Joint Multi-Anatomy Training of a Variational Network for Reconstruction of Accelerated Magnetic Resonance Image Acquisitions

Patricia M. Johnson[1], Matthew J. Muckley[1], Mary Bruno[1], Erich Kobler[2], Kerstin Hammernik[2,3], Thomas Pock[2], and Florian Knoll[1]

[1] Center for Biomedical Imaging, NYU Langone Health, Radiology Department, New York, NY, USA
[2] Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria
[3] Department of Computing, Imperial College London, London, United Kingdom

Abstract. Magnetic resonance imaging is a leading image modality for many clinical applications; however, a significant drawback is the lengthy data acquisition. This motivates the development of methods for reconstruction of sparsely sampled image data. One such technique is the Variational Network (VN), a machine learning method that generalizes traditional iterative reconstruction techniques, learning the regularization term from large amounts of image data. Previously, with the VN technique, reconstruction of 4-fold accelerated knee images was shown to be highly successful. In this work we extend the VN approach to applications beyond knee imaging and evaluate the classic VN and a newly developed Unet-VN in 5 different anatomical regions. We evaluate the networks trained individually for each anatomical area as well as jointly trained with data from all anatomical areas. The VN and Unet-VN were trained to reconstruct 4-fold accelerated images of knees, brains, hips, ankles and shoulders. SSIM was calculated to quantitatively evaluate the reconstructed images. Results show that the Unet-VN outperforms the classic VN, both quantitatively – in terms of structural similarity – and qualitatively. The networks jointly trained with multi-anatomy data approach the performance of the individually trained networks and offer the simplicity of a single network for a range of clinical applications which has substantial benefit for clinical translation.

Keywords: MR image reconstruction, Variational Network, Machine Learning

1 Introduction

The acquisition of Magnetic Resonance Image (MRI) data is an inherently slow process due to the high sampling requirements. Reconstructing images with sparser sampling has been, and continues to be, an active area of research in MRI. The major developments that have contributed to faster imaging are parallel imaging [1-3] and compressed sensing [4]. With parallel imaging techniques, the known sensitivities of multiple receive coils contribute to spatial encoding, and ultimately allow for an image reconstruc-

Preprint. See published version at https://doi.org/10.1007/978-3-030-33843-5_7.

tion from sparser sampling. Compressed sensing reconstruction is an extension of traditional iterative reconstruction methods which estimate images from under-sampled data by enforcing consistency with acquired data and applying regularization – a model of a-priori information about the reconstruction. In compressed sensing specifically, the regularization term enforces sparsity in some transform domain. Effective regularization is a key element for solving the under-sampled image reconstruction problem, however traditional regularization terms are often an over-simplification of MR image structure and offer limited a-priori information.

Recently, machine learning based approaches for sparsely sampled image reconstruction were introduced [5-9]; some of these methods use a convolutional neural network to learn the regularization term of an iterative reconstruction [5, 7]. They were designed to generalize the concept of compressed sensing and learn the entire reconstruction procedure for multi-channel MR data. One such method is the Variational Network, which has been demonstrated for successful reconstruction of 4-fold accelerated knee images [7, 10], and 3-fold accelerated abdominal images [11].

The first objective of this work is to extend the VN approach to applications beyond knee and abdominal imaging and evaluate the performance of a VN jointly trained with data of multiple anatomical regions. The simplicity of a single network for a wide range of applications would be a substantial benefit for clinical workflow. The second objective is to evaluate a newly developed version of the VN which consists of a higher model capacity regularizer.

2 Methods

2.1 Image Acquisition

All scans were performed on a clinical 3T system (Siemens Magnetom Skyra), with different receive coils ranging from 12 to 26 elements. Fifty fully-sampled anatomical images were obtained from 5 anatomical areas, these areas – ranked in order of perceived image SNR – were brain, knee, hip, ankle, and shoulder. The study was approved by our institutional review board. The sequence parameters were as follows:

Ankle – Sagittal fat-saturated proton-density (PD-FS): TR = 2800 ms, TE = 30 ms, turbo factor (TF) = 5, matrix size = 384 x 384, in-plane resolution 0.42 x 0.42 mm^2, slice thickness = 3.0 mm.

Brain – Axial T2: TR = 6000 ms, TE= 113 ms, TF = 18, matrix size 384 x 384, in - plane resolution = 0.57 x 0.57 mm^2, slice thickness = 5.0 mm

Hip – Coronal PD: TR = 3000 ms, TE = 32 ms, TF = 5, matrix size = 320 x 320, in - plane resolution = 0.5 x 0.5 mm^2, slice thickness = 3.0 mm

Knee – Coronal PD: TR = 2750 ms, TE = 32 ms, TF = 4, matrix size = 320 x320, resolution = 0.44 x 0.44 mm^2, slice thickness = 3.0 mm

Shoulder – Coronal fat-saturated T2: TR = 4540 ms, TE = 54 ms, TF = 12, matrix size = 320 x 320, in-plane resolution = 0.44 x 0.44 mm^2, slice thickness = 3.0 mm

The fully sampled images were then retrospectively under-sampled; the under-sampling was applied such that the center 24 lines of raw k-space data, and every fourth line beyond this center region were retained. The remaining k-space lines were set to zero. The center 24 lines were used for the ESPIRiT [12] estimation of coil sensitivities.

2.2 Variational network

Experiments were performed with two versions of the VN. The first is the classic VN, described in Hammernik et al. [7] and the second is a version in which the regularizer is replaced with a Unet network [13] (Unet-VN).

For this study, we implemented the classic VN in Pytorch, and replaced the IPALM optimizer [14], which was traditionally used for VN training, with the Adam optimizer [15]. The regularizer in this network is a single convolutional layer with 48 11x11 convolutional kernels. The activation functions are a learned set of Gaussian radial basis functions, and the model capacity is approximately 131, 000 parameters.

In addition to the classic VN network, we also evaluated a Unet-VN network which was designed to have much higher model capacity (1.2 million parameters). For this architecture, we replace the regularizer in the classic model with a Unet network; otherwise the VN method was unchanged. Our Unet implementation has 3 encoding convolutional layers followed by 3 decoding convolutional layers, with 24,48,96,48,24, and 12 3x3 convolutional kernels respectively. Max-pooling and bi-linear interpolation were used for dimensionality reduction and expansion respectively. We used ReLU for the non-linear activation function, and instance normalization was applied during training.

2.3 Network training

Individual trainings of the VN and Unet-VN were performed with 30 volumes of each anatomical region. Ten volumes for each dataset were reserved for a validation set and another 10 volumes were reserved for testing. Joint multi-anatomy training was performed with 6 volumes of each of the 5 anatomical regions for a total of 30 training cases. The Adam optimizer was used with a batch size of 1 and a learning rate of $3\text{x}10^{-4}$. We used Mean squared error as the loss function. Convergence (validation loss stops decreasing) for each training was achieved at a different number of epochs ranging from 60 to 100. Training was performed on a Tesla P100 GPU.

2.4 Evaluation of reconstructed images

We tested the trained networks on data from 10 image volumes per anatomical region. These cases were not included in the training set. We compare the VN and Unet-VN reconstructions with the fully-sampled reference, the zero-filled reconstruction and a combined Parallel Imaging, Compressed Sensing reconstruction method based on Total Generalized Variation (PI-CS TGV)[16] For all of the PI-CS TGV reconstructions, the regularization parameter was set to $4\ \text{x}10^{-6}$ and the number of iterations was 1000. We compared the reconstruction results quantitatively in terms of structural similarity index (SSIM)[17].

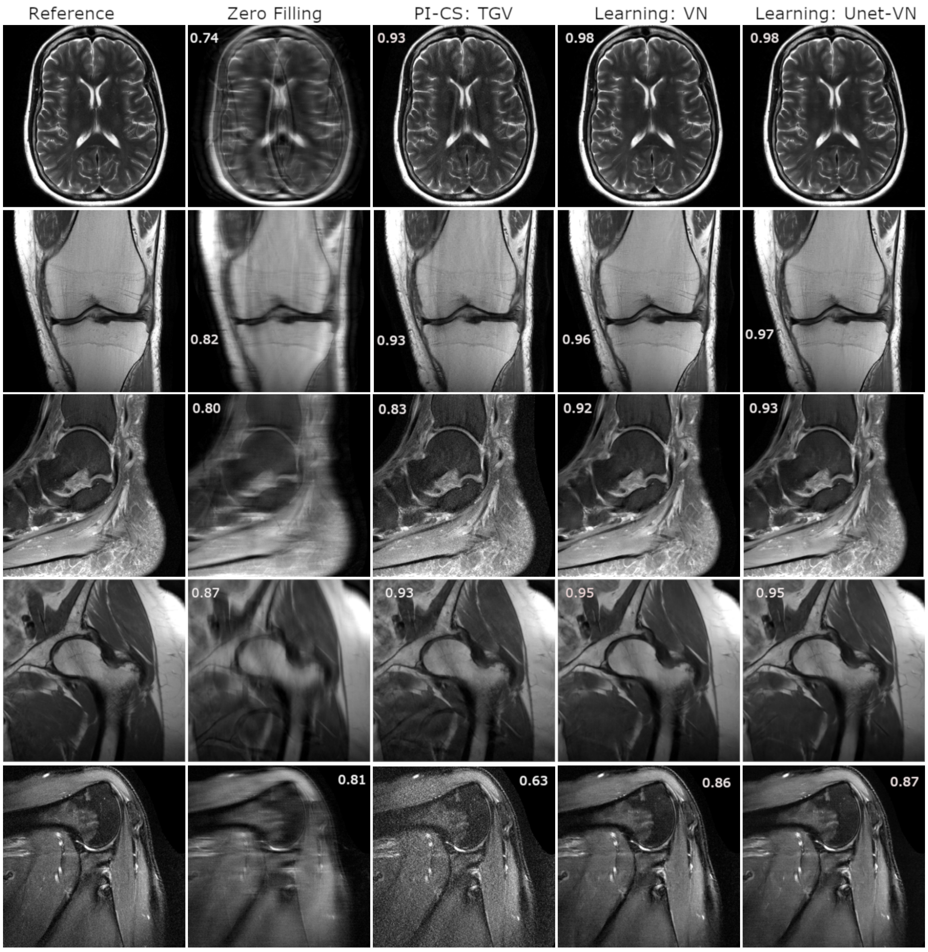
Table 1. *Structural similarity index was calculated for the 10 volumes in each test set; the mean and standard deviations are reported. Each of the 5 test sets were evaluated on all 12 trained networks*. The row labels are the training sets used, and the column labels are the test set data.

		Mean structural similarity of predicted images				
		brain	knee	ankle	hip	shoulder
VN training set	brain	0.976 (0.013)	0.965 (0.022)	0.948 (0.012)	0.948 (0.021)	0.830 (0.060)
	knee	0.971 (0.013)	0.974 (0.024)	0.947 (0.011)	0.948 (0.021)	0.843 (0.049)
	ankle	0.951 (0.010)	0.952 (0.012)	0.966 (0.006)	0.950 (0.017)	0.917 (0.023)
	hip	0.950 (0.009)	0.932 (0.013)	0.961 (0.007)	0.961 (0.013)	0.907 (0.027)
	shoulder	0.958 (0.010)	0.954 (0.016)	0.962 (0.006)	0.952 (0.017)	0.924 (0.020)
	all	0.970 (0.012)	0.964 (0.019)	0.966 (0.006)	0.958 (0.015)	0.922 (0.020)
		brain	knee	ankle	hip	shoulder
Unet-VN training set	brain	0.979 (0.013)	0.942 (0.012)	0.957 (0.007)	0.933 (0.019)	0.876 (0.040)
	knee	0.968 (0.015)	0.981 (0.021)	0.956 (0.007)	0.945 (0.020)	0.890 (0.031)
	ankle	0.951 (0.013)	0.866 (0.021)	0.970 (0.005)	0.941 (0.018)	0.917 (0.024)
	hip	0.899 (0.025)	0.893 (0.023)	0.899 (0.025)	0.965 (0.012)	0.888 (0.026)
	shoulder	0.925 (0.023)	0.909 (0.011)	0.950 (0.013)	0.939 (0.020)	0.929 (0.019)
	all	0.976 (0.014)	0.969 (0.017)	0.967 (0.005)	0.960 (0.015)	0.926 (0.019)

3 Results

The SSIM results for the VN and Unet-VN reconstructed images are reported in Table 1. We report the SSIM for all combinations of training and test data. For all anatomical regions, the highest SSIM is achieved with the individual, anatomy-specific, trained network. In these cases where the training and test anatomy are matched, the Unet-VN outperforms the classic VN. Image reconstruction results for the matched training and test sets are shown in Figure 1. The VN and Unet-VN both outperform the PI-CS TGV method.

Fig. 1. Brain, knee, ankle, hip and shoulder reconstructions with 4-fold acceleration. The learned reconstructions appear sharper and have less residual artefacts than the PI-CS TGV reconstructions. The displayed SSIM values were calculated for the presented slices.



When the training data and test data are not matched we observe an increase in residual artefacts in the reconstructed image and a decrease in SSIM. This is demonstrated in Figure 2 where we show knee images reconstructed with the VN individually trained with knee, brain, ankle and hip images. Another general trend that we observe when the training and test data is not matched is over-smoothing in the reconstructed images when training SNR < test SNR. When the opposite is true – training SNR > test SNR, noise amplification is observed. This effect is demonstrated in Figure 3. When the training and test data are not matched, the classic VN outperforms the Unet-VN for the majority of the training set/ test set combinations (16/20).

The performance of the joint multi-anatomy trained networks approached that of the individual trainings for each anatomy and the Unet -VN consistently outperformed the classic VN. Image results for the multi-anatomy training are shown in Figure 4.

Fig. 2. Coronal PD weighted knee scan with 4-fold acceleration. The top row depicts the reconstructed results for the classic VN trained with knee, brain, ankle and hip images. The bottom row shows the difference images compared to the fully sampled reference.

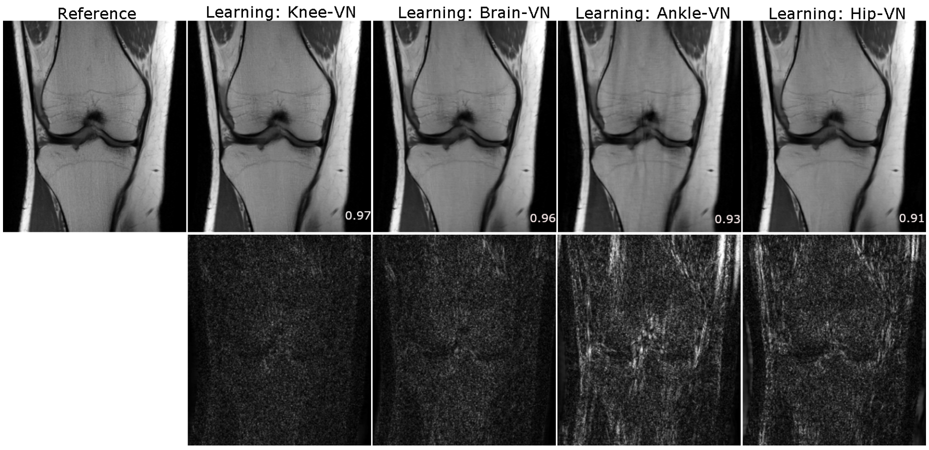


Fig. 3. Sagittal PD-FS ankle scan with 4-fold acceleration. Reconstruction results for the VN trained with ankle, brain (high SNR), and shoulder (low SNR). These results illustrate the trend that when training SNR > test SNR, the images suffer from noise amplification, and when training SNR < test SNR, the images appear over-smoothed.

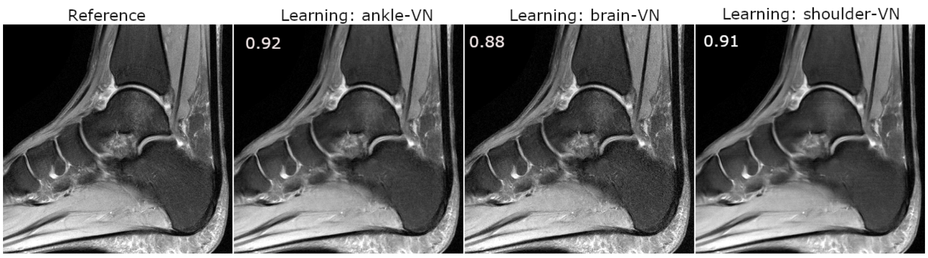
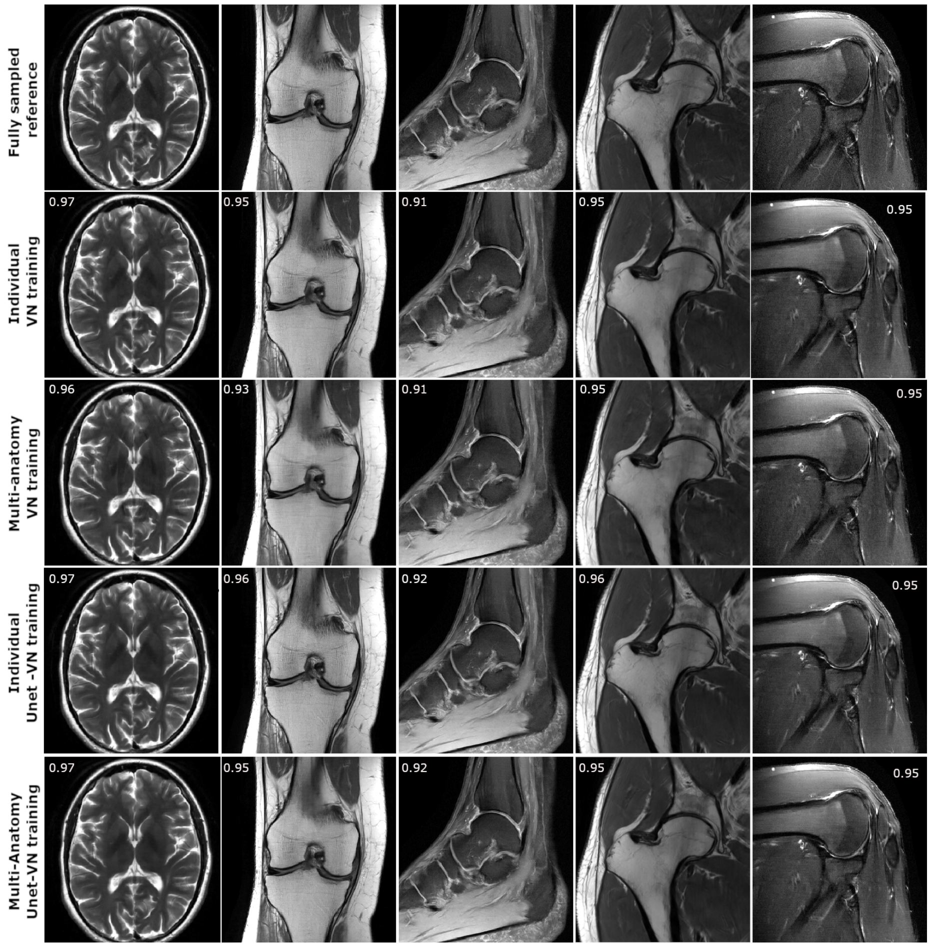


Fig. 4. Brain, knee, ankle, hip and shoulder reconstructions with 4-fold acceleration. The joint multi-anatomy trained networks result in similar reconstructed image quality as the individually trained networks. The Unet-VN matches or exceeds the classic VN for individual anatomy and multi-anatomy training



4 Discussion

The variational network outperformed the PI-CS TGV algorithm for reconstructions of 4-fold accelerated knee, brain, hip, ankle and shoulder images. The Unet-VN which has a higher model capacity regularizer than the classic VN, outperforms the classic VN for individual trainings when the test and training data are matched. In addition to higher model capacity, the Unet regularizer – with multiple convolutional layers – has a larger perceptive field than the classic single-layer regularizer. This may also contribute to the improved performance. The training time of the Unet-VN is approximately 25% longer than the training time of the classic VN. The Unet-VN network does not perform as well in most cases when the image being reconstructed is not represented in the training set, suggesting that the Unet-VN does not generalize as well to anatomical regions not previously seen by the network.

A specific trend is observed when there is a mismatch in the SNR of the training set and the test set; when the SNR of the training data is lower than the SNR of the test data, we observe over-smoothing in the reconstructed images. When the SNR of the training data is higher than the SNR of the test data, we see noise amplification in the reconstructed images. These findings are in agreement with a previous study that made a similar observation with fat-saturated (lower snr) and non – fat saturated (higher snr) knee images [10].

The networks that were jointly trained with multi-anatomy data have similar performance to those trained with a single anatomy, and again the Unet-VN outperforms the classic VN. A single network that can be used for many different clinical applications is not only beneficial for clinical workflow but also presents the opportunity for much larger training sets. In this study we used 30 images for joint multi-anatomy training in order to make fair comparisons with individual trainings; this approach does not take advantage of the 5x more training data that were available.

5 Conclusion

In this work, the classic VN and a newly developed Unet-VN were demonstrated for 4-fold acceleration of ankle, brain, hip and shoulder images and out-performed the PI-CS approach. The Unet-VN, with a higher model capacity regularizer, outperformed the classic VN for individual trainings as well as for joint multi-anatomy trainings. The networks jointly trained with multi-anatomy data had similar performance to those trained for a specific anatomy. Our findings suggest that the VN approach is a promising clinical tool for accelerated MR image reconstruction.

Acknowledgements

This work was supported by the US National Institute of Health under grants NIH/NIBIB R01EB024532 and P41EB017183 and the European Research Council starting grant HOMOVIS, No. 640156.

References

1. Griswold, M.A., et al., *Generalized autocalibrating partially parallel acquisitions (GRAPPA).* Magn Reson Med, 2002. **47**(6): p. 1202-10.
2. Pruessmann, K.P., et al., *SENSE: sensitivity encoding for fast MRI.* Magn Reson Med, 1999. **42**(5): p. 952-62.
3. Sodickson, D.K. and W.J. Manning, *Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays.* Magn Reson Med, 1997. **38**(4): p. 591-603.
4. Lustig, M., D. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging.* Magn Reson Med, 2007. **58**(6): p. 1182-95.
5. Aggarwal, H.K., M.P. Mani, and M. Jacob, *MoDL: Model-Based Deep Learning Architecture for Inverse Problems.* IEEE Trans Med Imaging, 2019. **38**(2): p. 394-405.
6. Akcakaya, M., et al., *Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging.* Magn Reson Med, 2019. **81**(1): p. 439-453.
7. Hammernik, K., et al., *Learning a Variational Network for Reconstruction of Accelerated MRI Data.* Magn Reson Med, 2018. **79**(6): p. 3055-71.
8. Yang, G., et al., *DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction.* IEEE Trans Med Imaging, 2018. **37**(6): p. 1310-1321.
9. Zhu, B., et al., *Image reconstruction by domain-transform manifold learning.* Nature, 2018. **555**(7697): p. 487-492.
10. Knoll, F., et al., *Assessment of the generalization of learned image reconstruction and the potential for transfer learning.* Magn Reson Med, 2019. **81**(1): p. 116-128.
11. Chen, F., et al., *Variable-Density Single-Shot Fast Spin-Echo MRI with Deep Learning Reconstruction by Using Variational Networks.* https://doi.org/10.1148/radiol.2018180445, 2018.
12. Uecker, M., et al., *ESPIRiT — An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA.* Magn Reson Med, 2014. **71**(3): p. 990-1001.
13. Ronneberger, O., P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science. **9351**.
14. Pock, T. and S. Sabach, *Inertial Proximal Alternating Linearized Minimization (iPALM) for Nonconvex and Nonsmooth Problems.* SIAM J. Imaging Sciences, 2016. **9**: p. 1756-1787.
15. Kingma, D.P. and J. Ba, *Adam: A Method for Stochastic Optimization.* arxiv, 2014. **1412.6980**.
16. Knoll, F., et al., *Second Order Total Generalized Variation (TGV) for MRI.* Magn Reson Med, 2011. **65**(2): p. 480-91.
17. Wang, Z., et al., *Image quality assessment: from error visibility to structural similarity.* IEEE Trans Image Process, 2004. **13**(4): p. 600-12.