On the use of Benford's law to detect GAN-generated images

POLITECNICO MILANO 1863

UNIVERSITÀ DEGLI STUDI DI PADOVA

Nicolò Bonettini*, Paolo Bestagini*, Simone Milani†, Stefano Tubaro*

*Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano

† Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova

Contacts: name.surname@{*polimi.it, †dei.unipd.it}

ICPR 2020 25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION Milan, Italy 10 | 15 January 2021

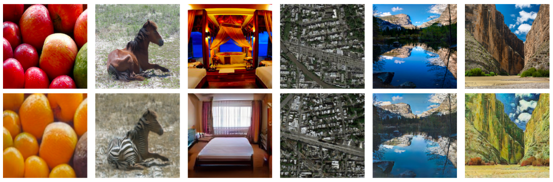
Detection of GAN-generated images

Problem

- Images generated by GANs can be very realistic

Goal

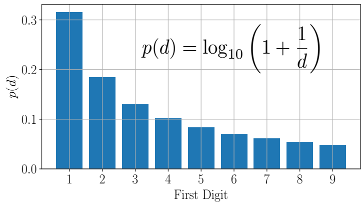
- To detect whether a picture is a natural one or it has been generated by a neural network



Method

Main idea

- GAN images may have different statistics from natural images
- Benford's law can capture these traces

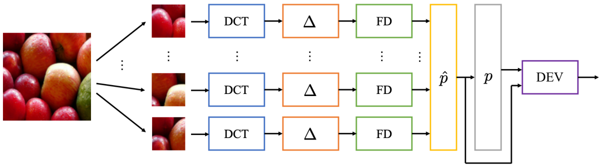


Proposed pipeline

1. Given a query image, compute FD of quantized DCT coefficients
2. Compute probability distribution through histogram computation
3. Fit theoretical Benford's curve
4. Use deviation between computed and theoretical distribution as a feature vector
5. Train a simple classifier (Random Forest) to discriminate between real and generated images

Feature vector

- Different feature vectors are generated by combining:
  - 9 possible DCT frequencies
  - 5 possible JPEG quality factors (QF)
  - 4 possible bases for computing the first digits



Dataset

- Publicly available dataset from [1]

Architecture	Dataset	Number of images
Cycle-Gan	orange2apple	1280
	photo2ukiyoe	4072
	winter2summer	1484
	zebra2horse	1670
	photo2cezanne	3978
	photo2vangogh	4099
	photo2monet	4765
	facades	259
	cityscapes	1996
	sats	684
ProGAN	lsun_bedroom	30770
	lsun_bridge	28768
	lsun_churchoutdoor	29120
	lsun_kitchen	42706
	lsun_tower	29020

Results

Uncompressed images

Dataset	Proposed	Xception	Steganalysis SVM [2]	Steganalysis RF
orange2apple	**98.13**	97.64	88.80	76.49
photo2ukiyoe	**100.00**	97.41	86.78	87.90
winter2summer	**100.00**	68.33	77.96	68.89
zebra2horse	**99.69**	89.58	91.01	77.00
photo2cezanne	**99.97**	95.91	95.88	93.17
photo2vangogh	**100.00**	93.75	94.68	92.93
photo2monet	**99.84**	94.08	94.80	89.87
facades	**100.00**	99.84	73.93	76.06
cityscapes	**100.00**	**100.00**	**100.00**	**100.00**
sats	**99.69**	73.00	90.92	96.93
lsun_bedroom	**100.00**	76.22	98.92	99.25
lsun_bridge	**99.89**	82.49	95.90	95.16
lsun_churchoutdoor	**99.99**	99.79	98.81	99.12
lsun_kitchen	**99.99**	87.26	99.49	99.59
lsun_tower	**99.98**	95.45	98.87	99.19
avg	**99.83**	89.64	91.03	90.11

Compressed images

QF	Dataset	Proposed	Xception
100	orange2apple	**94.50**	92.56
	photo2ukiyoe	**100.00**	98.50
	cityscapes	**100.00**	100.00
	lsun_tower	**100.00**	94.64
95	orange2apple	82.01	**90.66**
	photo2ukiyoe	97.00	**98.42**
	cityscapes	**99.99**	99.32
	lsun_tower	**99.80**	99.48
90	orange2apple	65.93	**85.61**
	photo2ukiyoe	92.01	**98.17**
	cityscapes	**100.00**	99.66
	lsun_tower	**99.60**	98.86

Faces (preliminary)

Dataset	Proposed
progan_celeba	79.75
stargan_black_hair	**97.26**
stargan_blond_hair	**96.56**
stargan_brown_hair	**96.76**
stargan_male	**96.24**
stargan_smiling	**96.06**
glow_black_hair	**86.56**
glow_blond_hair	**88.26**
glow_brown_hair	**86.18**
glow_male	**87.11**
glow_smiling	83.04
stylegan2-0.5	77.18
stylegan2-1	72.63
avg	**87.96**

[1] F.Marra, D.Gragnaniello, L.Verdoliva, G.Poggi, "Do GANs Leave Artificial Fingerprints?" *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2019

[2] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-Generated Fake Images over Social Networks," *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2018.