Paper Review 8DM50
Multi-Contrast Super-Resolution MRI Through a Progressive Network [1]

Sander Derwig
Computational Biology
Eindhoven University of Technology
s0901783
s.derwig@student.tue.nl

Manon van Erp
IMAG/e
Eindhoven University of Technology
s0943658
m.v.erp@student.tue.nl

Rutger Hendrix
IMAG/e
Eindhoven University of Technology
s0902650
r.hendrix@student.tue.nl

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

Djennifer Madzia-Madzou
Image Sciences Institute
Utrecht University
s1526960
d.madziamazou@student.tue.nl

I. Application Domain

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. 1) A one-level non-progressive neural network and 2) two-level progressive neural network.

II. Methods

To train the neural networks they have to be fed low resolution (LR) images. The (LR) images are acquired by down-sampling 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.

To reach SR, 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 system 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 system.

In the reviewed paper the following 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 (SSIM), peak signal-to-noise ratio (PSNR) and information fidelity criterion (IFC). The SSIM is a fidelity metric which compares the terms of luminance, contrast and structural similarity. The second fidelity measure PSNR relates to the MSE and can be expressed as the ratio between the square of the maximum value and the MSE value. IFC measures image quality based on natural scene statistics. IFC utilizes the mutual information between the reference and a distorted image to evaluate image quality.

III. Discussion and Recommendations

One common problem in training a GAN 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 and regularization factors. 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.

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.

The perceptual loss part in the objective function measures image similarity in a high-level feature space. The pretrained VGG16 model on ImageNet is used to extract features from a given image. However these images are normal images of for example cars and so are not biomedical images. In our opinion extracting features and training a model on biomedical images would be more reliable.

PSNR is no longer regarded as a reliable indicator of image quality degradation, and so is a weak point [2]. 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 [2].

References

[1] Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C. T., Chan, M., Wang, G. (2020). Multi-contrast super-resolution mri through a progressive network. IEEE Transactions on Medical Imaging.

[2] 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.