4050

Voxel-wise Temporal Attention Network and Simulation-Driven Dynamic MRI Sequence Reconstruction

Shouchang Guo[1], Jeffrey A. Fessler[1], and Douglas C. Noll[2]
[1]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United States, [2]Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States

Synopsis

Inspired by two open questions of dynamic MRI reconstruction, we propose a novel voxel-wise attention network for temporal modeling for the undersampled reconstruction. The voxel-wise design of the network enables voxel-wise training, and we further propose a two-stage training scheme that pretrains the network with voxel-wise simulated data when dynamics are easy to obtain with physical models. With a factor of 12 undersampling, our proposed model outperforms other reconstructions with higher PSNR and better fMRI performance.

Introduction

Deep learning-based approaches have been successful for structural MRI undersampled reconstruction [1,2]. However, there are fewer works on learning-based dynamic MRI reconstruction [3,4] with two main open questions: 1) what would be a good learning-based approach for temporal or spatial-temporal signal modeling, 2) for dynamic MRI sequence of images, how to get enough training data for the learning schemes that are data hungry? Inspired by these two questions, we propose a voxel-wise attention network based on the emerging attention mechanism [5, 6] for temporal modeling, together with a matched transfer learning approach to handling the problem of limited amounts of training data.

Our work has three novel contributions: 1) incorporate an attention mechanism for temporal learning and mapping, 2) propose a voxel-wise network architecture based on attention and Transformers for spatial-temporal undersampled reconstruction, 3) propose a two-stage learning scheme that pretrains the network with voxel-wise simulated data, and then fine-tunes with human temporal data for dynamic MRI.

Proposed Model

(1) Attention Mechanism

De-aliasing of a dynamic sequence with undersampling artifacts can be viewed as mapping a temporal sequence with aliasing to a sequence without aliasing. The self-attention mechanism [5, 6] maps a sequence of input vectors to another sequence of output vectors with a weighted combination of all the input vectors for each of the output vectors. It can be expressed as:

$$\text{Attention}(X) = \text{softmax}\left(\frac{XW_Q(XW_K)^T}{\sqrt{d}}\right)XW_V \in \mathbb{R}^{t\times d},$$

where $X \in \mathbb{R}^{t\times d}$ is an input sequence of time dimesion t and feature dimension d. W_Q, W_K, W_V are learned matrices. We use the attention mechanism as a key component for the dynamic sequence reconstruction.

(2) Proposed Voxel-Wise Attention Network

We propose a voxel-wise attention network with Transformers as building blocks. The voxel-wise attention network is composed of three components: 1) an encoder that brings the input time-series to the feature domain, 2) two consecutive Transformer blocks [6, 7] that consist of attention, feed-forward operations, and residual connections, and maps the sequence of temporal features to another sequence of temporal features, 3) a decoder that brings the transformer sequence to the image domain. We used 1×1 convolutions in both the encoder and decoder to ensure the voxel-wise operations of the network.

(3) Two-Stage Training and Data Simulation

The overall framework with an attention network and data consistency is presented in Fig. 1. We performed two-stage training to handle the problem with limited human data for learning. We pretrain the attention network with voxel-wise simulated temporal sequences (which could be easier to simulate than spatial-temporal sequences). After pretraining, we fine-tune the attention network together with data consistency in an end-to-end fashion. For simulated data, we generated the ground truth sequence [8] using Bloch simulation with varying physics parameters, and the inputs for the network are complex Gaussian noise corrupted sequences.

We compared our proposed approach to 3D U-Net [9] that takes spatial-temporal images as 3D volumes for processing. Because 3D U-Net is a spatial-temporal network, we cannot easily pretrain the network with simulated data.

Experimental Methods

The human data were acquired with OSSI sequence [10]. We formed a human data training set with 10 oscillatory temporal images for each subject and 22 subjects in total. The k-space data are multi-coil and undersampled with spiral trajectories with an undersampling factor of 12. We used data-shared initializations as inputs for the network. The simulated data contains 8,662,000 voxel-wise time courses of dimension 10×1. We pretrained the network with simulated data for 60 epochs and fine-tuned the network with human data patches for 60 epochs.

In the testing stage, we reconstructed every 10 dynamic images of the OSSI fMRI data using the proposed network, and L2-combined every reconstructed 10 images to get fMRI images for evaluation. The functional task was a left/right reversing-checkerboard visual stimulus.

Results

Fig. 2 presents attention map visualization for simulated and human temporal sequence mapping. Each sample of the output sequence is formed based on a weighted combination of all the samples in the input sequence, and the weights are given by the rows of the attention map.

For reconstruction, Fig. 3 shows that the proposed method leads to less structure in the difference maps than other reconstruction methods such as 3D U-Net. Every 10 reconstructed images are combined with L2-norm for fMRI. Fig. 4 provides functional maps for the reconstructions. The proposed model results in fewer false positives and cleaner time courses compared to the fully sampled data. Fig. 5 summarizes quantitative evaluations of the reconstruction and functional performance. The proposed model improves the PSNR by about 2 dB and presents lower NRMSE. The proposed method also provides the largest area under the ROC curve.

Conclusions

We propose a novel voxel-wise attention network for dynamic MRI temporal modeling. The voxel-wise network design enables pretraining with voxel-wise simulated data that might be easier to obtain than spatial-temporal data, and resolves the training data limitation for dynamic imaging. Our proposed model reconstructs dynamic MRI images with a factor of 12 undersampling, provides high-quality reconstruction and functional maps, and without spatial smoothing. The proposed voxel-wise, attention-based model can potentially be used for MR fingering reconstruction and other dynamic reconstruction applications.

Acknowledgements

We wish to acknowledge the support of NIH Grant U01EB026977.

References

[1] Hammernik, Kerstin, et al. "Learning a variational network for reconstruction of accelerated MRI data." Magnetic resonance in medicine 79.6 (2018): 3055-3071.

[2] Aggarwal, Hemant K., Merry P. Mani, and Mathews Jacob. "MoDL: Model-based deep learning architecture for inverse problems." IEEE transactions on medical imaging 38.2 (2018): 394-405.

[3] Schlemper, Jo, et al. "A deep cascade of convolutional neural networks for dynamic MR image reconstruction." IEEE transactions on Medical Imaging 37.2 (2017): 491-503.

[4] Qin, Chen, et al. "Convolutional recurrent neural networks for dynamic MR image reconstruction." IEEE transactions on medical imaging 38.1 (2018): 280-290.

[5] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

[6] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

[7] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[8] Guo, Shouchang, Jeffrey A. Fessler, and Douglas C. Noll. "Manifold Model for High-Resolution fMRI Joint Reconstruction and Dynamic Quantification." arXiv preprint arXiv:2104.08395 (2021).

[9] Çiçek, Özgün, et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016.

[10] Guo, Shouchang, and Douglas C. Noll. "Oscillating steady-state imaging (OSSI): A novel method for functional MRI." Magnetic resonance in medicine 84.2 (2020): 698-712.

Figures

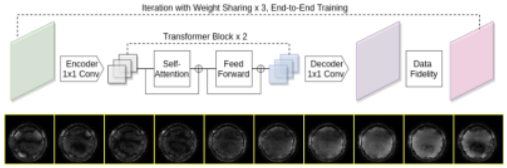

Fig. 1. Our proposed voxel-wise temporal attention network architecture and the dynamic OSSI MRI images (with temporal dimension = 10) to be reconstructed. The data fidelity contains 2 iterations of CG-SENSE for multi-coil NUFFT reconstruction. The main part of the network (encoder-Transformer-decoder) can take voxel-wise simulations or spatial images/patches from human data as inputs.

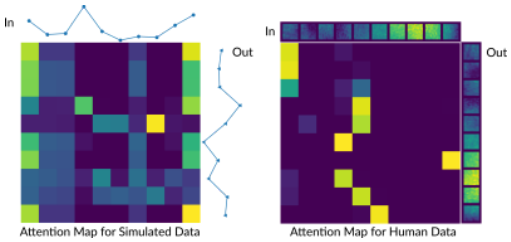

Fig. 2. Attention map visualization at the testing stage for voxel-wise simulation data (left) and human data patch mapping (right). In the attention mechanism, each output value in a 10x1 sequence is generated with a weighted combination of all the values in the input sequence, and the learned weights are given by each row of the 10x10 attention maps for each output value. The figure presents absolute values of the complex input/output for illustration while the proposed network inputs real and imaginary parts and uses deep representations from the encoder for attention calculation.

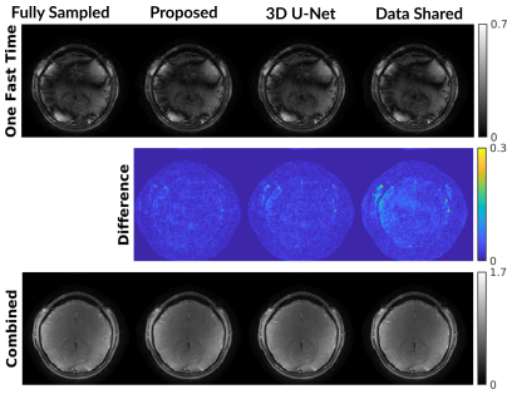

Fig. 3. The proposed voxel-wise model presents less residual in the difference maps than spatial-temporal reconstruction using 3D U-Net.

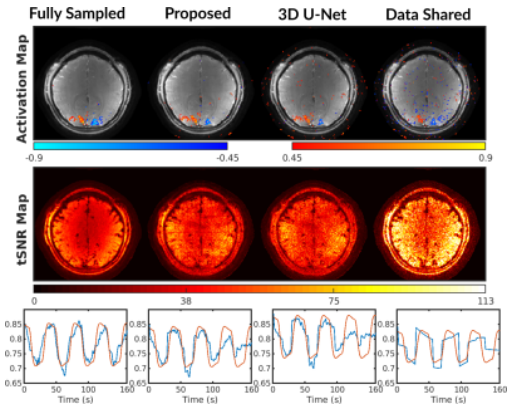

Fig. 4. The proposed approach results in fewer false positives in the activation map, less noisy temporal SNR map, and a time course more similar to the ground truth.

Reconstruction	Proposed	3D U-Net	Data Shared
NRMSE	0.13	0.15	0.16
PSNR	34.4	33.8	32.5

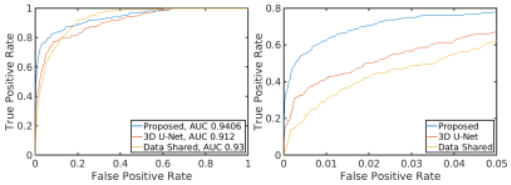

Fig. 5. The table presents NRMSE and PSNR values for the dynamic undersampled reconstructions. Both quantitive values for reconstruction and the ROC curves for fMRI demonstrate that the proposed model outperforms other reconstructions.