Learning Apparent Diffusion Coefficient Maps from Undersampled Radial k-Space Diffusion-Weighted MRI in Mice using a Deep CNN-Transformer Model in Conjunction with a Monoexponential Model

Yuemeng Li¹,², Hee Kwon Song², Miguel Romanello Joaquim², Stephen Pickup², Rong Zhou², Yong Fan¹,²*

¹Center for Biomedical Image Computing and Analytics (CBICA), Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
²Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA

*Corresponding author:
Name Yong Fan
Department Department of Radiology
Institute University of Pennsylvania
E-mail Yong.Fan@pennmedicine.upenn.edu

Approximate word count: 180 (Abstract) 5000 (body)
Abstract:

**Purpose:** To accelerate radially sampled diffusion weighted spin-echo (Rad-DW-SE) acquisition method for generating high quality of apparent diffusion coefficient (ADC) maps.

**Methods:** A deep learning method was developed to generate accurate ADC map reconstruction from undersampled DWI data acquired with the Rad-DW-SE method. The deep learning method integrates convolutional neural networks (CNNs) with vison transformers to generate high quality ADC maps from undersampled DWI data, regularized by a monoexponential ADC model fitting term. A model was trained on DWI data of 147 mice and evaluated on DWI data of 36 mice, with undersampling rates of 4x and 8x.

**Results:** Ablation studies and experimental results have demonstrated that the proposed deep learning model can generate high quality ADC maps from undersampled DWI data, better than alternative deep learning methods under comparison, with their performance quantified on different levels of images, tumors, kidneys, and muscles.

**Conclusions:** The deep learning method with integrated CNNs and transformers provides an effective means to accurately compute ADC maps from undersampled DWI data acquired with the Rad-DW-SE method.

**Keywords:** Parametric Estimation, Self-Attention, Monoexponential Model, Diffusion weighted MRI, Apparent diffusion coefficient, Convolutional Neural Network.
1 Introduction

Diffusion weighted MRI (DWI) provides quantitative metrics related to the translational mobility of water hindered by microstructures present in biological tissues (1,2). Apparent diffusion coefficient (ADC) of water derived from DW images at multiple b-values has been employed extensively as a biomarker in neurological and oncological applications (3-6). Since DWI pulse sequences typically collecting DW images at multiple b-values are sensitized to motion on the micrometer scale, macroscopic (millimeter) scale movement of tissue/organ due to respiratory motion can introduce artefacts, resulting in errors in quantitative measurement of ADC in the affected tissue. In clinical DWI, respiratory motion could be mitigated by employing single-shot echo planar imaging (EPI) with parallel acquisition (7). In DWI of mice, however, due to higher respiration rate and magnetic susceptibility effects increasing with magnetic field strength, EPI-based DWI performed on preclinical MRI instruments leads to greater levels of distortions and artifacts (8). By leveraging intrinsic motion-insensitive property of radial k-space sampling, previous studies have shown that radially sampled diffusion weighted spin-echo (Rad-DW-SE) acquisition method effectively suppresses respiratory motion artifacts in DW-MR images of mouse abdomen over a wide range of b-values (8,9). However, compared to the single-shot EPI acquisition, the acquisition time of Rad-DW-SE is substantially longer. An effective means to shorten the Rad-DW-SE scanning time is to acquire under-sampled k-space data. However, undersampling of Rad-DW-SE k-space data degrades the image quality (e.g., signal-to-noise ratio, SNR) dramatically, especially at higher b-values due to lower signal-to-noise ratios, and subsequently degrades the derived ADC maps, as illustrated in Figure. 1.

Two approaches can be adopted to generate high quality ADC maps from under-sampled DW images: 1) generating high quality DW images followed by fitting a monoexponential model (10); or 2) directly generating ADC maps. High quality DW images can be generated using deep learning methods that have achieved promising performance in MR image reconstruction from undersampled k-space data in k-space domain (11,12), image domain (13-16), or both (17-19). However, the performance of such an indirect method is hinged on the quality of the reconstructed DW images at different b-values with varied signal-to-noise ratios. On the other hand, directly generating high quality ADC maps from under-sampled ADC maps can be implemented using a deep learning model under a supervised learning setting. However, such an approach does not utilize information provided by the under-sampled DW images.

In this study, we develop a deep learning model, referred to as DeepADC-Net, to generate high quality ADC maps from radially undersampled DWI data, in conjunction with a monoexponential model that estimates the ADC maps from the DW images. Our deep learning model takes the undersampled DW images of multiple b-values and derived ADC map as a multi-channel input to generate high quality ADC maps. The deep learning model is trained to optimize two complementary loss functions 1) the
difference between the generated ADC maps learned by deep learning model and their counterparts
derived from fully-sampled DW images using a monoexponential model, and 2) the difference between
fully-sampled DW images and the estimated DW images from the learned ADC maps (10). Different from
the existing deep learning based MR image reconstruction methods that are typically built on
convolutional Neural Networks (CNNs), our deep learning method is an integration of CNNs with vision
transformers (20) that has shown great potential to learn the global context information as a self-attention
module in conjunction with CNNs for feature extractions (21). Extensive ablation studies and
experimental results have demonstrated that the monoexponential model and the integration of CNNs
with vision transformers could enhance deep learning to generate high quality ADC maps from
undersampled data. To the best of our knowledge, this is the first time that a deep learning network has
been applied to accelerate radially sampled DWI for quantitative ADC estimation.

2 Methods

2.1 Datasets

All animal handling protocols were reviewed and approved by our institute’s IACUC. All Genetically
engineered mouse model of pancreatic ductal adenocarcinoma was used and DWI data were acquired
on a Bruker 9.4T animal scanner using a Rad-DW-SE acquisition technique (8). Details of mouse
preparation, anesthesia, and monitoring of vital signs during MRI exam are found in (22). Rad-DW-SE
data were acquired using 2D contiguous slices spanning the abdominal cavity with the following settings:
FOV=32 × 32 mm², 96 readout points, 403 views, thickness=1.5 mm, TR=750 msec, TE=28.7 msec, b-
values= 24.25, 536.86, 1072.62, 1482.07, 2144.69 s/mm², total acquisition time=25 min. The number of
slices in each Rad-DW-SE animal scans were varied: 8, 16 or 19 largely based on tumor size.

Based on the fully sampled Rad-DW-SE data, we evaluated our proposed model with two different
undersampling factors of 4x and 8x respectively. The undersampled 4x DWI data were generated by
sampling one out of every four radial views in k-space, resulting in a total of 100 views and an acceleration
factor of 4, whereas the undersampled 8x DWI data were generated by sampling one out of every eight
radial views in k-space, resulting in a total of 51 views and an acceleration factor of 8. DW images were
reconstructed from both the fully sampled and undersampled k-space data with a matrix size of
96 × 96 and 32 x 32 mm² FOV using re-gridding implementation in Python. ADC maps were computed from both
the undersampled and the fully sampled DW images by Least Squares Fitting the monoexponential
model of Equation (1). The ADC maps derived from the fully sampled DW images were used as ground
truth, and values were excluded if outside of [0, 0.0032] mm²/sec, since 0.0032 corresponds to ADC of
freely diffusing water. We split the entire dataset into a training subset with 147 animal scans consisting of a total of 2255 slices, and a testing subset with 36 animal scans with a total of 557 slices.

2.2 DeepADC-Net

DeepADC-Net is constructed to generate high quality ADC maps from undersampled DWI data collected with the Rad-DW-SE sequence, as schematically illustrated in Figure 2. The input to the deep learning model includes the undersampled DW images and their corresponding ADC maps. The model consists of two parts in the training setting: 1) generating an ADC map and its corresponding DWI scan at a b-value of 0, denoted as $S_0$, from undersampled DW images of different b-values and their corresponding ADC map, computed by fitting a monoexponential model (10), as a multi-channel input, illustrated in Figure 2 (a) and (b); 2) estimating DW images from the generated ADC map and $S_0$ with the monoexponential model, illustrated in Figure 2 (c). In the inference setting, deep learning model is applied to undersampled DW images and their corresponding ADC maps to generate high quality ADC maps, as shown in Figure 2 (a) and (b).

2.2.1 Problem formulation

Given Rad-DW-SE scans collected at $n$ different b-values, an ADC map can be computed from the DW images by fitting a monoexponential model $\mathcal{M}$:

$$S_i = \mathcal{M}(ADC, S_0) = S_0 e^{-b_i \times ADC}, \quad i = 1, \ldots, n,$$

where $S_i$ is signal intensity of a DWI scan at a b-value of $b_i$, $i = 1, \ldots, n$. $ADC$ and $S_0$ are the ADC value and signal intensity of the DWI scan at a b-value of 0, respectively. Based on the monoexponential model, the ADC map can be calculated from DW images collected with at least two $b$-values:

$$ADC_{\text{fit}} = -\frac{\ln(S_j) - \ln(S_i)}{b_j - b_i},$$

where $b_i$ and $b_j$ are two different b-values, and $S_i$ and $S_j$ are their corresponding DWI images. To make the ADC estimation robust, DW images are typically collected at 3 or more b-values, and a Least-Squares-Fitting algorithm is then adopted to estimate the ADC values.

Given undersampled DW images, we aim to optimize the DeepADC-Net to generate a high-quality ADC map close to those estimated from their corresponding full-sampled DW images:

$$\max_{\theta} \text{Similarity}(ADC_{\text{full}}, ADC_{\text{cnn}} = F_{\text{CNN}}(S_i, ADC_{\text{us}}|\theta)),$$

where $ADC_{\text{cnn}} = F_{\text{CNN}}(S_i, ADC_{\text{us}}|\theta)$ is a deep learning model with parameter $\theta$, its input consists of the undersampled DW images $S_i, i = 1, \ldots, n$ and their corresponding ADC map $ADC_{\text{us}}$ computed by fitting the monoexponential model, $ADC_{\text{full}}$ is the fully sampled ADC map computed using the monoexponential model.
model from their corresponding fully sampled DW images, and \( \text{Similarity}(\cdot, \cdot) \) is a similarity measure between two ADC maps. Moreover, multiple regularization terms are adopted to optimize the deep learning model, including one derived from the monoexponential model.

2.2.2 DeepADC Network architecture

**Encoder-Decoder architecture:** High quality ADC maps are generated using an Encoder-Decoder network that is a U-Net with 5 densely connected blocks for both the encoder and decoder (23). This network takes both the undersampled DW images \( \text{DWI}_{\text{us}} \in \mathbb{R}^{H \times W \times n} \) at \( n \) b-values and their corresponding undersampled ADC map \( \text{ADC}_{\text{us}} \in \mathbb{R}^{H \times W \times 1} \) as a multi-channel input, where \( H \) and \( W \) are spatial size of the input. The last Decoder branch has two paralleled output heads \( D_{\text{adc}} \) and \( D_{S_0} \) to generate two outputs: a high quality ADC map \( \overline{\text{ADC}} \in \mathbb{R}^{H \times W \times 1} \) and its corresponding DWI scan at a b-value of 0, denoted by \( \overline{S_0} \in \mathbb{R}^{H \times W \times 1} \), respectively.

To generate ADC values within a physiologically plausible range, we adopt a scaled sigmoid activation function, instead of commonly used activation functions, in the decoder’s output head \( D_{\text{adc}} \), formulated as:

\[
\overline{\text{ADC}} = \text{ADC}_{\text{min}} + \frac{1}{1 + e^{-X_{\text{adc}}}} \times (\text{ADC}_{\text{max}} - \text{ADC}_{\text{min}}),
\]  

where \( \text{ADC}_{\text{min}} \) and \( \text{ADC}_{\text{max}} \) are the lower and upper boundaries of ADC values respectively, and \( X_{\text{adc}} \in \mathbb{R}^{H \times W \times 1} \) is the output of \( D_{\text{adc}} \) (i.e., the one before the activation layer).

According to the monoexponential model specified in Equation (1), the signal intensities of DW images at different b-values are positive, decreasing as the b-value increases. Therefore, the output of \( D_{S_0} \) should be equal to or larger than its corresponding DWI scan collated at the lowest b-value, denoted by \( S_1 \). Accordingly, the output of \( D_{S_0} \) is formulated as:

\[
\overline{S_0} = (1 + \text{Max}(0, X_{S_0})) \times S_1,
\]  

where \( S_1 \) is the DWI scan collated at the lowest b-value, and \( X_{S_0} \in \mathbb{R}^{H \times W \times 1} \) is a feature map generated by the decoder’s output head \( D_{S_0} \).

Given \( \overline{\text{ADC}} \) and \( \overline{S_0} \), DW images at different b-values can be calculated to regularize the output ADC maps \( \overline{\text{ADC}} \) by encouraging the generated DW images \( \tilde{S} = [\tilde{S}_1, \ldots, \tilde{S}_n] \in \mathbb{R}^{H \times W \times n} \) close to their corresponding fully sampled DW images \( S = [S_1, \ldots, S_n] \in \mathbb{R}^{H \times W \times n} \). According to Equation (1), DW images \( \tilde{S} \) can be computed with the generated \( \overline{\text{ADC}} \in \mathbb{R}^{H \times W \times 1} \) and \( \overline{S_0} \in \mathbb{R}^{H \times W \times 1} \) by the monoexponential model, referred to as \( \mathcal{M}(\text{ADC}, S_0) \), at \( n \) b-values used for collecting the DW images.

**Bottleneck Self Attention:** Let \( x \in \mathbb{R}^{C_{\text{in}} \times H \times W} \) be the input feature map, \( q = W_q x \) be the queries, \( k = \)
\( W_kx \) be the keys, and \( v = W_vx \) be the values, the output \( y \in \mathbb{R}^{C_{out} \times H \times W} \) from self-attention layer can be computed as:

\[
y_{ij} = \sum_{n=1}^{H} \sum_{w=1}^{W} \text{softmax}(q_{ij}^n k_{nw}) v_{hw},
\]

where \( i \in \{1, ..., H\} \) and \( j \in \{1, ..., W\} \) represent the arbitrary location, and \( W_q \in \mathbb{R}^{C_{in} \times C_{out}}, W_k \in \mathbb{R}^{C_{in} \times C_{out}}, W_v \in \mathbb{R}^{C_{in} \times C_{out}} \) are the learnable attention weights. To make the self-attention mechanism become sensitive to the positional information, the learnable position encoding is incorporated to the self-attention layer. The Multi-Head self-attention module is applied by taking different query, key and value matrices to enable the attention layers focus on different part of the input feature maps.

**Loss functions:** Multiple loss functions are adopted to optimize the network for generating high quality ADC maps, including:

\[
L_{\text{adc}} = \frac{1}{M} \sum_{j=1}^{M} \| ADC_j - \widehat{ADC}_j \|_1, L_{S_0} = \frac{1}{M} \sum_{j=1}^{M} \| S_{0j} - \widehat{S}_{0j} \|_1, L_{\text{dwi}} = \frac{1}{M} \sum_{j=1}^{M} \| S_j - \widehat{S}_j \|_1,
\]

where \( \widehat{ADC}_j \) is the generated ADC map and \( \widehat{S}_{0j} \) is DW images at b-value of 0, \( \widehat{S}_j \) are the DW images computed from generated \( \widehat{ADC}_j \) and \( \widehat{S}_{0j} \) using \( \mathcal{M}(\widehat{ADC}_j, \widehat{S}_{0j}) \) with \( n \) b-values according to equation (1), and \( ADC_j, S_{0j} \) and \( S_j \) are their counterparts of the fully sampled data. The overall loss function is:

\[
L_{\text{total}} = \alpha L_{\text{adc}} + \beta L_{S_0} + \gamma L_{\text{dwi}},
\]

where \( \alpha, \beta, \) and \( \gamma \) are regularization parameters. We set \( \alpha = 1 \) and \( \beta = \gamma = 0.1 \).

### 2.3 Implementation Details and Evaluation Metrics

We performed our experiment on a single NVIDIA TITAN RTX GPU with Pytorch implementation. We utilized the Adam optimizer with a learning rate of \( 1 \times 10^{-5} \), and a weight decay of \( 1 \times 10^{-4} \). We chose the head size of 4 for multi-head self-attention module (21) in our implemented bottleneck transformer. The model was trained in total of 1000 epochs with approximately 2 hours. Based on our dataset, the dynamic range of DW images with 5 b-values are largely varied, and its corresponding ADC maps are in the range of \([0, 0.0032]\), where the maximal value corresponds to ADC of free water at 37°C. In order to normalize the DW images and its corresponding ADC maps into feasible scale to feed into the deep learning model, we therefore clipped the DWI data with a maximum value of 99 percentile, and further normalized the clipped DWI data into a \([0, 1]\) range. The ADC maps were normalized into the range of \([0, 1]\) during training and the predicted ADC maps were scaled back to its original ranges.

To evaluate ADC map reconstruction, we utilized all testing images with background region excluded by masking out the non-tissue regions to reduce the influence of noise during evaluation. We also
evaluated the ADC map reconstruction on the testing scans into total of 31 DW images out of 36 testing scans in specific regions of interests, including tumor, muscle, and kidney. To comprehensively understand the performance of the generated ADC maps, correlation coefficients could be an effective measurement that focuses on the pixel correlations between the reconstructed versus ground truth ADC maps. Therefore, quantitative metrics, including correlation coefficient (CC), structural similarity (SSIM) index, peak signal-to-noise ratio (PSNR), and normalized mean square error (NMSE) were used to compare ADC maps of inference datasets derived from the deep learning models with those computed from the fully sampled imaging data.

2.4 Ablation studies

2.4.1 Ablation study on network inputs

We implemented all our ablations on a DenseU-Net backbone (23) with the same training and inference settings. All our ablations are performed using undersampled DW images with an undersampling factor of 4. As shown in Table 1, the baseline DenseU-Net model takes the input of the undersampled ADC map to generated high-quality ADC maps. We first evaluate if the undersampled DW images at multiple b-values could provide the anatomic information to improve ADC map estimation, where we used undersampled DW images at multiple b-values and their derived ADC maps as a multi-channel input, referred to as DenseU-ADC.

2.4.2 Ablation study on loss functions

We studied the effectiveness of the monoexponential model. Particularly, we implemented a DenseU-Net with network supervision on reconstructing DW images from the generated ADC map and the auxiliary output $S_0$ using the monoexponential model, referred to as DenseU-DWI. In this setting, rather than optimizing the CNN models based on ADC maps, DenseU-DWI was trained by directly calculating image discrepancies between reconstructed DWIs and fully sampled DWIs at multiple b-values in an unsupervised setting.

To evaluate how this monoexponential model improves the model performance as a regularizer, we implemented the DenseU-Net with network supervision on both reconstructed ADC maps, $S_0$ and DWI, referred to as DenseU-BOTH. In this study, the loss function $L_{adc}$ on reconstructed ADC maps is utilized as supervision to generate the high-quality ADC maps, where we leveraged the reconstructed auxiliary output $S_0$ and the calculated DWI using the monoexponential model. Therefore, besides generating high-quality ADC maps, the loss functions $L_{S_0}$ and $L_{	ext{dwI}}$ were functioned as regularization terms to enforce the network reconstructing high-quality DW images from those predicted ADC maps and $S_0$. 
2.4.3 Ablation study on self-attentions

Besides comparing different network inputs and loss functions, we studied how the self-attention module extracting feature attentions in the network could assist to predict ADC maps. The multi-head self-attention (MHSA) module was implemented in the network bottleneck, which contains three densely connected convolutional layers, and the MHSA module was applied after the first and second convolutional layers. The effectiveness of the self-attention module was evaluated by directly comparing the ablation model DenseU-BOTH with the proposed DeepADC-Net.

2.4.4 Ablation study on numbers of filter sizes and convolutional layers

We investigated how the network filter size affects the model performance by implementing the models with the different filter sizes of 64, 128, 196 and 256 for each densely connected blocks in all encoders and decoders. We then fixed the network filter size to test the effectiveness of different numbers of consecutive convolutional layers inside each densely connected block. In those ablations, all models contained the same training setting with the presence of 1) the DW images and ADC maps as multi-channel inputs, 2) supervisions on ADC maps, $S_0$ and DW images ($L_{adc}$, $L_{S_0}$ and $L_{dw}$) during training and 3) the transformer module at network bottleneck.

2.5 Comparison with state-of-the-art Methods

We performed all our experiments using undersampling factors of 4x and 8x. We compared our method, namely DeepADC-Net, with state-of-the-art deep learning method, including: 1) U-Net (24), 2) DenseU-Net (23), 3) FBP-ConvNet (13), and 4) Att-UNet (14). All those methods were implemented with the same network architectures as reported in their corresponding papers to generate high quality ADC maps from the undersampled ADC maps. We utilized same training and inference setting for all models, where the best models are saved based on the best correlation coefficient score from the training dataset, and we evaluated the models’ overall performance in all four criteria on the testing dataset. Specifically, the U-Net model contained encoders and decoders, each with 4 convolutional blocks; the DenseU-Net model contained encoders and decoders, each with 5 densely-connected blocks; the FBP-ConvNet model used the U-Net based architecture with a skip connection between the input and the output; and the Att-UNet model utilized a channel attention mechanism within the U-Net backbone.

We evaluated our proposed network and alternative methods for entire ADC map cross-section estimation. The evaluation on the whole scans could demonstrate the generalization and robustness of the model, but it could also overlook the degraded performance on specific Region-of-Interests (ROIs) while having the high performance on the surrounding tissues. To further demonstrate that the proposed
model could accurately estimate and reconstruct ADC maps in specific regions of interest, we quantitively and qualitatively compared all network performances by limiting to a smaller ROIs, including tumor, muscle, and kidney-only regions. All ROIs were manually labeled.

3 Results

3.1 Comparison with state-of-the-art Methods on the Entire Cross-Section

For 4x undersampled testing dataset as shown in Table 2, the curve fitting of ADC maps from undersampled DW images shows severe degradation compared to fully sampled ADC maps. U-net, DenseU-Net, Att-UNet, and FBP-ConvNet methods could successfully estimate the ADC parametric map but still missed the detailed tissue information and thus yield lower performance. We could observe a large performance improvement between the curve fitting method and CNN based methods. By comparing our proposed DeepADC-Net with the alternative CNN methods, DeepADC-Net obtained better performance than the second-best method with improvement by 3.15%, 0.13%, 1.29 in correlation coefficient, SSIM, and PSNR respectively and a decrease in NMSE by 0.63%.

The ADC map estimation for 8x undersampled testing dataset proved to be a more challenging task due to the additional loss of image information from undersampled scans, where the curve fitting of ADC maps from undersampled DW images showed worse performance for all evaluation criteria. The CNN based methods improved the ADC map estimation and yielded the ADC maps with significantly improved performance. All the alternative methods achieved similar performance in SSIM, PSNR and NMSE. Furthermore, our proposed DeepADC-Net demonstrated superior performance compared to the other methods for all measures.

The visualization of representative cases with best performance for 4x and 8x undersampled dataset is shown in Figure. 3. The demonstration of absolute error map with range displayed up to 75% maximum difference are shown in the second and fifth rows of each figure. The direct curve fitting of ADC maps from undersampled DW images is incapable of estimating tissues near the edges with fine details of the image missing because of low SNR. The CNN based method including U-net, DenseU-Net, Att-UNet, and FBP-ConvNet could successfully reconstruct the complete ADC maps but yielded higher error maps. By comparing the performance of DeepADC-Net with alternative CNN methods, DeepADC-Net generated ADC maps performed best with smaller error and higher similarity to the fully sampled dataset.

Besides demonstrating the case with best performance, Figure. 4 shows the representative case with worst performance of ADC map estimation for 4x and 8x undersampled dataset. For both datasets, the curve fitting of ADC maps from undersampled DW images shows the partial scans with significant missing of the organs and boundary tissues. Alternative CNN methods could reconstruct the ADC maps.
with more smooth estimations and less details in the tissue region. Even for the case with lowest performance, the proposed DeepADC-Net outperformed from the alternative CNN methods, maintaining better issue details.

3.2 Comparison with Competing Methods on Region of Interests

We evaluated the network performance in specific tissue ROIs, including tumor, muscle and kidney, to determine their performance when limited to a smaller region. The networks were trained on the entire cross-section, and were subsequently applied only in the tissue ROIs. For the 4x undersampled testing dataset as shown in Table 3, the curve fitting showed degraded performance on all three organs. By comparing the alternative CNN based methods, Att-UNet achieved best performance in correlation coefficient, whereas the FBP-ConvNet yielded best performance in SSIM, PSNR and NMSE for all three organs. Furthermore, the proposed DeepADC-Net showed best performance in tumor, muscle, and kidney region and outperformed from the second-best method, especially on the criteria of Correlation Coefficient, PSNR and NMSE. Similarly, for 8x undersampled testing dataset as shown in Table 4, alternative CNN based methods yielded better performance compared to curve-fitting method. Our proposed DeepADC-Net outperformed second best method with the increments of 8.19%, 9.32% and 8.78% for tumor, muscle and kidney estimation respectively in correlation coefficient, and showed best performance among all methods in SSIM, PSNR and NMSE by a large margin for all three organs under comparison.

The third and sixth rows from Figure. 3 present the visualization of absolute error map from the best representative cases with range displayed up to 25% maximum difference on tumor for 4x and 8x undersampled dataset. The tissues outside of the tumor regions are masked out. Specifically, for 4x undersampled dataset, the absolute error maps for curve fitting methods yielded similar qualitative results compared to CNN based methods, with U-Net method showing slightly larger errors across the tumor region. For 8x undersampled dataset, we could observe a significant difference between curve fitting method and the CNN based methods, where the curve fitting method showed large errors in the edges of the tumor. For all alternative methods under comparison, U-Net and DenseU-Net showed declined performance on reconstructing the upper left tumor region, whereas Att-UNet and FBP-ConvNet reconstructed tumor regions with less errors. For both 4x and 8x undersampled datasets, the proposed DeepADC-Net presented lowest absolute error difference among all alternative methods, and thus yielded best results for high-quality ADC map estimation on smaller regions.

Moreover, the third and sixth rows from Figure. 4 demonstrate the case with worst performance on tumor regions for 4x and 8x undersampled dataset. Different from the best representative cases displayed in Figure. 3, the undersampled ADC maps generated from curve fitting methods showed large
discrepancies and higher errors in the tumor region, and thus it is harder for CNN based methods to generate accurate tumor region reconstructions from those undersampled ADC maps. Therefore, all alternative CNN based methods yielded similar performance reconstruction results as curve fitting methods for both datasets. Although all alternative CNN methods showed worst performance in this representative case, our proposed DeepADC-Net still outperformed all alternative methods for tumor ADC map estimation with relatively less error.

3.3 Ablation study on network inputs, loss functions and self-attentions

DeepADC-Net utilizes the DenseU-Net backbone, consisting of five densely connected blocks for both encoder and decoder, and the DenseU-Net served as the baseline backbone in all our ablation experiments. Table 5 summarizes ADC map estimation performance obtained by deep learning models with different settings of network inputs, loss functions and self-attentions. The DenseU-ADC network with undersampled DW images at multiple b-values and their derived ADC maps as a multi-channel input showed large improvements comparing to DenseU-Net, which shows the effectiveness of the undersampled DWI as auxiliary input.

We then compared the network performance among the models incorporated with different loss functions but contained the same undersampled multi-channel input. The DenseU-DWI model built upon reconstructing DW images had worse performance than those models with supervisions on ADC map reconstruction. This declined result is reasonable because this model solely focuses on reconstructing DW images with loss function $L_{dwi}$ using the generated ADC maps and $S_0$. Thus, DW images without direct supervision on generated ADC maps yielded worse performance compared to those with supervisions on generated ADC maps.

The goal of optimizing the DenseU-BOTH model is focuses on reconstructing high-quality ADC maps, as well as minimizing the differences between the generated DW images and fully-sampled DW images, where the high-quality $S_0$ and DW images are the auxiliary output. As shown in Table 5, DenseU-BOTH with the presence of all loss functions improved the model performance of both DenseU-ADC and DenseU-DWI. By directly comparing DenseU-BOTH and DenseU-ADC, the results indicated that the supervisions on reconstructing high-quality $S_0$ and DW images as auxiliary output are beneficial during network training and thus yield superior results.

The proposed DeepADC-Net model was built with the undersampled DWI as auxiliary input, the supervisions on reconstructing high-quality ADC maps, $S_0$ and DW images, and the self-attention module in the network bottleneck. By comparing the performance of DeepADC-Net and DenseU-BOTH, we observed that the bottleneck transformer with multi-head self-attention module showed further improved
the results by effectively extracting feature attentions. Moreover, the proposed DeepADC-Net yielded best performance among all ablations under comparison for high-quality ADC map estimation.

3.4 Ablation study on numbers of filter sizes and convolutional layers

To investigate how the network filter size affects the model performance, we evaluated DeepADC-Net models built with different numbers of filter sizes for each densely connected blocks in all encoders and decoders. As summarized in Table 6 top four rows, filter size of 128 yielded the best performance for Correlation Coefficient; 256 performed best for SSIM; and 64 did best for PSNR and NMSE. We adopted a filter size of 64 for our proposed network due to its computational efficiency and overall competitive performance.

Table 6 bottom two rows summarize the performance of DeepADC-Net models trained with different numbers of consecutive convolutional layers inside each densely connected block. We evaluated the network backbone with 3 and 5 consecutive convolutional layers, and these results indicated that the model with 3 consecutive convolutional layers yielded best results for SSIM, PSNR and NMSE. We therefore adopted convolution layers size of 3 for our proposed network.

4 Discussion

We have proposed a new deep leaning method, referred to as DeepADC-Net, to reconstruct apparent diffusion coefficient maps from undersampled diffusion-weighted MR data. To evaluate our method, we down-sampled 4x and 8x of fully sampled DW images of various b-values and ADC maps calculated from under-sampled DW images as a multi-channel input. Our proposed network contains two complementary loss functions 1) the difference between ADC maps learned by deep learning (DL) model and those computed from fully-sampled DW images ($L_{adc}$) and 2) the difference between fully-sampled DW images and DL-derived DW images, which are computed from the generated $S_0$ and ADC maps ($L_{S_0}$ and $L_{dw_i}$).

The bottleneck transformer with multi-head self-attention module further improved model performance by effectively extracting the feature representations. Quantitative metrics have demonstrated that our method could achieve better ADC map estimations than alternative state-of-the-art DL methods. To the best of our knowledge, this is the first report that utilizes deep learning-based method for quantitative metric estimation from radially undersampled DW-MRI.

We have compared our method with conventional curve-fitting and state-of-the-art DL based methods on both entire cross-section and specific ROIs, including tumor, muscle and kidney. In particular, we directly compared our method with the curve fitting algorithm, U-Net (24), DenseU-Net (23), FBP-ConvNet (13), and Att-UNet (14) under the same training and inference settings. Comparison results
summarized in Tables 2, 3, and 4 demonstrated that DeepADC-Net obtained the best results among all methods under comparison for ADC map reconstruction on both the entire cross-section and specific tissue regions. Figure. 3 and Figure. 4 qualitatively showed the visual differences between DeepADC-Net and alternative methods on whole image cross-section as well as on small regions of interests, with DeepADC-Net generating high-quality ADC maps with minimum errors and less discrepancy from full-sampled ADC maps. We not only demonstrated that the proposed DeepADC-Net could achieve best performance on the case with highest evaluation scores, but also proved that even for the worst case with lowest evaluation scores, DeepADC-Net still outperformed all alternative methods and yielded visually better results.

The objective of the present study was to compute high quality ADC map from undersampled radial k-space DWI data, and thus is different from general image super-resolution. Though many deep learning algorithms have been developed for MRI data reconstruction, our method is a first-of-its-kind for computing parametric maps, i.e., ADC maps. The approach thus is fundamentally different from existing DL methods focusing on reconstruction of MRI images from undersampled k-space data. Our method adopts the monoexponential model to directly compute ADC maps in the deep learning process, facilitating end-to-end learning of the ADC maps with improved quality and efficiency. Since no alternative deep learning method is available for a direct comparison, we evaluated the proposed method through extensive ablation studies. In our ablation studies, we first compared the baseline DenseU-Net model by taking different inputs. The results demonstrated that the undersampled DWI and their derived ADC maps as multi-channel inputs could significantly improve the model performance compared to the model with ADC maps as its input, where the undersampled DWI could still provide crucial anatomical information to facilitate ADC map reconstructions.

Different from conventional methods that solely optimize the differences between DL-generated ADC maps and those computed from fully-sampled DW images, our results showed that minimizing the discrepancy between the DW images reconstructed from DL-the generated ADC maps and fully-sample DW images could regularize the reconstructed ADC maps more effectively to achieve greater accuracy. Furthermore, the proposed DeepADC-Net could specifically extract anatomical representations from the multi-head self-attention module incorporated in the bottleneck and potentially make the network more robust, especially for reconstructing smaller regions of interests. We have also studied how different numbers of filter size and convolutional layers inside each densely connected block have impacted the model performance, and the experimental results indicated that our model could achieve competitive results with minimum numbers of filter size and consecutive convolutional layers.

Moreover, we showed that the DWI regularization helps increase the quality of predicted ADC maps in the evaluation, and in our future work, generating high-quality DWI could be another interesting topic
to study for DWI reconstruction from undersampled DW images by applying ADC loss as regularization.

5 Conclusion

We developed a deep learning method, referred to as DeepADC-Net, to reconstruct apparent diffusion coefficient maps from undersampled diffusion-weighted MR data. In particular, our proposed network takes the undersampled DWI with various b-values and undersampled ADC maps as a multi-channel input. The proposed DeepADC-Net integrating a densely connected Encoder-Decoder architecture with a vision transformer has shown promising results for computing ADC maps from radially diffusion weighted spin-echo (Rad-DW-SE) k-space data. Consequently, this DL model can accelerate Rad-DW-SE acquisition by 4 to 8-fold.

Acknowledgements

This work was supported in part by the National Institutes of Health [grant numbers: MH120811, EB022573, AG066650, and CA253377].
Table 1. Illustration of the compared ablations with (✓) and without (✗) listed implementations.

| Models        | DWI inputs | $L_{adc}$ | $L_{S0}$ | $L_{dwi}$ | Self-Attention |
|---------------|------------|-----------|----------|-----------|----------------|
| DenseU-Net    | ✗          | ✓         | ✗         | ✗         | ✗              |
| DenseU-ADC    | ✓          | ✓         | ✗         | ✗         | ✗              |
| DenseU-DWI    | ✓          | ✗         | ✗         | ✓         | ✗              |
| DenseU-BOTH   | ✓          | ✓         | ✓         | ✓         | ✗              |
| DeepADC-Net   | ✓          | ✓         | ✓         | ✓         | ✓              |
Table 2. Quantitative comparison of ADC maps on DeepADC-Net and alternative state-of-the-art methods for both 4x and 8x undersampled testing datasets. Results are shown as (Mean ± Standard Deviation).

| Models          | Sampling Factor | Correlation ($\times 10^{-2}$) | SSIM ($\times 10^{-2}$) | PSNR | NMSE ($\times 10^{-2}$) |
|-----------------|-----------------|-------------------------------|-------------------------|------|-------------------------|
| Curve fitting   | 4x              | 68.35 ±6.25                   | 96.13 ±1.60             | 14.98 ±1.69 | 13.39 ±3.76            |
| FBPConvNet      |                 | 87.28 ±3.66                   | 99.49 ±0.15             | 21.89 ±1.11 | 2.45 ±0.34             |
| AttUnet         |                 | 87.76 ±3.41                   | 99.45 ±0.08             | 21.73 ±0.84 | 2.61 ±0.38             |
| DenseUnet       |                 | 87.68 ±3.48                   | 99.47 ±0.10             | 21.85 ±0.96 | 2.47 ±0.30             |
| Unet            |                 | 87.41 ±3.57                   | 99.48 ±0.11             | 21.89 ±1.09 | 2.46 ±0.32             |
| DeepADC-Net     |                 | 90.91 ±2.28                   | 99.62 ±0.06             | 23.18 ±0.90 | 1.82 ±0.19             |
| Curve fitting   | 8x              | 46.00 ±7.90                   | 84.86 ±4.44             | 9.60 ±1.53  | 43.15 ±9.34            |
| FBPConvNet      |                 | 76.01 ±5.45                   | 99.02 ±0.18             | 19.35 ±0.92 | 4.41 ±0.63             |
| AttUnet         |                 | 76.76 ±5.49                   | 99.00 ±0.18             | 19.30 ±0.91 | 4.52 ±0.79             |
| DenseUnet       |                 | 76.16 ±5.31                   | 99.03 ±0.19             | 19.40 ±0.92 | 4.37 ±0.64             |
| Unet            |                 | 76.27 ±5.43                   | 99.03 ±0.18             | 19.39 ±0.92 | 4.36 ±0.62             |
| DeepADC-Net     |                 | 85.77 ±3.15                   | 99.37 ±0.09             | 21.06 ±0.75 | 2.97 ±0.49             |
Table 3. Quantitative comparison of ADC maps on DeepADC-Net and alternative state-of-the-art methods for 4x undersampled testing datasets on different ROIs. Results are shown as (Mean ± Standard Deviation).

| Models         | ROIs   | Correlation ($\times 10^{-2}$) | SSIM ($\times 10^{-2}$) | PSNR   | NMSE ($\times 10^{-2}$) |
|----------------|--------|---------------------------------|--------------------------|--------|------------------------|
| Curve fitting  | Tumor  | 67.80 ±8.2                      | 96.69 ±1.42              | 15.16 ±1.85 | 11.85 ±4.11          |
| FBPConvNet     |        | 87.72 ±3.0                      | 99.52 ±0.11              | 21.96 ±1.00 | 2.29 ±0.40           |
| AttUnet        |        | 88.17 ±2.9                      | 99.48 ±0.11              | 21.81 ±0.84 | 2.43 ±0.56           |
| DenseUnet      |        | 87.98 ±2.9                      | 99.49 ±0.09              | 21.84 ±0.86 | 2.35 ±0.39           |
| Unet           |        | 87.76 ±3.0                      | 99.51 ±0.11              | 21.95 ±1.03 | 2.30 ±0.40           |
| DeepADC-Net    |        | 90.32 ±2.4                      | 99.61 ±0.08              | 22.89 ±0.94 | 1.84 ±0.31           |
| Curve fitting  | Muscle | 70.32 ±11.2                     | 96.66 ±2.14              | 15.27 ±2.38 | 11.78 ±5.66          |
| FBPConvNet     |        | 88.09 ±4.7                      | 99.51 ±0.17              | 21.93 ±1.53 | 2.37 ±0.63           |
| AttUnet        |        | 88.54 ±4.7                      | 99.47 ±0.13              | 21.86 ±1.57 | 2.49 ±1.27           |
| DenseUnet      |        | 88.39 ±4.6                      | 99.48 ±0.19              | 21.89 ±1.39 | 2.42 ±0.72           |
| Unet           |        | 88.16 ±4.8                      | 99.49 ±0.20              | 21.89 ±1.63 | 2.41 ±0.74           |
| DeepADC-Net    |        | 90.83 ±3.6                      | 99.62 ±0.11              | 22.95 ±1.30 | 1.86 ±0.43           |
| Curve fitting  | Kidney | 66.05 ±10.7                     | 96.17 ±2.52              | 14.66 ±2.32 | 13.17 ±6.61          |
| FBPConvNet     |        | 86.10 ±5.0                      | 99.50 ±0.15              | 21.71 ±1.37 | 2.38 ±0.58           |
| AttUnet        |        | 86.53 ±4.9                      | 99.45 ±0.28              | 21.54 ±1.46 | 2.56 ±1.25           |
| DenseUnet      |        | 86.40 ±4.8                      | 99.46 ±0.16              | 21.55 ±1.25 | 2.47 ±0.62           |
| Unet           |        | 86.19 ±5.0                      | 99.48 ±0.17              | 21.64 ±1.35 | 2.43 ±0.60           |
| DeepADC-Net    |        | 89.62 ±2.9                      | 99.58 ±0.11              | 22.55 ±1.17 | 1.95 ±0.42           |
Table 4. Quantitative comparison of ADC maps on DeepADC-Net and alternative state-of-the-art methods for 8x undersampled testing datasets on different ROIs. Results are shown as (Mean ± Standard Deviation).

| Models        | ROIs | Correlation ($\times 10^{-2}$) | SSIM ($\times 10^{-2}$) | PSNR | NMSE ($\times 10^{-2}$) |
|---------------|------|---------------------------------|-------------------------|------|--------------------------|
| Curve fitting | Tumor| 41.83 ±12.16                    | 86.53 ±4.39             | 9.74 ±1.44 | 39.28 ±10.83 |
| FBPConvNet    |      | 77.27 ±4.45                     | 99.12 ±0.18             | 19.54 ±0.91 | 3.98 ±0.64   |
| AttUnet       |      | 77.56 ±4.54                     | 99.08 ±0.20             | 19.42 ±0.88 | 4.12 ±0.78   |
| DenseUnet     |      | 76.92 ±4.67                     | 99.11 ±0.19             | 19.53 ±0.92 | 4.00 ±0.72   |
| Unet          |      | 77.04 ±4.85                     | 99.12 ±0.18             | 19.54 ±0.89 | 3.98 ±0.69   |
| DeepADC-Net   |      | **85.75 ±2.89**                 | **99.41 ±0.08**         | **21.16 ±0.70** | **2.74 ±0.46** |
| Curve fitting | Muscle| 42.84 ±10.70                    | 85.21 ±5.31             | 9.42 ±1.58  | 41.83 ±12.08 |
| FBPConvNet    |      | 74.94 ±6.14                     | 99.08 ±0.19             | 19.39 ±1.03 | 4.06 ±0.66   |
| AttUnet       |      | 74.89 ±7.36                     | 99.03 ±0.25             | 19.18 ±1.07 | 4.30 ±0.88   |
| DenseUnet     |      | 74.32 ±6.43                     | 99.06 ±0.22             | 19.29 ±1.02 | 4.16 ±0.75   |
| Unet          |      | 74.63 ±6.26                     | 99.07 ±0.21             | 19.33 ±0.99 | 4.12 ±0.73   |
| DeepADC-Net   |      | **84.17 ±3.59**                 | **99.36 ±0.12**         | **20.81 ±0.77** | **2.92 ±0.47** |
| Curve fitting | Kidney| 44.48 ±10.44                    | 86.31 ±3.90             | 9.78 ±1.41  | 40.26 ±93.44 |
| FBPConvNet    |      | 77.14 ±4.35                     | 99.04 ±0.17             | 19.33 ±0.87 | 4.40 ±0.71   |
| AttUnet       |      | 77.59 ±4.53                     | 99.00 ±0.18             | 19.23 ±0.76 | 4.59 ±1.14   |
| DenseUnet     |      | 77.14 ±4.37                     | 99.04 ±0.16             | 19.34 ±0.84 | 4.36 ±0.72   |
| Unet          |      | 76.97 ±4.44                     | 99.04 ±0.17             | 19.34 ±0.86 | 4.37 ±0.67   |
| DeepADC-Net   |      | **86.37 ±2.79**                 | **99.39 ±0.08**         | **21.10 ±0.58** | **2.92 ±0.70** |
Table 5. Quantitative comparison of ablation studies of ADC maps for 4x undersampled testing datasets. Results are shown as (Mean ± Standard Deviation).

| Models       | Correlation ($\times 10^{-2}$) | SSIM ($\times 10^{-2}$) | PSNR    | NMSE ($\times 10^{-2}$) |
|--------------|-------------------------------|-------------------------|---------|------------------------|
| DenseU-Net   | 87.68 ± 3.48                  | 99.47 ± 0.10            | 21.85 ± 0.96 | 2.47 ± 0.30          |
| DenseU-ADC   | 90.15 ± 2.65                  | 99.59 ± 0.06            | 22.81 ± 0.91 | 1.96 ± 0.21          |
| DenseU-DWI   | 84.37 ± 3.13                  | 99.37 ± 0.10            | 20.80 ± 0.70 | 3.15 ± 0.48          |
| DenseU-BOTH  | 90.63 ± 2.46                  | 99.62 ± 0.07            | 23.05 ± 0.88 | 1.87 ± 0.24          |
| DeepADC-Net  | 90.91 ± 2.28                  | 99.62 ± 0.06            | 23.18 ± 0.90 | 1.82 ± 0.19          |
Table 6. Ablation studies of ADC maps with different filter sizes and different numbers of convolutional layers for 4x undersampled testing datasets. Results are shown as (Mean ± Standard Deviation).

| Convolution Layers | Filter Size | Correlation ($\times 10^{-2}$) | SSIM ($\times 10^{-2}$) | PSNR | NMSE ($\times 10^{-2}$) |
|--------------------|-------------|---------------------------------|------------------------|------|------------------------|
| 3                  | 64          | 90.91 ±2.28                     | 99.62 ±0.06            | 23.18 ±0.90 | 1.82 ±0.19            |
| 3                  | 128         | 90.93 ±2.53                     | 99.62 ±0.07            | 23.08 ±1.02 | 1.85 ±0.25            |
| 3                  | 196         | 90.63 ±2.64                     | 99.60 ±0.07            | 22.96 ±0.89 | 1.93 ±0.37            |
| 3                  | 256         | 90.69 ±2.67                     | 99.63±0.08             | 23.10 ±1.07 | 1.86 ±0.29            |
| 3                  | 64          | 90.91 ±2.28                     | 99.62 ±0.06            | 23.18 ±0.90 | 1.82 ±0.19            |
| 5                  | 64          | 90.97 ±2.58                     | 99.61 ±0.08            | 23.04 ±1.06 | 1.87 ±0.27            |
Figure 1. Diffusion-weighted images and ADC maps from fully and undersampled RAD-DWI-SE scans at different b-values. The undersampled images are obtained by down-sampling the fully sampled data by a factor of four and eight, and the ADC maps are computed from their corresponding DW images by fitting a monoexponential model. Compared with their fully-sampled counterparts, the undersampled DW images are noisy, especially at higher b-values, and the derived ADC maps lose anatomical details at locations with white arrows.
Figure. 2. DeepADC-Net flowchart: a) the input consists of multiple channels, including undersampled DW images and their corresponding ADC maps generated by fitting a monoexponential model; b) a densely connected encoder-decoder backbone that contains a bottleneck transformer with the self-attention; c) the output includes high quality ADC map and $S_0$ where the high quality DW images are reconstructed from those outputs using the monoexponential model.
Figure 3. Visualization of best cases of 4x and 8x undersampled dataset on DeepADC-Net and alternative methods. The first and fourth rows show the ground truth and the generated ADC maps. The second and fifth row shows the absolute error maps with range displayed up to 75% of the maximum difference. The third and sixth rows show the absolute error maps in the tumor region with range displayed up to 25% maximum difference.
Figure 4. Visualization of worst cases of 4x and 8x undersampled dataset on DeepADC-Net and alternative methods. The first and fourth rows show the ground truth and the generated ADC maps. The second and fifth row shows the absolute error maps with range displayed up to 75% of the maximum difference. The third and sixth rows show the absolute error maps in the tumor region with range displayed up to 25% maximum difference.
Reference

1. White NS, McDonald C, Farid N, Kuperman J, Karow D, Schenker-Ahmed NM, Bartsch H, Rakow-Penner R, Holland D, Shabaik A, Bjornerud A, Hope T, Hattangadi-Gluth J, Liss M, Parsons JK, Chen CC, Raman S, Margolis D, Reiter RE, Marks L, Kesari S, Mundt AJ, Kane CJ, Carter BS, Bradley WG, Dale AM. Diffusion-weighted imaging in cancer: physical foundations and applications of restriction spectrum imaging. Cancer Res 2014;74(17):4638-4652.

2. Le Bihan D. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. Radiology 2013;268(2):318-322.

3. Guo Y, Cai YQ, Cai ZL, Gao YG, An NY, Ma L, Mahankali S, Gao JH. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 2002;16(2):172-178.

4. Marini C, Iacconi C, Giannelli M, Cilotti A, Moretti M, Bartolozzi C. Quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesion. European radiology 2007;17(10):2646-2655.

5. Sinha S, Lucas-Quesada FA, Sinha U, DeBruhl N, Bassett LW. In vivo diffusion-weighted MRI of the breast: potential for lesion characterization. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 2002;15(6):693-704.

6. Tong D, Yenari M, Albers G, O'brien M, Marks M, Moseley M. Correlation of perfusion-and diffusion-weighted MRI with NIHSS score in acute (< 6.5 hour) ischemic stroke. Neurology 1998;50(4):864-869.

7. Bammer R, Keeling SL, Augustin M, Pruessmann KP, Wolf R, Stollberger R, Hartung HP, Fazekas F. Improved diffusion-weighted single-shot echo-planar imaging (EPI) in stroke using sensitivity encoding (SENSE). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 2001;46(3):548-554.

8. Cao J, Song HK, Yang H, Castillo V, Chen J, Clendenin C, Rosen M, Zhou R, Pickup S. Respiratory Motion Mitigation and Repeatability of Two Diffusion-Weighted MRI Methods Applied to a Murine Model of Spontaneous Pancreatic Cancer. Tomography 2021;7(1):66-79.

9. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. Journal of Magnetic Resonance Imaging 2015;42(4):887-901.
10. Neil JJ, Bretthorst GL. On the use of bayesian probability theory for analysis of exponential decay data: An example taken from intravoxel incoherent motion experiments. Magnetic resonance in medicine 1993;29(5):642-647.

11. Akçakaya M, Moeller S, Weingärtner S, Uğurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. Magnetic resonance in medicine 2019;81(1):439-453.

12. Wang S, Cheng H, Ying L, Xiao T, Ke Z, Zheng H, Liang D. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. Magnetic Resonance Imaging 2020;68:136-147.

13. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. IEEE Transactions on Image Processing 2017;26(9):4509-4522.

14. Lee J, Kim H, Chung H, Ye JC. Deep learning fast MRI using channel attention in magnitude domain. 2020. IEEE. p 917-920.

15. Quan TM, Nguyen-Duc T, Jeong W-K. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE transactions on medical imaging 2018;37(6):1488-1497.

16. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, Feng D, Liang D. Accelerating magnetic resonance imaging via deep learning. 2016. IEEE. p 514-517.

17. Du T, Zhang H, Li Y, Pickup S, Rosen M, Zhou R, Song HK, Fan Y. Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation. Medical Image Analysis 2021;72:102098.

18. Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magnetic resonance in medicine 2018;80(5):2188-2201.

19. Souza R, Lebel RM, Frayne R. A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. 2019. PMLR. p 437-446.

20. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929 2020.

21. Srinivas A, Lin T-Y, Parmar N, Shlens J, Abbeel P, Vaswani A. Bottleneck transformers for visual recognition. 2021. p 16519-16529.

22. Joaquim MR, Furth E, Fan Y, Song HK, Pickup S, Cao J, Choi H, Gupta M, Cao Q, Shinohara R, McMenamin D, Clendenin C, Karasick T, Duda J, Gee J, O'Dwyer P, Rosen M, Zhou R. DWI
Metrics Differentiating Intraductal Papillary Mucinous Neoplasms from Pancreatic Cancer: A study in GEM Models 2022 May 7-12, 2022. p page 3430.

23. Li Y, Li H, Fan Y. ACEnet: Anatomical context-encoding network for neuroanatomy segmentation. Medical Image Analysis 2021;70:101991.

24. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 2015. Springer. p 234-241.