Paper Review 8DM50

Sander Derwig
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Manon van Erp
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Rutger Hendrix
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Dirk Loeffen
IMAG/e
Eindhoven Technical University
s1252399
d.w.m.loeffen@student.tue.nl

Djennifer Madzia-Madzou
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

I. Application Domain (81)

One of the limitations of the resolution of magnetic resonance (MR) imaging is acquisition time. Acquisition time has to be limited in order for MR imaging remain convenient for patients. To increase resolution without increasing acquisition time, one could try one of several super-resolution (SR) techniques. A more recent technique to reach super-resolution is deep learning. The reviewed paper demonstrates two super resolution networks.

II. Methods (295)

To train the neural networks they have to be fed low resolution (LR) images. The (LR) images are acquired by downsampling and zero-filling the high resolution (HR) images in Fourier space. By doing so in Fourier space, the image size is unchanged but the image quality is degraded.

In this paper, two networks are trained. Both networks are a general adversarial network (GAN), thus they consist of a generator and a discriminator. The generator trains the SR, whereas the discriminator trains to qualify images as real or fake. The difference between the two networks lays in the generator. In the first network, the generator consists of an encoder, a decoder and a feature extraction network that is applied to other MR sequences. The features extracted from the other MR sequences are then fed into the network between the encoder and the decoder. The second network, applies the encoder-decoder network two times. In the second network the features from the other MR sequences are also fed into the network twice, once in each encoder-decoder network.

In the reviewed paper, four loss functions are used:

1) The adversarial loss in the generative adversarial network framework is used to train the generator.
2) The Mean-Squared Error evaluates the difference between the output of the generator and the corresponding ground truth at pixel-wise level. It can greatly improve the signal-to-noise ratio of generated images.
3) The perceptual loss overcomes the problem that some details may be lost due to over-smoothed SR results. The perceptual loss recovers more details by measuring image similarity in a high-level feature space.
4) The texture matching loss contributes to generate an image with great similarity between the output of the generator and the ground truth by statistically matching extracted features.

The image quality evaluation metrics that are used in this paper are structural similarity, peak signal-to-noise ratio and information fidelity criterion.

III. Discussion (179)

One common problem in training a Generative Adversarial Network is that it is highly unstable. This is inherently the case since often the generator and the discriminator are trained simultaneously while competing against each other. In the aforementioned paper the discriminator was trained four times before the generator was trained once. This might help stabilize the model.

Another strong point of the methodology is the use of multiple loss functions. Each loss function has its own (dis)advantages. By using the four loss functions, the power of each one is combined and results in a generally better applicable loss function for this application.

Using peak signal-to-noise ratio as a metric is sometimes problematic, since a higher PSNR usually indicates a reconstruction of higher quality, but this is not guaranteed. PSNR is also proven to be outperformed by most other popular evaluation metrics, so the use of PSNR does not seem necessary in this paper. But on the other hand it is beneficial to use multiple evaluation metrics to ensure that the model is working optimally and correctly.

IV. Recommendations (216)

Although MSE leads to a high signal-to-noise ratio in reference to the ground truth, it tends to produce over-smoothed SR results. Therefore they also use perceptual loss. Is there a way to combine these two to avoid one problem being created that has to be solved with another loss function.

Also two distorted images with the same MSE may have very different types of errors, some of which are much more visible than others. One error might be preferred over the other, which is now not addressed with the use of MSE.

PSNR is no longer regarded as a reliable indicator of image quality degradation [1]. Although a higher PSNR generally indicates that the reconstruction is of higher quality, in some cases it may not. Bear in mind that none of these objective measures are particularly good at predicting human visual response to image quality. Sometimes PSNRs vary wildly between two almost indistinguishable images. SSIM is recommended. So since this is already used we propose just skipping PSNR.

The Cognitive Interaction Problem. It is widely known that cognitive understanding and interactive visual processing (e.g., eye movements) influence the perceived quality of images. For example, a human observer will give different quality scores to the same image if given different instructions [1].

REFERENCES

[1] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.