Evaluating generative networks using Gaussian mixtures of image features

Lorenzo Luzi[1,2], Carlos Ortiz Marrero[2], Nile Wynar[2]
Richard G. Baraniuk[1], Michael J. Henry[2]

[1]Rice University, [2]Pacific Northwest National Laboratory

Summary

We show that Inception-v3 features are not Gaussian and hence FID is misspecified. To remedy this problem, we model the featurized images using Gaussian mixture models (GMMs) and compute the **2**-Wasserstein distance restricted to GMMs, which we call WaM.

FID does not capture higher order moments

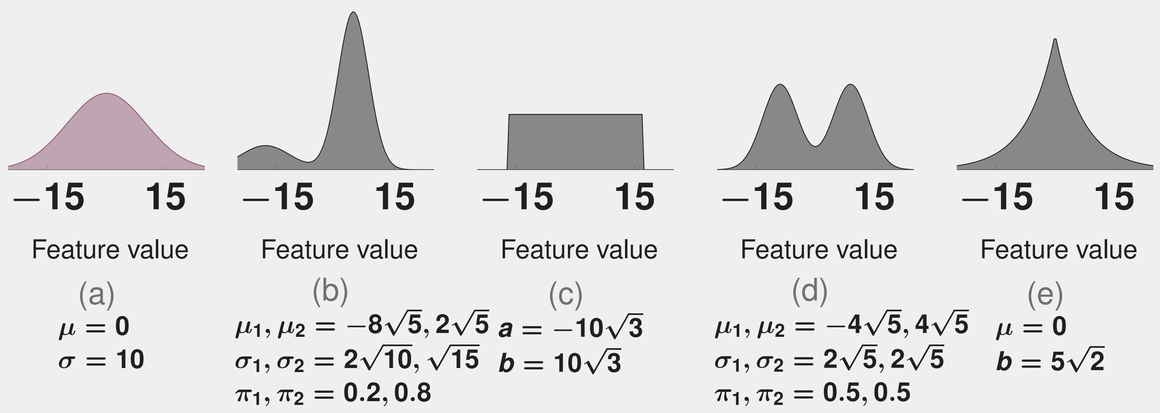


Figure 1: The FID score between each pair of the distributions shown above is zero although they are clearly different distributions. This is because FID is only defined for Gaussians, and treats any input distribution as Gaussian, even if it is not. All that is required for the FID score between two distributions to be zero is that their first two moments match. Figure 1a is the only Gaussian distribution. Figures 1b and 1d are Gaussian mixtures with two components, Figure 1c is a uniform distribution, and Figure 1e is a Laplace distribution. In contrast, GMMs can fit these distributions easily.

Inception-v3 Features are not Gaussian

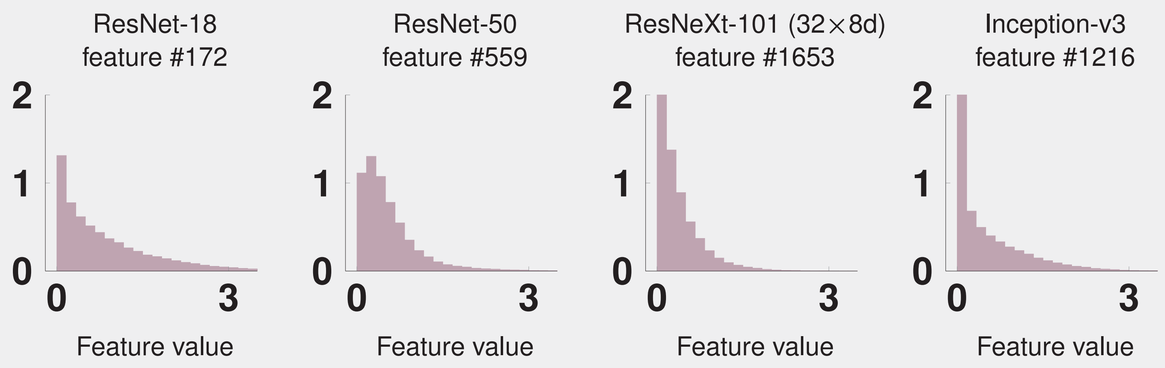


Figure 2: Histograms showing non-Gaussianity of randomly chosen features from the ImageNet validation dataset featurized by ResNet-18, ResNet-50, ResNeXt-101 (32×8d), and Inception-v3. They are non-negative because these features are passed through a ReLU and then average pooled; for this reason, we have a spike around **0**.

WaM - A Wasserstein-type metric on GMMs of image features which captures higher order moments

A closed form solution for the **2**-Wasserstein distance between GMMs is not known. However, if we restrict ourselves to the relaxed problem of only considering joint distributions over GMMs, then the resulting **2**-Wasserstein distance of this new space is known [1]:

$$\mathrm{MW}_2^2(\boldsymbol{P}, \boldsymbol{Q}) = \inf_{\gamma} \int_{\boldsymbol{X}\times\boldsymbol{X}} \boldsymbol{d}(\boldsymbol{x}, \boldsymbol{y})^2 \boldsymbol{d}\gamma(\boldsymbol{x}, \boldsymbol{y})$$

where the infimum is over all joint distributions γ which are also GMMs. **This reduces to a discrete optimal transport problem.** We use this metric on features from Inception-v3 (and other networks) to obtain our performance measure **WaM** which can capture higher order moments, past the mean and covariance.

Since WaM and FID are on different scales, we compare the two by seeing how much they change under certain perturbations. We define

$$\boldsymbol{R}_{\mathrm{FID}} = \frac{\mathrm{FID}_{\mathrm{PERT}}}{\mathrm{FID}_{\mathrm{ORIG}}}, \quad \boldsymbol{R}_{\mathrm{WaM}} = \frac{\mathrm{WaM}_{\mathrm{PERT}}}{\mathrm{WaM}_{\mathrm{ORIG}}}, \quad \text{and} \quad \boldsymbol{R} = \frac{\boldsymbol{R}_{\mathrm{FID}}}{\boldsymbol{R}_{\mathrm{WaM}}}.$$

Acknowledgements and References

Rice University affiliates were partially supported by NSF grants CCF-1911094, IIS-1838177, and IIS-1730574; ONR grants N00014-18-12571, N00014-20-1-2534, and MURI N00014-20-1-2787; AFOSR grant FA9550-18-1-0478; and a Vannevar Bush Faculty Fellowship, ONR grant N00014-18-1-2047.

References

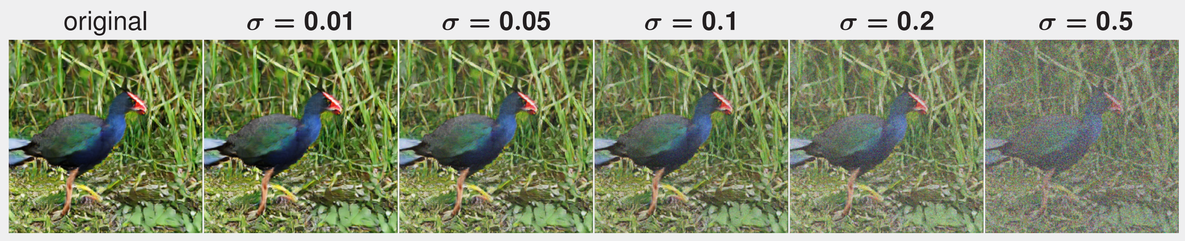
1. Delon, J. & Desolneux, A. A Wasserstein-type distance in the space of Gaussian mixture models. *SIAM Journal on Imaging Sciences* **13,** 936–970 (2020).

WaM is less sensitive to imperceptible perturbations

Original (BigGAN)	Perturbed (BigGAN)	Original (ImageNet)	Perturbed (ImageNet)
FID = **55.71**	FID = **154.19**	FID = **3.66**	FID = **46.63**
WaM^2 = **378.37**	WaM^2 = **424.29**	WaM^2 = **237.05**	WaM^2 = **280.02**
$\boldsymbol{R}_{\mathrm{FID}}$ = **2.77**		$\boldsymbol{R}_{\mathrm{FID}}$ = **12.74**	
$\boldsymbol{R}_{\mathrm{WaM}}$ = **1.12**		$\boldsymbol{R}_{\mathrm{WaM}}$ = **1.18**	
$\boldsymbol{R}$ = **2.47**		$\boldsymbol{R}$ = **10.78**	

Figure 3: Samples of images showing targeted perturbations which target the **feature means**. The two original images above are randomly selected from a set of 50,000 images generated by BigGAN and a set of 50,000 images of the ImageNet validation dataset. We cannot visually perceive the difference between the original and perturbed images, despite the datasets from which they were selected clearly demonstrating a drastic change in FID. The FID, WaM, and $\boldsymbol{R}$ values were calculated using Inception-v3.

WaM and FID perform similarly on random noise added to semi-realistic data (BigGAN generated)



	σ = **0.01**	σ = **0.05**	σ = **0.1**	σ = **0.2**	σ = **0.5**
FID(orig)	24.14	24.14	24.14	24.14	24.14
FID(pert)	24.37	27.10	33.55	51.10	114.94
WaM^2(orig)	504.30	504.30	504.30	504.30	504.30
WaM^2(pert)	539.54	516.75	628.68	748.65	1328.01
$\boldsymbol{R}_{\mathrm{FID}}$	1.01	1.12	1.39	2.12	4.76
$\boldsymbol{R}_{\mathrm{WaM}}$	1.07	1.02	1.25	1.48	2.63
$\boldsymbol{R}$	**0.94**	**1.10**	**1.11**	**1.43**	**1.81**

Figure 4: $\boldsymbol{R}$ values for BigGAN-generated images using additive isotropic Gaussian noise showing that FID is more sensitive than WaM to random noise perturbations of generated images. The noise perturbations in this experiment are all greater in magnitude than the targeted perturbations above. The original image on the left was randomly selected from a set of 50,000 images generated by BigGAN. The FID, WaM, and $\boldsymbol{R}$ values were calculated using ResNet-18.

WaM outperforms FID on random noise added to real data (ImageNet)

	σ = **0.01**	σ = **0.05**	σ = **0.1**	σ = **0.2**	σ = **0.5**
FID(orig)	3.61	3.61	3.61	3.61	3.61
FID(pert)	5.07	21.79	52.30	120.05	322.84
WaM^2(orig)	208.45	208.45	208.45	208.45	208.45
WaM^2(pert)	219.49	316.06	549.03	1081.28	4007.29
$\boldsymbol{R}_{\mathrm{FID}}$	1.41	6.04	14.49	33.26	89.45
$\boldsymbol{R}_{\mathrm{WaM}}$	1.05	1.52	2.63	5.19	19.22
$\boldsymbol{R}$	**1.34**	**3.98**	**5.50**	**6.41**	**4.65**

Figure 5: $\boldsymbol{R}$ values for real images (ImageNet validation data) using additive isotropic Gaussian noise showing that FID is significantly more sensitive than WaM to noise perturbations of real images. The noise perturbations in this experiment are all greater in magnitude than the targeted perturbations above. The original image on the left was randomly selected from a set of 50,000 images of the ImageNet validation dataset. In contrast to Figure 4, we see that **FID is more sensitive to these perturbations when the images look more realistic.** The FID and WaM values were calculated using ResNet-18.

{carlos.ortizmarrero,nile.wynar,michael.j.henry}@pnnl.gov, {enzo,richb}@rice.edu