Evaluating Image Enhancement using Semantic Tasks

Leonardo Galteri, Lorenzo Seidenari, Marco Bertini, Alberto Del Bimbo
{leonardo.galteri,lorenzo.seidenari,marco.bertini,alberto.delbimbo}@unifi.it,

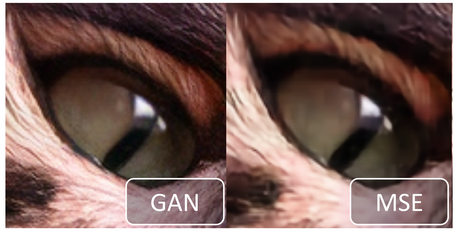
Overview

- Evaluation of generative models in the visual domain is often performed providing anecdotal results to the reader.
- Using signal based metrics often leads to counterintuitive results:
  - highly natural crisp images may obtain worse scores than blurry ones.

- Human based image assessment is expensive and time consuming.
- We advocate the use of language generation tasks to evaluate the quality of restored images.

Image Restoration

- We train a Generative Adversarial Network inspired by [1] for artifact removal.



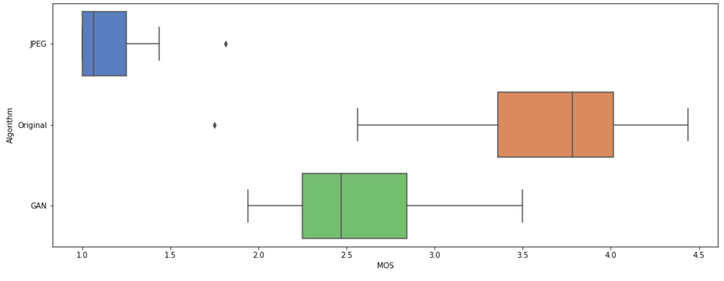
Evaluation Protocol

- Captioning maps the semantics of images into a much finer and rich label space represented by short sentences.

- We devise the following evaluation protocol for image enhancement:
  1. We pick an image captioning algorithm $\mathcal{A}$ [2] .
  2. We generate a sequence of words describing the image in detail.
  3. We look at performance of a captioning algorithm $\mathcal{A}$ on different versions of dataset: compressed, original and restored.

Results

Mean Opinion Scores



- According to human viewers, GAN is able to produce images perceptually of much higher quality than the images from which they are originated.

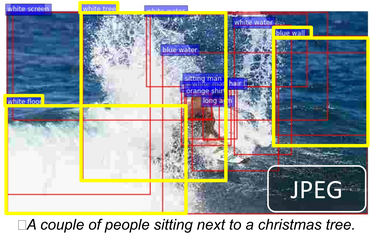
Reference and No-Reference Metrics

	PSNR	SSIM	NIQE	BRISQUE
JPEG10	**24.8245**	**0.7852**	6.36	53.17
GAN	23.8412	0.7605	**4.27**	**19.65**
ORIG	-	-	4.35	24.32

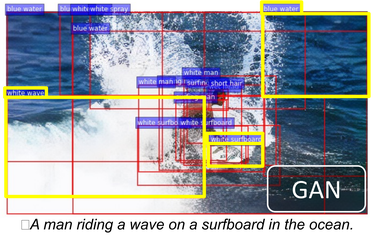
Language scores

	BLEU_1	BLEU_2	BLEU_3	BLEU_4	METEOR	ROUGE_L	CIDEr	SPICE	VIFIDEL
JPEG 10	0.685	0.500	0.360	0.250	0.220	0.490	0.810	0.150	0.309
GAN	**0.770**	**0.600**	**0.450**	**0.330**	**0.260**	**0.540**	**1.090**	**0.200**	**0.313**
ORIG	0.800	0.630	0.480	0.360	0.280	0.570	1.200	0.210	0.313

- Image details that are compromised by the strong compression induce errors in the captioning algorithm.
- GAN is able to recover an image which is not only pleasant to the human eye but recovers details which are also semantically relevant to a captioning algorithm.



A couple of people sitting next to a christmas tree.



A man riding a wave on a surfboard in the ocean.

Correlation

Metric	ρ
NIQE	-.84
BRISQUE	-.89
CIDEr	**.96**

- Correlation coefficient ρ between BRISQUE, NIQE, CIDEr and MOS for all versions of the images.
- A fine-grained semantic task as image captioning is a good proxy of real human judgment.

Bibliography

[1] Galteri et al., “Deep Universal Generative Adversarial Compression Artifact Removal.” *IEEE Transactions on Multimedia* (2019).
[2] Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering.“ *Computer Vision and Pattern Recognition* (2018).