Autoencoding Low-Resolution MRI for Semantically Smooth Interpolation of Anisotropic MRI

Paper by: Jörg Sander, Bob D. de Vos, Ivana Išgum
Article synthesis by:
Hamdi BEL HADJ HASSINE HAMDI.BELHADJHASSINE@ENSAE.FR

Unless otherwise specified, super-resolution in this summary refers to through-plane super-resolution, i.e. increasing the resolution of anisotropic MRI by upsampling in the low-resolution direction of the 3D images. 'Resolution' also refers to the through-plane resolution of the images.

1. Background

1.1. Methodological topic

This paper deals with interpolation of magnetic resonance images (MRI), which can be linked to medical image super-resolution and upsampling. It proposes an algorithm that can generate intermediate slices that smoothly interpolate between two spatially adjacent MRI slices.

1.2. Medical problem

MRI acquisition involves a trade-off of spatial and temporal resolutions, signal-to-noise ratio and acquisition time and therefore cost. Anisotropic MRI have lower through-plane resolution than in-plane because increasing the through-plane resolution substantially raises the time and cost of the acquisition. However, having high through-plane resolution is also useful for many applications such as improving the accuracy of structure volume measurements, data augmentation for training neural networks, and generally enhancing the quality of the images for downstream analysis of the development of the studied structures and disease diagnosis. As a result, significant research efforts have been dedicated to MRI super-resolution in order to obtain high through-plane resolution from images acquired with low resolution.

1.3. State-of-the-art on the studied problem

MRI super-resolution problem can be linked to generation, compressed sensing and registration. Here the authors use an autoencoder for slice generation which is similar to what we have seen in the generation course, and the problem can also be addressed with other generation methods such as GANs (e.g. used in (Pham et al., 2019)). In the course we have also seen how compressed sensing and dictionary learning can be used in a denoising context to recover information from undersampled k-space observations assuming sparsity, and such methods can be adapted to improve the spatial resolution of MRI (e.g. as in (Lustig et al., 2007)). Registration is another concept from the course that can be applied here to match the structures in adjacent slices and interpolate between their boundaries.

The paper provides an extensive literature review. Most of the approaches to this problem can be divided into image combination methods and learning methods. Conventional methods (linear, B-spline, Lanczos resampling) fall within the first case and are still used due to their ease of use and being readily applicable with no particular requirements on data acquisition. Other combination methods such as (Rousseau et al., 2006) typically use registration to interpolate between MRI slices and achieve good results when the slices are aligned. The first learning methods were based on dictionary learning such as (Bhatia et al., 2014) and aim to produce a non-linear mapping between LR and HR samples. Recently these methods were outperformed by deep learning models, using different architectures such as GANs or U-Nets (Masutani et al., 2020), along with some unsupervised or self-supervised methods like (Zhao et al., 2018) which learn the structural information available in the samples to interpolate.

Although diverse, existing approaches have a few limitations: conventional methods are local transformations that do not exploit high-level contextual information from the images such as the shape or the texture of the captured structures. They also lead to cross-fade artifacts where misaligned objects fade and reappear instead of smoothly moving from the first to the second position. Image combination methods typically perform poorly when the slices are misaligned, and learning-based methods usually require a large amount of high-resolution data to effectively learn the patterns, and such data can be hard to obtain and may not present the same characteristics (luminosity, resolution, etc..) as the data to which the methods are applied, which decreases their performance. To overcome this, the authors propose an unsupervised method that can be easily applied to the low-resolution images we want to upsample.

2. Objectives and main contributions

2.1. Objective of the paper

The authors propose an unsupervised method to generate intermediate slices that interpolate between two spatially adjacent MRI slices in a smooth and semantically meaningful way without requiring high-resolution training data.

2.2. Main contributions

- Unsupervised method: doesn't need a high-resolution training dataset
- Simple to implement and fast to train and to infer
- Smooth interpolation (outperforms cubic B-spline)
- Able to generate an arbitrary number of slices

2.3. Data

Different public MRI datasets were used to train and validate the model in order to provide a more comprehensive assessment of the proposed method. The voxel values of all datasets were scaled to [0,1] to harmonize their intensities.

- Cardiac cine MRI dataset: A collection of two datasets: The first is taken from the MICCAI 2017 Automated Cardiac Diagnosis Challenge and comprises 100 cardiac cine MRIs. These MRIs have an in-plane resolution of 1.37 to 1.68 mm and a through-plane resolution of 5 to 10mm. The second is taken from the Sunnybrook Cardiac dataset and comprises 45 samples with in-plane resolution of 1.25mm and through-plane resolution of 8mm. Both datasets contain both healthy and pathological samples.

- Neonatal brain MRI dataset: This dataset is provided by the Human Connectome Project and comprises samples for 508 newborns. These images have an isotropic resolution of 0.5mm

- adult brain MRI dataset: This dataset is provided by the OASIS project and comprises MRIs from 416 patients, with a 1mm resolution.

3. Methodology

3.1. Description of the methodology

The main idea in this paper is to train an autoencoder to obtain latent space representations of the MRI slices and perform the interpolation in the latent space, as illustrated in Fig. 1 of the paper. We denote by f the encoder and by g the decoder. Given two input slice images x_1 and x_2, the encoder produces their latent representations $z_1 = f(x_1)$ and $z_2 = f(x_2)$. Then depending on the number K of intermediate slices to be generated, the interval $]z_1,z_2[$ is split into K equally spaced points $z'_1, .., z'_K$ which are then decoded using the decoder. For example, to generate 1 slice y_1, we set $z'_1 = \frac{1}{2}z_1 + \frac{1}{2}z_2$ then the generated slice is $y_1 = g(z'_1)$, and to generate 2 slices we set $z'_1 = \frac{2}{3}z_1 + \frac{1}{3}z_2$ and $z'_2 = \frac{1}{3}z_1 + \frac{2}{3}z_2$, then we have $y_1 = g(z'_1)$ and $y_2 = g(z'_2)$. In other words, the interpolation is done using evenly spaced convex combinations of the two input slices in the latent space, then decoded.

The model used in this paper is a convolutional autoencoder (CAE) (Masci et al., 2011), which differs from the original autoencoder architecture by using convolutional layers instead of fully-connected layers. The authors also compare with other architectures but show that CAEs perform better for this task. The authors implement a two-block encoder architecture where each block includes two layers with leaky ReLU activations, batch normalization, average pooling, then for the second block two additional convolutional layers. The decoder architecture mirrors that of the encoder, replaces the average pooling with nearest neighbor upsampling and uses a sigmoid function to scale the outputs of the last layer.

The authors introduce a loss function with two terms: $\mathcal{L} = \mathcal{L}_{\text{Reconstruction}} + \mathcal{L}_{\text{Synthesis}} = d_{MSE}(x_i, g(f(x_i))) + \lambda d_{LPIPS}(x_i, y_i)$. The first term is the usual reconstruction loss computed as the MSE of the autoencoder's output, while the second is the Learned Perceptual Image Patch Similarity (LPIPS) distance metric introduced in (Zhang et al., 2018) computed between the ground truth training slice x_i and its corresponding synthetic slice y_i generated by interpolating between x_i's adjacent slices x_{i-1} and x_{i+1}.

3.2. Motivation behind the components used

The main idea of the paper is to perform the interpolation in a latent space that can better represent the transition between the slices, which is why the paper relies on autoencoders

to produce a latent representation of the inputs. The use of a CAE originates from the idea that convolutions help better capture the high-level information in the images, and this is empirically proven in the paper's experiments. Besides, the use of the synthesis loss induces a semantic similarity measure between the interpolated images and the corresponding ground truth and incites the model to produce smooth and realistic transitions.

We believe that the choices made by the authors are coherent and properly motivated, except for the autoencoder hyperparameters (number of blocks, layers, kernel sizes) which are not explained nor validated in the paper.

4. Validation and results

4.1. Validation procedure

The authors first validate the choice of the autoencoder architecture and compare VAE (Kingma and Welling, 2013), ACAI (Berthelot et al., 2018) and this paper's model on the task of interpolating rotated MNIST digits.

Once the architecture choice is validated, the authors compare their method's performance to cubic B-spline interpolation on the 3 chosen datasets. For each dataset, training, validation and test subsets are created then the model was trained on randomly sampled patches from the training set. The test set was converted to low through-plane resolution by excluding one of every two adjacent slices and using it as ground truth. The model was then quantitatively evaluated using standard similarity metrics, namely Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR) and Visual Information Fidelity (VIF) (Sheikh et al., 2005). Qualitative performance was also evaluated by visually assessing the smoothness and anatomical plausibility of the interpolations. For all datasets, the paper's model was compared to cubic B-spline interpolation, and also with two other supervised methods on the second dataset (3D CNN (Pham et al., 2017), GAN (Pham et al., 2019)) and two unsupervised methods on the third dataset (Fourier Burst Accumulation (Jog et al., 2016), self-supervised CNN (Zhao et al., 2018)). An ablation study was also performed where the synthesis loss is removed.

4.1.1. Shortcomings

In medical imaging, error metrics are usually not enough to prove that a method is medically valuable and we appreciate that the paper also provides qualitative assessment of the interpolations. However it would have been better to blindly provide true and synthetic images to a radiologist and measure their ability to distinguish between the images, although we understand that this can be costly. Another possibility for better quantitative validation would be to compare lesion contrast between original and synthetic images, or use the synthetic images in a tumour classifier and compare its accuracy with a classifier using original images. This would have allowed for a better assessment of whether medically valuable information is well-preserved in the generated images.

4.2. Main results

The qualitative results of first experiment (MNIST rotation interpolation) confirms that the CAE architecture performs better than VAE and ACAI on this task, since the digits are successfully rotated with less cross-fade and blur artifacts.

The paper then provides a thorough analysis of the model's quantitative and qualitative performance on the 3 MRI datasets. To summarise, we observe that this method outperforms cubic B-spline interpolation on all 3 datasets quantitatively (statistically significant higher similarity metrics with the ground-truth images) and qualitatively (smoother interpolation with less cross-fade and aliasing artifacts). A few example synthetic slices can be seen in appendix A, Figure 1 and confirm that this method provides visually smooth and realistic interpolations.

The proposed method also provides similar or better performance than the 4 other methods compared on the second and third datasets, although those methods are either supervised to harder to train. Finally in the ablation study, the authors observed a decrease in the quality of the interpolation when the synthesis loss in not used, which validates the choice of this loss.

5. Conclusions

5.1. Main conclusions of the paper

This paper exploits the latent space encodings produced by autoencoders to perform slice interpolation of MRIs and proves that such encodings coupled with an appropriate loss function can successfully capture semantic information of the images and provide smooth interpolations that outperform standard unsupervised methods and perform on par with supervised methods which require high-resolution training data and more computing power.

Both visual and quantitative evaluation support the paper's claims, although sometimes the generated images are slightly blurry but overall they approximate the ground-truth images better than B-spline. No unsubstantiated claims were identified in this paper.

5.2. Limitations of the study

Although this method significantly reduces cross-fade artifacts, the authors acknowledge that they are still present and more work will need to be done to better address them. They also recognise that this method shows decreased performance when the slices are misaligned, which is a common problem for MRI interpolation methods. The authors aim to address this issue in future work by adding more interpolation endpoints.

In addition, the authors use a simple autoencoder with no fancy preprocessing or postprocessing, which is both a strength of this paper (it doesn't need complex tricks to outperform other methods) and a weakness (its performance could be improved with further tweaking of the architecture or postprocessing). For example, it could have been interesting to test transposed convolution or bicubic upsampling instead of nearest neighbour upsampling for the decoder. The ablation study was also limited (only removing the synthesis loss) and didn't compare different numbers of layers or blocks in the autoencoder. But we understand that the authors chose to leave such optimizations for future research.

References

David Berthelot, Colin Raffel, Aurko Roy, and Ian Goodfellow. Understanding and improving interpolation in autoencoders via an adversarial regularizer, 2018. URL https://arxiv.org/abs/1807.07543.

KK Bhatia, AN Price, W Shi, JV Hajnal, and D Rueckert. Super-resolution reconstruction of cardiac mri using coupled dictionary learning. pages 947–950. IEEE, 2014. URL http://gateway.webofknowledge.com/gateway/Gateway.cgi?GWVersion=2&SrcApp=PARTNER_APP&SrcAuth=LinksAMR&KeyUT=WOS:000392750900235&DestLinkType=FullRecord&DestApp=ALL_WOS&UsrCustomerID=1ba7043ffcc86c417c072aa74d649202.

Amod Jog, Aaron Carass, and Jerry L. Prince. Self super-resolution for magnetic resonance images. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, pages 553–560, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46726-9.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013. URL https://arxiv.org/abs/1312.6114.

Michael Lustig, David Donoho, and John M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. doi: https://doi.org/10.1002/mrm.21391. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.21391.

Jonathan Masci, Ueli Meier, Dan Ciresan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. pages 52–59, 06 2011. ISBN 978-3-642-21734-0. doi: 10.1007/978-3-642-21735-7_7.

Evan M. Masutani, Naeim Bahrami, and Albert Hsiao. Deep learning single-frame and multiframe super-resolution for cardiac mri. *Radiology*, 295(3):552–561, 2020. doi: 10.1148/radiol.2020192173. URL https://doi.org/10.1148/radiol.2020192173. PMID: 32286192.

Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. Brain mri super-resolution using deep 3d convolutional networks. pages 197–200, 04 2017. doi: 10.1109/ISBI.2017.7950500.

Chi-Hieu Pham, Carlos Tor-Díez, Hélène Meunier, Nathalie Bednarek, Ronan Fablet, Nicolas Passat, and François Rousseau. Simultaneous super-resolution and segmentation using a generative adversarial network: Application to neonatal brain mri. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 991–994, 2019.

François Rousseau, Orit Glenn, Bistra Iordanova, Claudia Rodriguez-Carranza, Daniel Vigneron, James Barkovich, and Colin Studholme. Registration-based approach for reconstruction of high-resolution in utero fetal mr brain images. *Academic radiology*, 13:1072–81, 10 2006. doi: 10.1016/j.acra.2006.05.003.

H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005. doi: 10.1109/TIP.2005.859389.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CoRR*, abs/1801.03924, 2018. URL http://arxiv.org/abs/1801.03924.

Can Zhao, Aaron Carass, Blake E. Dewey, and Jerry L. Prince. Self super-resolution for magnetic resonance images using deep networks, 2018. URL https://arxiv.org/abs/1802.09431.

Appendix A.

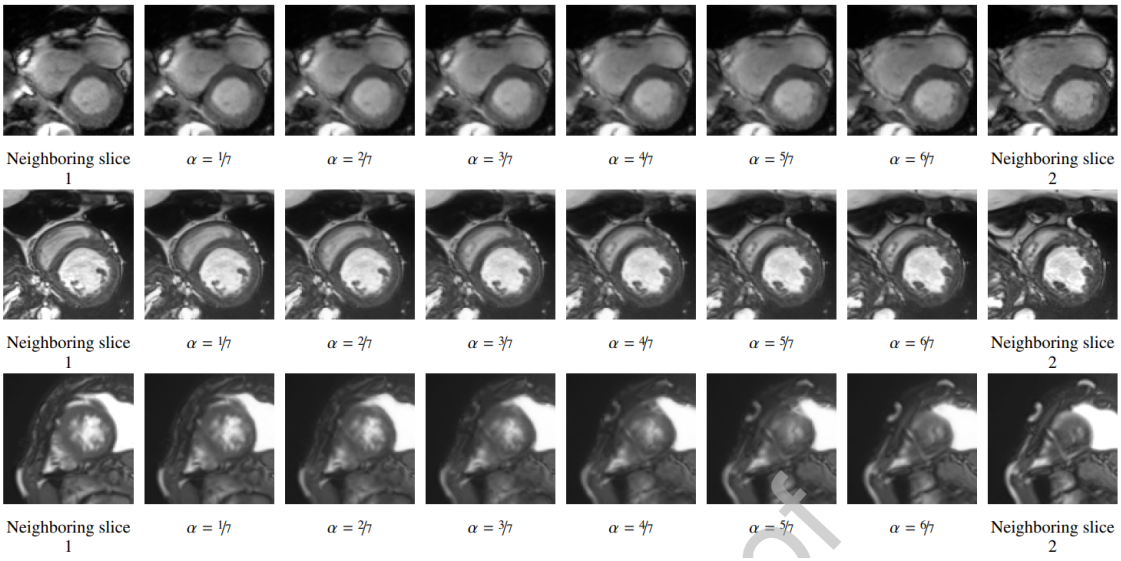

Figure 1: Examples of generated interpolation slices