2026 JULY 19–22 VANCOUVER, BC

JOINT AAPM | COMP MEETING

"Comparison of Two Deep-Learning Neural Networks of Single-Image Super-Resolution to Enhance Low-Resolution 3D Cine MR Images for Reconstruction of Time-Resolved 4DMRI"

Danial Kelsey, BS[1], Josh Korb, BS[1], Shaune Faraday, BS[1], Jana Elwassif, BS[1], Xingyu Nie, PhD[2], and Guang Li, PhD[1]*

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, NY 10065

[2]Department of Radiology, University of Kentucky School of Medicine, Lexington, KY 40536

ABSTRACT

Object: To enhance the image quality of low-resolution 3D cine magnetic resonance (MR) images acquired at 2Hz by applying two single-image super-resolution deep-learning neural networks, so that the enhanced 3D cine images serve as better templates for more accurate deformable-image-registration-based reconstruction of time-resolved 4DMRI images.

Methods: The Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) and Very-Deep Super-Resolution (VDSR) neural networks were applied to enhance low-resolution 3D cine MR images of 20 subjects. First, the performance of the two models was tested using published optimal hyperparameter sets: The models were trained with 800 pairs of low- and high-resolution 2D photographic images and tested on 100 additional picture pairs. Second, MR image pairs with diaphragm domes within 0-2mm were identified and processed using ImageJ. Patient-specific models were built with 250-400 2D slice pairs and tested with 50-100 slices per subject in three subjects. 5000 iterations, 10 epochs, 0.1 learning rate, 32 patch size, and Adam optimizer were applied to train both models. The enhanced image quality was compared with cubic B-Spline interpolation referenced to the high-resolution ground truth, using Peak-Signal-to-Noise-Ratio (PSNR), Structural Similarity Index (SSIM), Natural Image Quality Evaluator (NIQE), Root-Mean-Square-Error (RMSE), and visual comparison.

Results: The ESRGAN and VDSR models outperform the 3D B-Spline interpolation in both pictures and MR slices. For pictures, VDSR consistently improved in all 4 indexes by 1-5% and visually, but ESRGAN has a significant improvement in NIQE (-39%) and in visual comparison. For MR slices, VDSR enhances images in all 4 indexes and visually by 2.5-12.5%, while SSIM (7.5%), NIQE (-30%), and visual show improvement.

Conclusion: We have demonstrated that ESRGAN is more effective than VDSR in enhancing the quality of 3D cine MR images. Although enhanced 3D cines are useful as TR-4DMRI, they can serve as better templates to produce more accurate time-resolved 4DMRI images.

CONTACT

Guang (George) Li, PhD, DABR

Department of Medical Physics
MSKCC, 1275 York Ave, NY, NY10065
lig2@mskcc.org

INTRODUCTION

Time-resolved 4-dimensional magnetic resonance imaging (TR-4DMRI) provides respiratory images over multiple breathing cycles without binning artifacts for better motion management. It allows assessment of breathing irregularities, offers more accurate internal tumor volume (ITV), and has the potential for respiratory-gated radiotherapy (RGRT) with improved treatment accuracy. To reconstruct TR-4DMRI images, a super-resolution technique has been applied to combine high-temporal and high-spatial images using deformable image registration (DIR).

To achieve a 2 Hz frame rate, several acceleration approximations were applied, decreasing the low-resolution image quality and affecting TR-4DMRI image reconstruction. In this study, we aim to build a 3D cine enhancer to improve the TR-4DMRI reconstruction. Two deep-learning-based, single-image super-resolution networks were applied to enhance 3D cine MR images, and their results were analyzed and compared.

METHODS AND MATERIALS

Two deep-learning-based neural networks - very deep super-resolution (VDSR)[1] and enhanced super-resolution generative adversarial network (ESRGAN)[2] - were applied to improve the resolution of the 3D cine MR images.

Four indices were used to quantify 2D/3D image enhancement, including the natural image quality evaluator (NIQE), structure similarity (SSIM), peak signal-to-noise ratio (PSNR), and root mean square error (RMSE).

METHODS AND MATERIALS (con't)

For 2D photographic images, they were used to compare the ability of the two networks to boost image resolution. 800 pictures were used for training and 100 for testing. For 3D cine MR images, paired low-high resolution images were selected between 3D cine and BH images. 250-400 slice pairs were used to build patient-specific VDSR and ESRGAN models, and 50-100 pairs were used for testing and evaluation. Similar hyperparameters were applied in both model trainings using the same image sets.

Two dedicated computer workstations (Intel Xeon CPU 2.11 GHz, 64GB RAM, 12GB GPU, and 2TB SSD) were used for this deep-learning study. The model training took up to 48 hours, depending on the training data size. The hyperparameter settings are described below: 20 convolutional layers,10 Epochs, SGDM optimizer, variable learning rate, and 5000 iterations.

RESULTS

Fig. 1 shows the comparison of the quality of 5 images, including low-res, B-Spline, VDSR, ESRGAN, and high-res ground truth. ESRGAN achieves the best image enhancement, while VDSR and B-Spline interpolation yield similar results. These findings are similar to those observed in 2D photographic images.

The high-resolution image was used as the ground truth in calculating the indices. The NIQE index can better distinguish image resolution differences among these images, showing a wider dynamic range, compared with SSIM, as shown in Fig. 2. The quantitative assessment is consistent with visual evaluation. The other two indices were also calculated but were inferior to NIQE and SSIM and were not reported here.

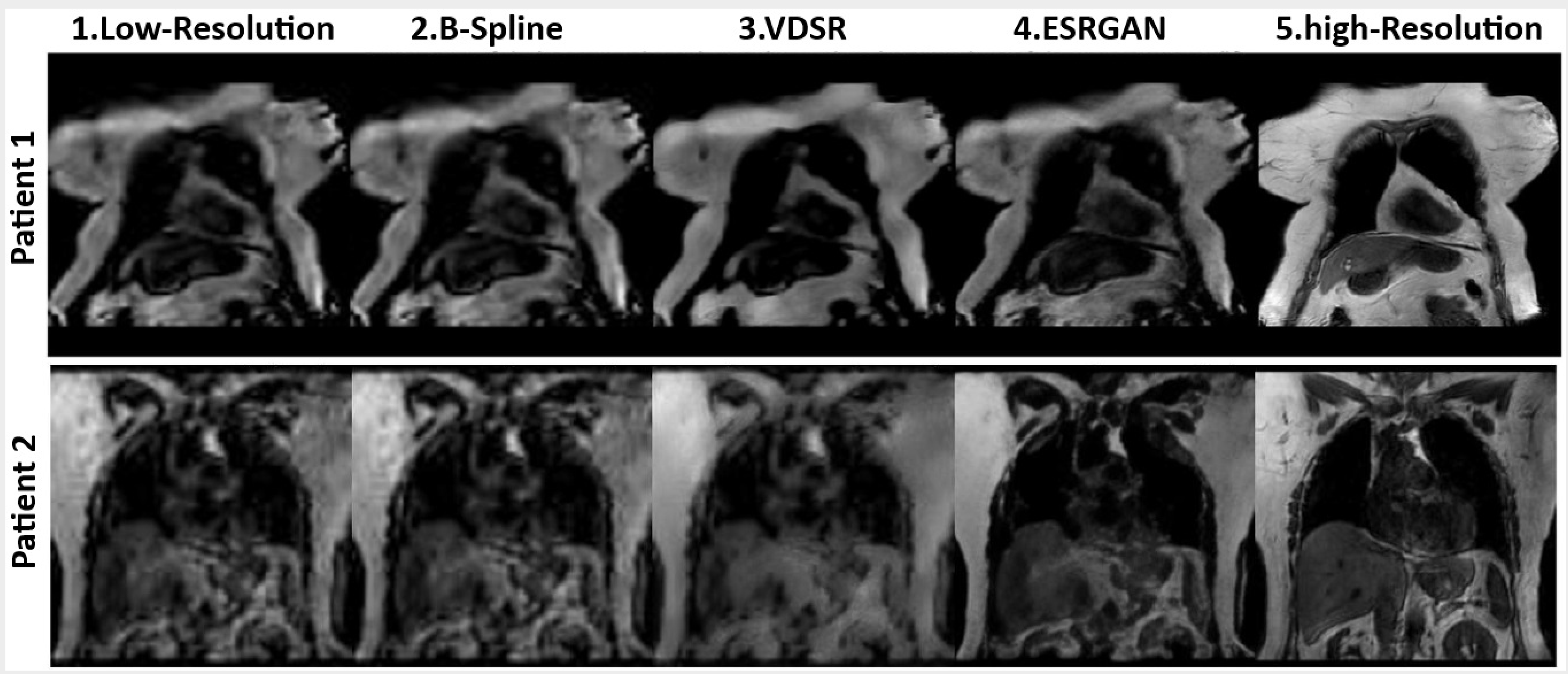

Figure 1. Comparison of image enhancement using B-spline, VDSR, ESRGAN techniques, and ground truth.

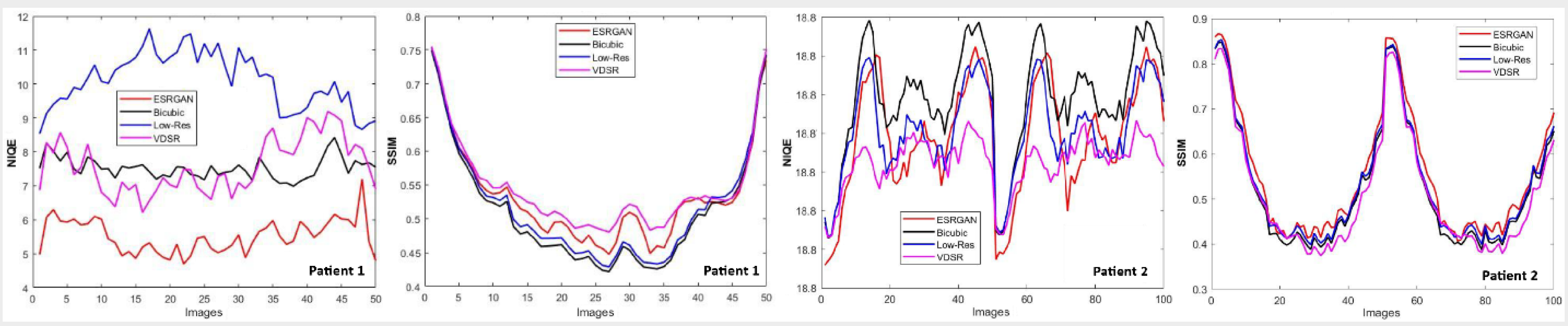

Figure 2. The SSIM increases, and the NIQE decreases, when VDSR and ESRGAN are applied to patients 1 & 2.

DISCUSSION

The preliminary results have shown that both VDSR and ESRGAN can enhance the image resolution to a certain degree, while ESRGAN outperforms VDSR. The conventional B-Spline interpolation was also applied, which can be used as a benchmark to assess the two deep-learning models. Currently, only the reported hyperparameter settings, suitable for photographic pictures, were used in the model training, and further study will explore optimal hyperparameters regarding medical MR images. As the model training takes a very long time (up to 48 hours per model), searching for optimal hyperparameters would be difficult, and using more powerful computer workstations with more GPU memory would be a good option.

This is an ongoing investigation to incorporate more patient MR data to maximize the enhancement of the 3D cine images using these models. Currently, the deep-learning models are patient-specific, as models are trained with one patient only. However, it would be more desirable to create population-based models so that the models are more generic and can be used for any patient. Therefore, we will aim for such models in the near future.

Originally, it was expected that the single-image super-resolution approach could fully recover the resolution from the low-resolution 3D cine images, so that no DIR would be required. However, the results from this study demonstrate that the ultimate goal may not be reachable at once. Therefore, the aim was set to improve the 3D cine images to provide an improved target image with a higher resolution and fewer artifacts. Therefore, the TR-4DMRI image quality would be improved. After the 3D cine image quality is improved, we will apply DIR to reconstruct the TR-4DMRI. Because the improved low-resolution 3D cine images will provide more landmarks for DIR, the overall quality of the TR-4DMRI is expected to be improved. Therefore, the improvement of the TR-4DMRI image quality would help to move one more step toward clinical applications.

CONCLUSIONS

The preliminary results have shown that both VDSR and ESRGAN can boost the 3D cine image quality to a certain degree, serving as better target images for deformable image registration in the reconstruction of TR-4DMRI images. The ESRGAN performs better than VDSR, and NIQE differentiates the resolution change better than SSIM, PSNR, and RMSE, showing the highest dynamic range to quantify the image enhancement. Although neither VDSR nor ESRGAN could boost image resolution near the ground truth, they can help to recover details that can be used as additional landmarks for alignment.

REFERENCES

1. 1. Hwang JJ, Jung YH, Cho BH, et al. Very deep super-resolution for efficient cone-beam computed tomographic image restoration. Imaging Sci Dent 2020;50:331–337.
2. 2. Zhou Z, Ma A, Feng Q, et al. Super-resolution of brain tumor mri images based on deep learning. J Appl Clin Med Phys 2022;23:e13758.