

Attention-based Convolutional Neural Network for MRI Gibbs-ringing Artifact Suppression [★]

Maksim Penkin¹[0000–0002–8027–9333], Andrey Krylov¹[0000–0001–9910–4501], and
Alexander Khvostikov¹[0000–0002–4217–7141]

Lomonosov Moscow State University, Moscow, Russia

penkin97@gmail.com

kryl@cs.msu.ru

khvostikov@cs.msu.ru

<https://imaging.cs.msu.ru/ru>

Abstract. Gibbs-ringing artifact is a common artifact in MRI image processing. As MRI raw data is taken in a frequency domain, 2D inverse discrete Fourier transform is applied to visualize data. Inability to take inverse Fourier transform of full spectrum (full k -space) leads to the insufficient sampling of the high frequency data and results in a well-known Gibbs phenomenon. It is worth to notice that truncation of high frequency information generates a significant blur, thus some techniques from other image restoration problems (for example, image deblur task) can be successfully used. We propose attention-based convolutional neural network for Gibbs-ringing reduction which is the extension of recently proposed GAS-CNN (Gibbs-ringing Artifact Suppression Convolutional Neural Network). Proposed method includes simplified non-linear mapping, amended by LRNN (Layer Recurrent Neural Network) refinement block with feature attention module, controlling the correlation between input and output tensors of the refinement unit. The research shows that the proposed post-processing refinement construction considerably simplifies the non-linear mapping.

Keywords: Gibbs-ringing artifacts · Magnetic resonance imaging · Attention CNN · Image deringing.

1 Introduction

Gibbs-ringing artifact reduction is an image restoration problem, that can be solved by mathematical methods of image processing.

Gibbs oscillations (Gibbs phenomenon) often occur near high-frequency image features, for example, edges. Artifacts can be observed while mapping image on a finer grid, during contrast enhancement, video compression and MRI data

[★] The reported study was funded by RFBR, CNPq and MOST according to the research project №19-57-80014 (BRICS2019-394).

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

visualizing. Slight image distortions can be left invisible, while severe Gibbs artifacts may even create obstacles in patients diagnosing, if we refer to Gibbs oscillations caused by k -space (Fourier space) truncation of MRI frequency domain (see Fig. 1).

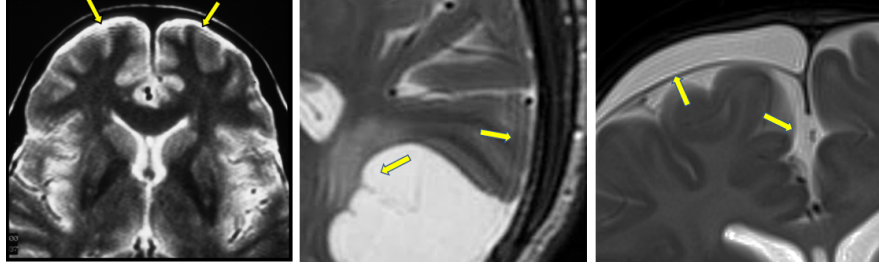


Fig. 1. Examples of Gibbs-ringing artifacts on MRI images. Arrows point out areas of distortion.

Simple finite real-valued periodic function can be observed to disclose mathematical reasons for Gibbs phenomenon:

$$\xi(t) = \begin{cases} a, & \text{if } t \in [-\tau/2, \tau/2] \\ 0, & \text{if } t \in [-T/2, T/2] \setminus [-\tau/2, \tau/2] \end{cases}, \quad (1)$$

$$\xi(t) = \xi(t + T),$$

where $\xi(t)$ is a basic model of a contrast edge, a is the amplitude of edge and T is the period of the model function.

Assuming Fourier transform in a complex form and $T = 2\tau$, (1) can be rewritten in a form:

$$\xi(t) = \sum_{k=-\infty}^{+\infty} d_k e^{i\omega_k t}, \quad (2)$$

where $d_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \xi(t) e^{-i\omega_k t} dt$, $\omega_k = \Omega k$, $\Omega = 2\pi/T$.

$$\xi(t) = 2 \cdot \frac{a}{2} \sum_{k=0}^{+\infty} (-1)^k \cdot \frac{1}{(2k+1)\pi} \cdot \cos(2k+1) \cdot \frac{2\pi}{T} t = \frac{a}{\pi} (\cos \Omega t - \frac{1}{3} \cos 3\Omega t + \dots), \quad (3)$$

where $\Omega = 2\pi/T$.

In practice it is impossible to include all terms in Fourier series (3), so Gibbs oscillations occur (see Fig. 2). The amplitude of Gibbs oscillations is constant for a given signal and doesn't depend on a chosen cut-off frequency.

In this paper we propose a new CNN architecture for MRI Gibbs-ringing suppression. It differs from recently introduced GAS-CNN [1] model by simplified architecture of non-linear mapping, followed by trainable LRNN [2] post-processing with attention block, which controls correlation between input and

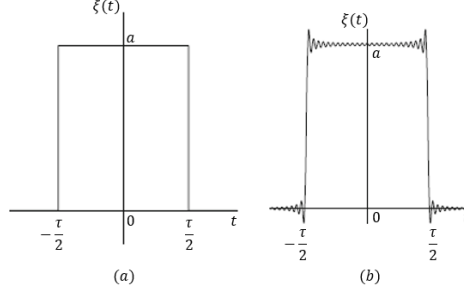


Fig. 2. Gibbs oscillations on a model edge function $\xi(t)$. (a) is artifact free signal, (b) is signal with Gibbs-ringing.

output tensors of the post-processing unit. The proposed architecture outperforms GAS-CNN on the generated synthetic testing dataset in terms of PSNR [3].

The remainder of this paper is organized as follows. In Section 2 we observe some known methods for MRI Gibbs-ringing suppression. In Section 3 we describe MRI dataset generation, give a detailed overview of the proposed architecture and show profit of involving our modifications to the architecture. In Section 4 the results and comparisons are presented. The work is concluded in Section 5.

2 Related Work

Gibbs-ringing reduction task has been solved by many methods so far. For example, the problem can be tackled as variational, and the solution can be searched as a function, which minimizes the stated functional in some functional space (L_2 or L_1 , for example):

$$J(u) = \frac{1}{2} \|u - u_0\|^2 + \lambda \int_{\Omega} |\nabla u(x)| dx \rightarrow \min_{u \in U}, \quad (4)$$

where u_0 is an input Gibbs-corrupted image, u is a searched Gibbs-free image from the chosen functional space U , Ω is the image's area and λ is the regularization parameter. The parameter can depend on the distance from the nearest image edge [4]. Joint ringing estimation and suppression can be performed using sparse representations [5].

Another recently introduced method is based on a search of optimal subpixels shifts [6]. The approach is intended to find a unique best shift for each pixel in terms of minimizing total variation in some predefined pixel's neighbourhood. They found the neighbourhood $K = [1, 3]$ to be sufficient for the most Gibbs-ringing cases. Proposed approach was visually compared by authors with median filter and Lanczos filtering, and it surpassed them.

Deep learning methods have acquired great popularity in computer vision and image processing nowadays. Convolutional neural networks map input images in

high dimensional feature spaces, implement filtering with a convolutional set and produce output images using final image reconstruction net.

GAS-CNN [1] is the example of very deep architecture used by authors to suppress Gibbs-ringing artifact on MRI images. Authors proposed it as the extension of the super-resolution model EDSR [7]. The following distinct model's features were presented by authors:

- external U-Net [8] like skip connections;
- decreasing the model's size (diminishing of the feature space dimension);
- flat architecture, as Gibbs oscillations are almost local phenomenon (rejection of spatial reduction layers, such as max pool or convolution with stride 2).

GAS-CNN maps input tensor into high dimensional feature space of depth 64 and then implements non-linear residual filtering with 32 ResBlocks [9]. Architecture is concluded with the simple reconstruction net, composed of one projection convolutional layer.

We chose this recently proposed model as a baseline and decided to conduct a research on the ways of non-linear mapping simplification with the maintenance of generalization ability. Despite of making quite extensive analysis of GAS-CNN, for example, showing the advantages in utilizing external skip connections and residual learning and making comparisons with other methods (sinc filtering, bilateral filtering [10], NLM [11], GARCNN [12]), authors of GAS-CNN didn't pay much attention on possible model's redundancy. It deserves to mention that the trend of recent years is to propose some hybrid refinement modules which make it possible to reduce amount of convolutions in the ensemble [2, 13, 14], as the straightforward excessive stacking of convolutional layers leads to learning degradation, vanishing gradients and so on.

So, in this article we demonstrate the way to shrink the number of convolutions in the non-linear mapping by two times and save (even improve) model's generalization ability, utilizing the proposed attention LRNN refinement module.

3 Proposed Architecture

The proposed architecture is shown in Fig. 3. We call it GAS14-ACNN (Gibbs-ringing Artifact Suppression Attention-based Convolutional Neural Network). It comprises of the following structural blocks: representing input corrupted samples into high dimensional feature space of depth 64 with the first convolution; performing non-linear mapping with 14 RCAN blocks [15] (two times less than in GAS-CNN); implementing trainable post-processing with the proposed attention LRNN refinement module and reconstructing output image with the final projection convolutional layer.



- apply Fourier transform to ground truth image 256×256 from IXI dataset;
- crop frequency spectrum: central $\frac{1}{9}$ part of frequency domain is saved;
- implement zero-padding, so that Gibbs-corrupted image fits the shape of ground truth image;
- apply inverse Fourier transform to get Gibbs-corrupted image.

Zero-padding is not a necessary step in Gibbs data generation, Gibbs-ringing can be synthesized just by cropping frequencies. In this work we include zero-padding to create image pairs $\{I_{gt}, I_{Gibbs}\}_{i=1}^N$ of the same spatial size. Zero-padding is often used before iFFT to project image on a finer grid, and zero-padding is often passed to project image on a coarse grid by inverse Fourier transform.

¹ <http://brain-development.org/ixi-dataset/>

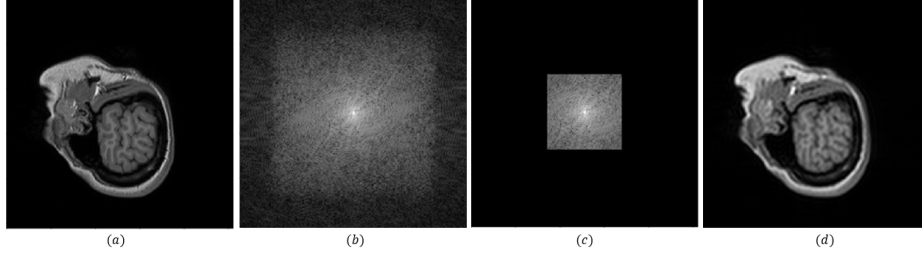


Fig. 4. Dataset generation pipeline. (a) is ground truth image from IXI dataset, (b) is Fourier spectrum of image (a), (c) is cropped and zero-padded Fourier spectrum, (d) is Gibbs-corrupted image.

tenth slice was obtained to produce pair $(I_{gt}, I_{Gibbs})_i$ [1]. So, training, validating and testing sets consist of 10427, 2016 and 1617 image pairs respectively. T1, T2 and PD have different data range, thus, maxmin normalization was used to map input features to a single band:

$$I_{min} = \min(I_{GT}), I_{max} = \max(I_{GT}), \quad (5)$$

$$I_{Gibbs\ normed} = \frac{I_{Gibbs} - I_{min}}{I_{max} - I_{min}}, \quad (6)$$

$$I_{GT\ normed} = \frac{I_{GT} - I_{min}}{I_{max} - I_{min}}, \quad (7)$$

where I_{Gibbs} is Gibbs-corrupted image, I_{GT} is ground truth image, $I_{Gibbs\ normed}$ is normed Gibbs-corrupted image and $I_{GT\ normed}$ is normed ground truth image.

3.2 Non-Linear Mapping

Images with Gibbs-ringing artifact obtain a significant blur also, as Gibbs-corrupted images are generated by high frequencies truncation. Noticed that RCAN structural module was successfully used in the recently published deep CNN architecture for image deblur [15], one of the proposed modifications to GAS-CNN is the replaced ResBlock [9] with RCAN (see Fig. 5) in non-linear mapping. The key difference of RCAN module is the presence of trainable weights for each slice of convolution output. These weights are generated applying global pooling operation (calculating expectation values over feature slices) and subsequently fusing acquired features by 1×1 ResBlock with sigmoid as a closing activation.

To the best of our knowledge the authors of GAS-CNN didn't provide code and weights, so to make fair comparisons we trained all presented here models ourselves, utilizing the same training procedure (refer to Section 4 for details) and the same synthetic generated dataset.

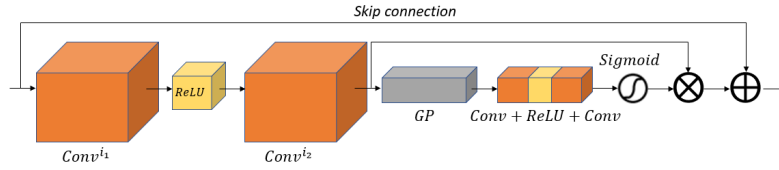


Fig. 5. RCAN module, used in non-linear mapping instead of GAS-CNN’s ResBlock.

GAS-CNN and RCAN-GAS-CNN were trained to evaluate RCAN performance. RCAN-GAS-CNN precisely matches GAS-CNN architecture with the only difference that RCAN module is used in non-linear mapping instead of an ordinary ResBlock. The performance growth is shown in Table 1 and in Fig. 6

Table 1. Average PSNR values on the generated testing set. Benefits estimation of RCAN block including to the non-linear mapping.

Model	Average PSNR
Initial Gibbs-corrupted images	29.79
GAS-CNN	32.04
RCAN-GAS-CNN	32.19

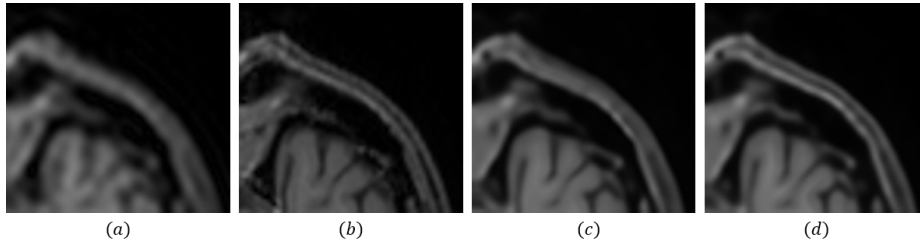


Fig. 6. RCAN influence on GAS-CNN architecture. (a) is Gibbs-corrupted image, (b) is ground truth image, (c) is GAS-CNN result, (d) is RCAN-GAS-CNN result.

3.3 Attention LRNN Refinement

LRNN refinement block was introduced in [2], and its ideas have been effectively incorporated into deblur [2] and depth generation pipelines [13].

We utilized LRNN approach in solving Gibbs-ringing problem and, moreover, extended it by attention mechanism, controlling correlation between LRNN input and output features. Such attention module aims to force LRNN to be a refinement block. Authors [14] used similar attention unit to improve biomedical image segmentation, nevertheless, our approach differs from the existing one [14] in the way we employ attention mechanism. For the current case, it has the sense of additional constraint to the refinement operation, whereas in [14]

it is applied to group features based on their correlation within a single feature tensor.

LRNN has two input tensors: feature tensor (non-linear mapping result), to be recursively processed, and weights tensor. We acquire weights by the auxiliary RNN weights generation net (see Fig. 3), which is trained end-to-end with the whole neural network. LRNN implements 4 recursive updates: left-to-right, right-to-left, top-to-down and down-to-top, applying the rule:

$$H_{t+1} := (1 - \omega) \cdot H_{t+1} + \omega \cdot H_t, \quad (8)$$

where H_t is the current processing row, if we refer to top-to-down or down-to-up operations. Subsequent concatenation and convolution fuse recursively processed tensors and conclude LRNN operation.

LRNN can be viewed as an alternative (hybrid) way to enlarge receptive field and to accumulate global spatial information within the layer. Despite of the locality of Gibbs oscillations, mentioned above, accounting overall information within the layer occurred to be very helpful and remarkably raised generalization ability of the architecture. We trained two extra CNNs to reveal LRNN advantages:

- RCAN8-GAS-CNN – model accurately matches with original GAS-CNN, but the number of blocks in non-linear mapping is heavily decreased by 4 times;
- RCAN8-GAS-CNN+LRNN $\times 2$ – previously stated model, extended by proposed LRNN refinement;

The obvious positive LRNN impact can be observed in Table 2 and in Fig. 7. RCAN8-GAS-CNN+LRNN $\times 2$ has the comparable performance with the original 4 times deeper GAS-CNN, whereas repealing of LRNN post-processing leads to algorithm’s degradation.

Table 2. Average PSNR values on the generated testing set. Benefits estimation of LRNN block including after the non-linear mapping.

Model	Average PSNR
Initial Gibbs-corrupted images	29.79
GAS-CNN	32.04
RCAN8-GAS-CNN	31.67
RCAN8-GAS-CNN+LRNN $\times 2$	32.19

Attention module is shown in Fig. 8. It gets three tensors as inputs: $x_1 \in \mathbb{R}^{C \times H \times W}$ – non-linear mapping output (LRNN input), $x_2 \in \mathbb{R}^{H \times W \times C}$ – LRNN output, $x_3 \in \mathbb{R}^{C \times H \times W}$ – transposed copy of x_2 . Attention block performs weighted regrouping of x_3 features in a manner to uplift feature at position i , mostly correlated with i^{th} feature of LRNN’s input tensor x_1 .

Assume f_1, \dots, f_C to be x_3 features. Introduce following latent variables $\mathcal{A} = \{a_1, \dots, a_C\}$. Value $a_i \in \{1, \dots, C\}$ defines x_3 feature, mostly correlated with

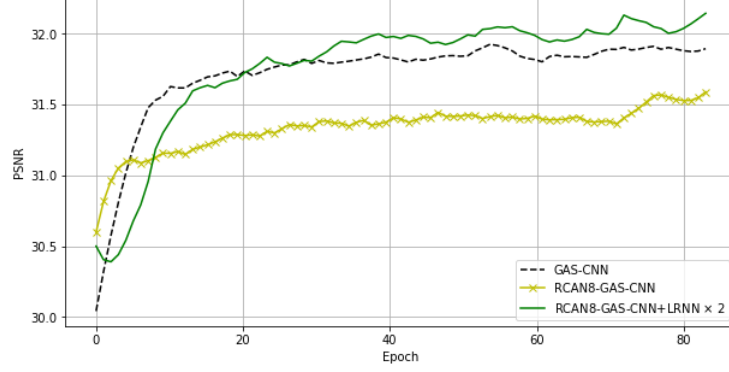


Fig. 7. PSNR on the validation set over epoch while training.

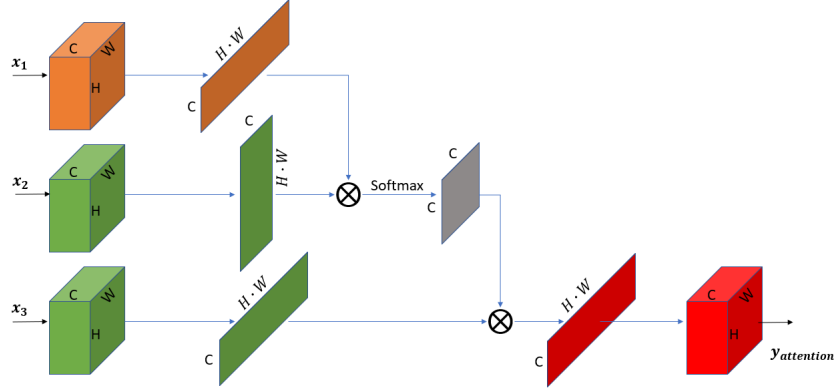


Fig. 8. Attention unit.

i^{th} feature of LRNN's input tensor x_1 . Weighted regrouping is executed via computed correlation tensor $C \in \mathbb{R}^{C \times C}$ (probability distribution map over the latent variables values).

It deserves mentioning that there is an analogy with word aligning task in classical machine learning (for example, IBM Model 1). Proposed attention LRNN unit performs like feature aligner of LRNN input and output tensors, forcing LRNN to be a refinement block.

Finally, we get the overall proposed architecture: RCAN14-GAS-CNN+Attention LRNN $\times 2$ (see Fig. 3). We call it GAS14-ACNN. Table 3 shows the increase of performance, caused by the attention unit.

Table 3. Average PSNR values on the generated testing set. Estimation of the performance growth, caused by the attention block.

Model	Average PSNR
Initial Gibbs-corrupted images	29.79
GAS-CNN	32.04
RCAN14-GAS-CNN+LRNN $\times 2$	32.57
GAS14-ACNN (proposed)	32.65

4 Experiments

Proposed attention-based convolutional neural network for Gibbs-ringing reduction was implemented in Python 3 with the use of deep learning framework Tensorflow 1.14. We provide implementation of our code at².

Models were trained with Adam optimizer [16] ($\beta_1=0.9$, $\beta_2=0.999$, $\varepsilon=1e-08$) on GPU NVIDIA GeForce RTX 2080 Ti. Learning rate had polynomial decay:

$$lr(x) = (lr_0 - lr_1) \cdot \left(1 - \frac{x}{M}\right)^p + lr_1, \quad (9)$$

where x – training step, lr_0 – initial learning rate value, lr_1 – final learning rate value, $M = N_e \cdot N_d / bs$ – amount of training steps, N_e – epochs number, N_d – amount of pairs in the training set, bs – number of pairs, fed to the algorithm on the current training step (batch size), p – polynomial power.

We used the following values of these parameters: $lr_0 = 10^{-4}$, $lr_1 = 0$, $N_e = 1000$, $bs = 20$, $p = 0.3$.

We applied L_1 loss function with l_2 weights regularization ($\gamma = 10^{-4}$) to prevent models’ overfit.

We utilized augmentation by rotations and flips for patches of shape (48×48) during training. 10 random patches were cropped from each training image before an augmentation. Validation and testing were performed on full-size images.

All convolutions’ kernels have spatial size 3×3 , except for one projection convolution just before LRNN refinement unit: it has kernel of spatial size 1×1 , and it projects features on some trainable manifold of the less dimension.

GAS-CNN, chosen baseline model, and the proposed GAS14-ACNN can be compared in Fig. 9 and Fig. 10. It takes approximately 1.03 sec. and 0.05 sec. to process one image (256×256) for GAS-CNN on CPU and GPU respectively³. And it takes approximately 1.13 sec. and 0.08 sec. to process one image (256×256) for our GAS14-ACNN on the same CPU and GPU respectively³.

5 Conclusion

We proposed the new attention-based convolutional architecture GAS14-ACNN for MRI Gibbs-ringing suppression. This architecture is the extension of recently

² <https://github.com/MaksimPenkin/GAS14-ACNN>

³ Intel(R) Core(TM) i7-8700; NVIDIA GeForce RTX 2080 Ti

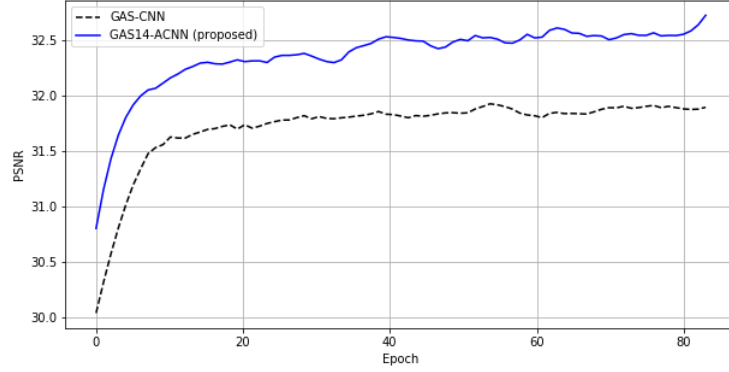


Fig. 9. PSNR on the validation set over epoch while training.

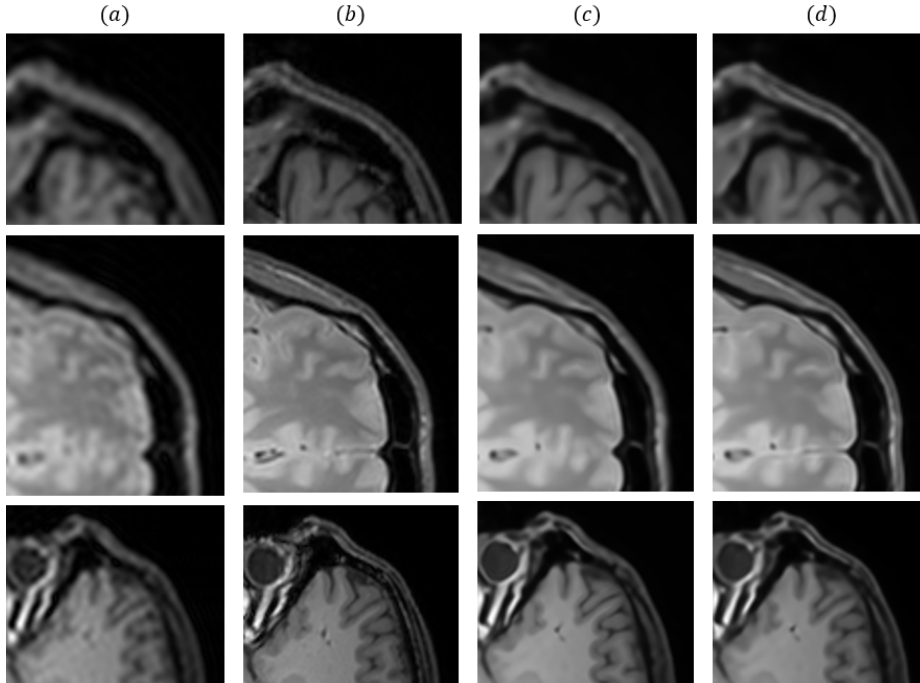


Fig. 10. Estimation of visual quality growth. (a) Gibbs-corrupted images, (b) ground truth images, (c) GAS-CNN results, (d) GAS14-ACNN results (proposed).

proposed GAS-CNN model with significantly simplified non-linear mapping, followed by attention LRNN unit to save generalization ability. The presented attention mechanism acts as auxiliary constraint for LRNN post-processing and as feature filtering module. The proposed GAS14-ACNN model outperforms baseline GAS-CNN on the generated synthetic testing set.

References

1. Zhao, X., Zhang, H., Zhou, Y., Bian, W., Zhang, T., Zou, X.: Gibbs-ringing artifact suppression with knowledge transfer from natural images to MR images. *Multimedia Tools and Applications*, 1–23 (2019)
2. Zhang, J., Pan, J., Ren, J., Song, Y., Bao, L., Lau, R. W., Yang, M. H.: Dynamic scene deblurring using spatially variant recurrent neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2521–2529. Salt Lake City, UT, USA (2018). <https://doi.org/10.1109/CVPR.2018.00267>
3. Al-Najjar, Y. A., Soong, D. C.: Comparison of image quality assessment: PSNR, HVS, SSIM, UIQI. *Int. J. Sci. Eng. Res* **3**(8), 1–5 (2012)
4. Sitdikov, I. T., Krylov, A. S.: Variational Image Deringing Using Varying Regularization Parameter. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications* **25**(1), 96–100 (2015)
5. Umnov, A. V., Krylov, A. S.: Sparse Approach to Image Ringing Detection and Suppression. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications* **27**(4), 754–762 (2017)
6. Kellner, E., Dhital, B., Kiselev, V. G., Reiser, M.: Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic resonance in medicine* **76**(5), 1574–1581 (2016)
7. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144. Honolulu, HI, USA (2017). <https://doi.org/10.1109/CVPRW.2017.151>
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778. Las Vegas, NV, USA (2016). <https://doi.org/10.1109/CVPR.2016.90>
10. Zhang, M., Gunturk, B. K.: Multiresolution bilateral filtering for image denoising. *IEEE Transactions on image processing* **17**(12), 2324–2333 (2008)
11. Manjón, J. V., Coupé, P., Buades, A., Fonov, V., Collins, D. L., Robles, M.: Non-local MRI upsampling. *Medical image analysis* **14**(6), 784–792 (2010)
12. Wang, Y., Song, Y., Xie, H., Li, W., Hu, B., Yang, G.: Reduction of Gibbs artifacts in magnetic resonance imaging based on Convolutional Neural Network. In: *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*, pp. 1–5. Shanghai, China (2017). <https://doi.org/10.1109/CISP-BMEI.2017.8302197>
13. Cheng, X., Wang, P., Yang, R.: Learning depth with convolutional spatial propagation network. *arXiv preprint arXiv:1810.02695* (2018)
14. Sinha, A., Dolz, J.: Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, arXiv preprint arXiv:1906.02849 (2020)
15. Park, D., Kim, J., Chun, S. Y.: Down-scaling with learned kernels in multi-scale deep neural networks for non-uniform single image deblurring. *arXiv preprint arXiv:1903.10157* (2019)
16. Da, K.: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)