Exploration and Implementations of Perceptual Image Metrics

Shawn Zixuan Kang
kangzx@stanford.edu

1 Introduction

Image compression faces the trade-off between the degree of compression (bit rate) and the distortion of a compressed image [1]. To evaluate the distortion, many image quality metrics have been developed. Metrics based on fundamental differences between pixel values are mean squared error (MSE) and peak-signal-to-noise ratio (PSNR). However, oftentimes value differences don't represent image compression quality. There are image quality metrics that focus on measuring the perceptual effects of images, such as SSIM [2], MS-SSIM [3], and VIF [4]. There are also emerging learned image perceptual metrics based on neural network, such as LPIPS [5].

In this project, I studied these important metrics, implemented them in Stanford Compression Library, and experimented on how different metrics reflect compression rate, distortion, and perceptual quality.

2 Literature

2.1 Image Compression Metrics

The most generic metric for comparing the errors of each corresponding pair of pixels between the reference image and the degraded noisy image, the Mean Squared Error (MSE) is represented as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||f(i,j) - g(i,j)||^2$$

where f and g represents the matrix data of the original image and the degraded image; m and n are the numbers of rows and columns of the image.

Based on this, peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value of a signal (value of the RGB channels in the case of image compression) and the power of distorting noise that affects the quality of its representation[6]. The intuition is to compare MSE with respect to the maximum value in decibel scale to adjust for the wide dynamic range. The higher the PSNR, the better the compressor in terms of reconstructing the original image.

$$PSNR = 20\log_{10}(\frac{\mathbf{MAX}_f}{\sqrt{MSE}})$$

While PSNR only relies on numeric comparison and does not actually take into account any level of biological factors of the human vision system, the Structural Similarity Index (SSIM) takes into accounts three key features from an image - luminance, contrast, and structure - that are relevant to how human perceive an image [7]. SSIM mathematically defines these three features and uses comparison functions (l, c, and s) to compare the features between the original and the degraded image.

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},\ c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},\ s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

where μ is the average pixel value (to represent luminance), σ is the standard deviation (to represent contrast), σ_{xy} is the covariance between the two images, and C_1, C_2 and C_3 are constants related with the dynamic range of the pixel values.

The SSIM score is given by:

$$SSIM = [l(f,g)]^{\alpha}[c(f,g)]^{\beta}[s(f,g)]^{\gamma}$$

where α, β, and γ are relative importance of each of the metrics. MS-SSIM (Mean Structural Similarity Index) divides the image into sections and applies a Gaussian weighting function to compute the local features, modeling the phenomenon that human focuses on the center more than the edge of the image, and finally computes the global features.

Visual Information Fidelity (VIF) is proposed to quantify the loss of image information in the distortion process to explore the relationship between image information and visual quality [4].

$$VIF = \frac{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^{N,j}; \overrightarrow{F}^{N,j} \mid s^{N,j})}{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^{N,j}; \overrightarrow{E}^{N,j} \mid s^{N,j})}$$

where $I(.)$ is the mutual information; $\overrightarrow{C}^{N,j}$ is the modeled image source; $\overrightarrow{F}^{N,j}$ and $\overrightarrow{E}^{N,j}$ are the outputs of the proposed human visual system model from image source and from after the distortion; $s^{N,j}$ is a realization of the scalars in the source model.

With the development of deep learning, we can look at images in a brand new way with the help of convolutional neural networks. Learned Perceptual Image Patch Similarity (LPIPS) utilizes exitet top-performing image neural network architectures (such as VGG and AlexNet) to compute the "perceptual distance" between two images [5]. Figure 1 illustrates the architecture of LPIPS.

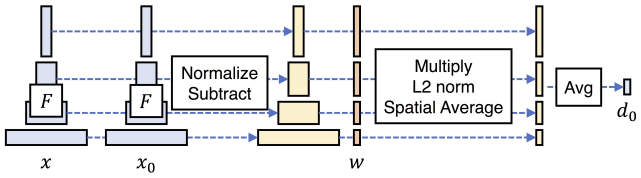


Figure 1: LPIPS Architecture

LPIPS takes the l_2 distance between the normalized activation layers in the network passed by x (original image) and x_0 (distorted image) scaled by w_i. It leverages the fact that neural network has extracted features from the images, which represents how human perceive images.

3 Methods

3.1 Implementation of metrics

I implemented the `ImageQualityMetrics` class in `utils/metrics.py` with documented APIs to evaluate the metrics results, retrieve all and each of the metric results as well as printable string formats.

The implementation includes PSNR, SSIM, MS-SSIM, and VIF. The code also uses APIs provided by the LPIPS paper's official repo [5] for the LPIPS metric. The `evaluate` method in `ImageQualityMetrics` class accepcts two images in NumPy array with form (H, W, C) and does format and dimension checks.

Verification of the implementation was done by running the metrics on example images from the dataset and the compression results, which is explained in the following subsections.

3.2 Dataset

In this project, to verify the metrics implementation and experiment with different metrics, I used the image database TID2013 [8]. This dataset has 25 reference images, and their distorted versions of 24 different types of distortion (Gaussian noise, contrast change, etc.), each with 5 different levels of intensity. Figure 2 shows a series of examples from TID2013 dataset.

Figure 2: Examples of a reference image under 5 levels of the same distortion (Gaussian noise)

This dataset allows us to first see how different metrics can reflect different types of distortion, which can occur in different image compression schemes. It also allows us to see how reactive each metric is under different levels but the same distortion, in the case of the same compressor but various compression rate.

3.3 Experiments

In the project, I intended to explore the metrics for:

- Images under different levels of intensity of common distortions incurred by image compressors
- Images compressed with different image compressors with different compression rates.

I conducted the following experiments with my metrics implementations:

- Run the metrics with images under JPEG compression with different compression rates
- Run the metrics with images under JPEG, BPG, and HiFiC (a more advanced learned image compressor [9]) compression with the same compression rate
- Run the metrics in TID2013 dataset and compare the results accross and within different types of distortion

In the TID2013 dataset experiments, I focused on the variances of each metrics under different levels but the same type of distortion. Since all images go through the same distortion schemes and each distortion has 5 different levels, the variances reflect how sensitive the metrics are to each type of distortion.

To visually better present the differences among the metric results, when plotting I normalized the metric values by dividing the maximum value of each metrics. I also reversed the LPIPS results since it's a distance metric and the smaller it is the closer the distorted image is perceptually to the original image, which is reverse to other metrics. See the next section for the results of the experiments.

4 Experiment Results

4.1 Metrics with JPEG under different compression rates

	Original	JPEG 20x	JPEG 30x	JPEG 40x	JPEG 76x
Image					
PSNR	-	-42.584	-42.583	-42.581	-42.696
SSIM	-	1.013	1.012	1.009	0.986
MS-SSIM	-	0.549	0.549	0.549	0.550
VIF	-	997	992	967	708
LPIPS	-	0.687	0.700	0.738	0.826

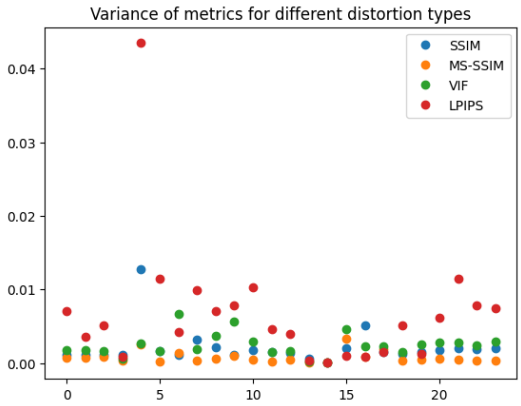
In this experiment, we found that PSNR, SSIM, or MS-SSIM can't signficantly show the increasing distortion incurred by the increasing compression rate under JPEG, while we can see significant changes in VIF and LPIPS results when the image quality changes, especially at compression rate of 76.

4.2 Metrics with JPG, BPG, and HiFiC under the same compression rate

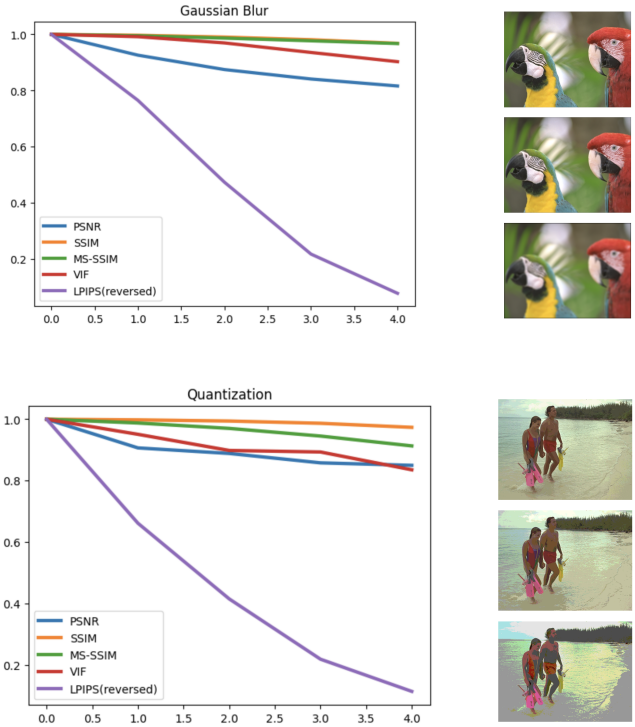
	Original	JPEG 76x	BPG 76x	HiFiC 76x
Image				
PSNR	-	-42.696	31.373	29.305
SSIM	-	0.986	2.997	2.996
MS-SSIM	-	0.550	0.989	0.985
VIF	-	708	0.911	0.926
LPIPS	-	0.826	0.166	0.038

In this experiment, we found that from all metrics we can see signifance changes from JPEG to BPG and HiFiC. However, while PSNR, SSIM, and MS-SSIM reflect that BPG is slightly better than HiFiC compressing this image, VIF and LPIPS show the opposite, and VIF and LPIPS in this case reflect what we actually perceive.

4.3 Metrics under different types of distortion in TID2013



This shows that in different distortion types (24 types, see [8]), LPIPS is the most sensitive to the level of distortion of the images.



Above are two common types of distortion in image compression, Gaussian blur and quantization, each with a sample series of images as examples. We found that while all metrics are able to reflect the level of distortion, we can see the most significant changes in LPIPS, followed by PSNR and VIF.

5 Conclusion

This report summarizes some of the major image quality metrics and the experiments I did about how they perform under different types of distortion and image compression schemes. In terms of different types of distortion, all metrics studied in this project are able to reflect the levels of distortion, with LPIPS being the most sensitive. However, in realistic image compression scenarios, the more advanced VIF and LPIPS show significant advantage in representing the compressed image quality.

In the future, I hope to evaluate more image compression algorithms with more metrics, in more detail exploring the advantages and disadvantages of different compressors in terms of the perceptual quality of the compressed images.

Github repo link: `https://github.com/shawwwwnK/SCL_perceptual_image_metrics`. This is a fork of the Stanford Compression `https://github.com/kedartatwawadi/stanford_compression_library`.

References

[1] Juan Carlos Mier, Eddie Huang, Hossein Talebi, Feng Yang, and Peyman Milanfar. Deep perceptual image quality assessment for compression, 2021.

[2] Dominique Brunet, Edward R. Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, 2012.

[3] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.

[4] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[5] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018.

[6] ni.com. Peak signal-to-noise ratio as an image quality metric. *ni.com*, 2020.

[7] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[8] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

[9] Fabian Mentzer, George Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression, 2020.

[10] Netflix Technology. Toward a practical perceptual video quality metric. *Netflix TechBlog*, 2016.

[3] [2] [8] [10]