim P_Z$ of the data by minimizing the Wasserstein distance

$$W_d(P_X, P_{G(Z)}) = \inf_{\Pi \in \mathcal{P}(P_X, P_{G(Z)})} \mathbb{E}_{(X,Y)\sim\Pi}[d(X, Y)]$$

via WGAN, WAE, or the new **Wasserstein++** algorithm (see below).

2. Learn the stochastic function $B\colon \mathcal{W} \to \mathcal{Z}$ and E by solving

$$\min_{B,E\colon B(E(X))\sim P_Z} \mathbb{E}_{X,B}[d(X, G^\star(B(E(X))))]$$

and set $D = G^\star \circ B$.

Theorem

Suppose $\mathcal{Z} = \mathbb{R}^m$ and let G be K-Lipschitz. Then,

$$\begin{aligned} W_d(P_X, P_{G^\star(Z)}) &\le \min_{B,E\colon B(E(X))\sim P_Z} \mathbb{E}_{X,B}[d(X, G^\star(B(E(X))))] \\ &\le W_d(P_X, P_{G^\star(Z)}) + 2^{-R/m} KC, \end{aligned}$$

where $C > 0$ is an absolute constant. Furthermore, it holds for all $R \ge 0$

$$W_d(P_X, P_{G^\star(B(E(X))}) = W_d(P_X, P_{G^\star(Z)}).$$

Unsupervised training via Wasserstein++

Wasserstein++ combines WGAN and WAE to learn G by jointly optimizing a convex combination of the two objectives

- **Robust to mode dropping**, structured latent space thanks to WAE (primal) part
- **Sharp samples** thanks to WGAN (dual) part

Experiments

- DCGAN architecture, CelebA and LSUN bedrooms data set (64×64)
- Baselines: BPG (engineered), compressive autoencoder (CAE), and global generative compression (GC) [1]
- Evaluate distortion, sample quality, and stochasticity via mean-squared error (MSE), Fréchet inception distance (FID), and conditional pixel variance (PV), respectively

Results

- **The proposed method effectively solves the PDLC problem**
- **Wasserstein++ provides the best tradeoff between sample and reconstruction quality**

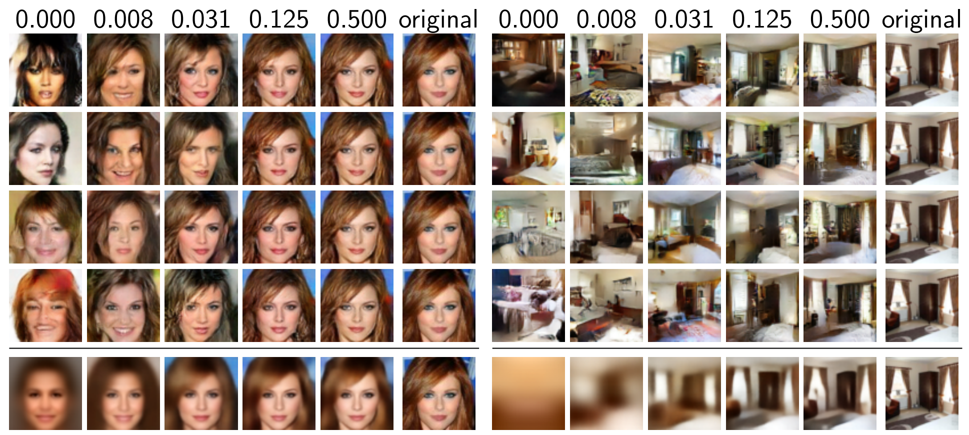

Figure 2: Example (test) reconstructions for CelebA (left) and LSUN bedrooms (right) obtained by our DPLC method based on Wasserstein++ (rows 1–4), and a CAE baseline (row 5), as a function of the bits per pixel.

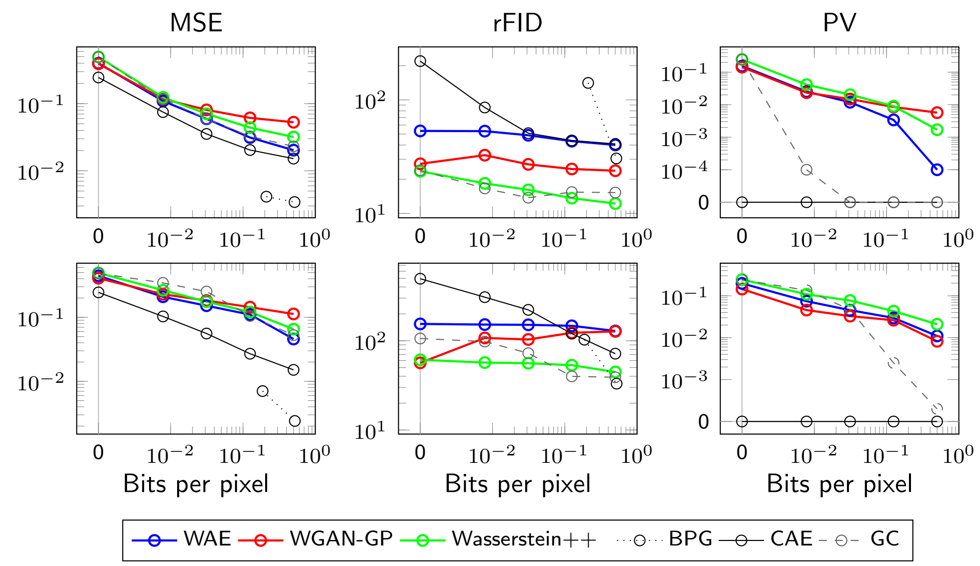

Figure 3: Testing MSE (smaller is better), reconstruction FID (smaller is better), conditional pixel variance (PV, larger is better) obtained by our DPLC model, for different generators $G^\star$, CAE, BPG, as well as GC [1], as function of the bitrate.

Table 1: Reconstruction FID and MSE (without rate constraint), and sample FID for the trained generators G, on CelebA and LSUN bedrooms (smaller is better).

	CelebA			LSUN bedrooms		
	MSE	rFID	sFID	MSE	rFID	sFID
WAE	0.0165	38.55	51.82	0.0099	42.59	153.57
WGAN-GP	/	/	22.70	/	/	45.52
Wasserstein++	**0.0277**	**10.93**	**23.36**	**0.0321**	**27.52**	**60.97**

Bibliography

[1] E. Agustsson et al. Generative adversarial networks for extreme learned image compression. arXiv:1804.02958, 2018.

[2] J.-Y. Zhuet al. Toward multimodal image-to-image translation. NIPS, 2017.