2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669582

AN ATTENTION BASED DEEP LEARNING MODEL FOR DIRECT ESTIMATION OF PHARMACOKINETIC MAPS FROM DCE-MRI IMAGES

Qingyuan Zeng
School of Medical Information Engineering
Guangzhou University of Chinese Medicine
Guangzhou, China
18928859910@163.com

Wu Zhou
School of Medical Information Engineering
Guangzhou University of Chinese Medicine
Guangzhou, China
zhouwu787@126.com

***Abstract*—Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful imaging technique that can quantitatively measure the pharmacokinetic (PK) parameters to characterize the microvasculature of tissues. Typically, the PK parameters are extracted by fitting the MR signal intensity of the pixels on the time series with the nonlinear least-squares method. The main disadvantage is that there are thousands of voxels in a single MR slice and the time consumption of voxels fitting to obtain the PK parameters is very large. Recently, deep learning methods based on convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) network have been applied to directly estimate the PK parameters from the acquired DCE-MRI image-temporal series. However, how to effectively extract discriminative spatial and temporal features within DCE-MRI for the estimation of PK parameters is still a challenging problem, due to the large intensity variation of tissue images in different temporal phases of DCE-MRI during the injection of contrast agents. In this work, we propose an attention based deep learning model for the estimation of PK parameters, which can improve the estimation performance of PK parameters by focusing on dominant spatial and temporal characteristics. Specifically, a temporal frame attention block (FAB) and a channel/spatial attention block (CSAB) are separately designed to focus on dominant features in specific temporal phases, channels and spatial areas for better estimation. Experimental results of clinical DCE-MRI from an open source RIDER-NEURO dataset with quantitative and qualitative evaluation demonstrate that the proposed method outperforms previously reported CNN-based and LSTM-based deep learning models for the estimation of PK maps, and the ablation study also demonstrates the effectiveness of the proposed attention-based modules. In addition, the visualization of the attention mechanism reflects interesting findings that are consistent with clinical interpretation.**

***Index Terms*—DCE-MRI, deep learning, attention, pharmacokinetic maps, contrast agent.**

I. Introduction

T1-weighted dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is an imaging technique that involves continuous image acquisition before, during and after the administration of a contrast agent, which has played an important role in the clinical research and applications of myocardial infarction, stroke and tumor related to perfusion and permeability of the vascular[1]. DCE-MRI provides a quantitative measure of pharmacokinetic (PK) parameters characterizing microvasculature of tissues. Typically, PK parameters are extracted by fitting the MR signal intensity of pixels over time series, in which the fitting is usually performed using a nonlinear least-squares approach[2]. However, the main disadvantage is that considering the thousands of voxels in a single MR slice, the calculation requirements for voxel model fitting are very high.

In order to alleviate the above problems of typical model fitting methods for PK parameter estimation, machine learning methods have recently been proposed and have achieved good results. The main idea of machine learning methods for PK parameter estimation is to learn the mapping function between image-time series and corresponding PK parameters. Ulas C. et al first utilized deep convolutional neural networks (CNN) to directly estimate the PK parameters from DCE-MRI, in which a local and global network structure based on CNN was adopted to extract discriminative features from the input DCE image-times series[3]. Subsequently, Kettelkamp J. et al adopted the same local and global network structure to extract discriminative features from DCE image-time series, but further embedded the patient's arterial input function (AIF) as input to improve the estimation performance of PK parameters[4]. Recently, Zou R. et al. proposed to extract long-term dependent features in DCE-MRI through a long short term memory (LSTM) network to estimate PK parameters[5], and further proposed an Attention Bidirectional Long Short-Term Memory(AB-LSTM)network[6], in which bidirectional LSTM cells with 32 features, a fully connected attention layer, and a fully connected layer are applied on the attention weighted features to generate an estimation of the PK parameters. Since the LSTM architecture is able to learn the long-term (time) dependence of the signal, the performance of PK parameter estimation is reported to be improved compared to CNN-based methods. However, how to effectively extract discrim-

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

inative spatiotemporal features in DCE-MRI to estimate PK parameters is still a difficult problem. On the one hand, due to the huge changes in the tissue image during the injection of the contrast agent, not all image content of the DCE-MRI tissue contributes equally to the final estimation. In addition, DCE-MRI includes image acquisition before, during, and after the administration of the contrast agent. The different time stages of DCE-MRI may cause uneven contributions to the final prediction. Therefore, it is foreseeable that the attention mechanism that can focus on the dominant features can potentially select the main features in a specific time stage, channel, and spatial region for better estimation.

In this work, we propose an attention-based deep learning model to estimate PK parameters, which can improve the estimation performance by focusing on the main spatiotemporal features. Specifically, temporal Frame Attention Block (FAB) and Channel/Spatial Attention Block (CSAB) are designed to focus on the dominant features in specific time phases, channels and spatial regions, respectively, for better estimation. Compared with the previously reported methods based on CNN and LSTM, the open source RIDER-NEURO dataset is used to prove the superiority of the proposed method. In addition, the visualization of the attention map is also used to reflect the different contributions of different time frames and DCE-MRI spatial regions to the estimation of PK parameters.

II. RELATED WORK

A. DCE-MRI

DCE-MRI plays an important role in lesion characterization and treatment management[7-10]. In order to obtain the DCE-MRI of the tissue, before or during the rapid imaging technique to obtain the T1-weighted image, the clinician injects a contrast agent (CA) into the plasma, and the CA will gradually spread from the plasma to the entire tissue. As the CA concentration in the tissue changes, the MR signal will also change accordingly. The analysis of the dynamic process of the MR signal can reflect the biological or physiological characteristics of the tissue. A mathematical PK model[11-15] designed to simulate the process of CA exchange between different compartments, by fitting the tissue concentration curve to estimate quantitative physiological parameters (such as vascular permeability, volume fraction). At present, the most widely used mathematical PK model is Tofts model proposed by Tofts. The Tofts model[2,13] regards EES (extravascular extracellular space) and plasma as two compartments, simulating the exchange of CA between these two compartments. Using the nonlinear least-squares method to solve the formula, the PK quantitative physiological parameters can be obtained, such as Ktrans (vascular permeability), Ve (EES volume fraction)[16-17]. However, the current clinical use of the PK model has three shortcomings as follows: First, the fitting of the PK model requires the vascular CA concentration (AIF), and AIF is often difficult to obtain due to problems including flow artifacts, inflow and nonlinear effects of high CA concentrations[18]; Second, the PK model uses tissue CA concentration instead of MR intensity. It is difficult to obtain the concentration (the conversion formula from MR intensity to CA concentration needs to be carefully designed)[19]; Third, the nonlinear least-squares method is calculated based on pixel level, and the tissue has a large number of pixels, while fitting point by point is too time-consuming[20].

B. Machine learning

In order to solve the three major shortcomings of the clinical use of the above PK model, existing studies have introduced machine learning methods. By learning the mapping relationship between the dynamic image sequence and the PK parameter map, it is possible to achieve no need for AIF, no need to convert MR intensity to CA concentration, and improve calculation speed. Ulas C. et al and Kettelkamp J. et al. both use the global+local structure of the CNN model to construct the mapping relationship between the image sequence and the PK parameter map, which is expanded by paralleling 3 dilated convolutional layers (ie global path) on the backbone, thereby extracting global spatial feature information[3-4]. However, only the use of dilated convolution at the spatial level to expand the receptive field to obtain spatial context information will result in the inability to effectively extract temporal context information. The core of DCE-MRI is the time information of the dynamic changes of the MR signal. Therefore, the lack of time information is the shortcoming of the global+local model in the PK parameter fitting problem. In addition, Zou R. et al. proposed the use of LSTM to perform DCE-MRI PK parameter fitting, and introduced an attention mechanism to allow the LSTM model to focus on frames with higher contributions, so as to better extract temporal context information[5-6]. However, it only embeds attention on the temporal level and uses LSTM for feature extraction. It does not consider spatial information and ignores the supplementary information between pixels at different spatial positions in the image. Moreover, in the DCE-MRI sequence with a large number of frames, the models of multi-layer LSTM and multi-layer two-way attention LSTM are too large and difficult to converge, resulting in high training computations.

C. Attention

The attention mechanism has been widely used in the field of deep learning. SE-net and CBAM are representative attention modules[21-22], which have been proven to be very effective in various fields and tasks. SE-NET and CBAM can be summarized as a three-step framework as follows: a) context modeling; b) transform; c) fusion. Context modeling is the most important step, which determines whether the context information of the input feature can be effectively extracted and used for the generation of attention weight[23]. SE-net and CBAM adopt GAP(Global Average Pooling) for context modeling, which has disadvantages[24] as follows: a) There are no trainable parameters to make context modeling adaptive; b) In channel or spatial dimensions, simply keep only one average value or maximum value, and discard the rest of a large amount of effective information.

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

III. METHOD

A. Tofts model

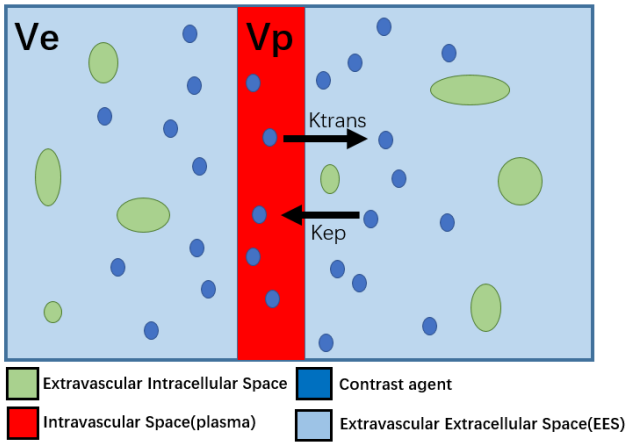

Fig. 1. schematic illustration of the contrast agent transfer in Tofts model between the central(plasma) compartment and the EES space with the Ktrans and Kep rates, respectively.

The PK parameter model of DCE-MRI is based on the nonlinear least-squares method to obtain various physiological PK parameters, and the Tofts model is the most commonly used PK model. As shown in the Figure 1, Tofts model is assumed that the contrast agent flows into the extravascular extracellular space (EES) from the plasma at a rate controlled by the transfer constant Ktrans(min^{-1}), and flows from the EES into the plasma at a rate controlled by Kep(min^{-1}). The Tofts model assumes that the CA concentration of the tissue comes only from EES (preserving Ve) and ignores the contribution of the intravascular space (set Vp=0), where Ve is EES volume per unit volume of tissue, Vp is blood plasma volumn per unit volumn of tissue. Tofts model can be summarized as the following expression:

$$C_t(t) = Ktrans \int_0^t C_p(t') \cdot e^{-(Ktrans/Ve)(t-t')} dt' \quad (1)$$

where $Ve = Ktrans/Kep$, Ct is the CA concentration of the tissue and Cp is the CA concentration after the contrast agent enters the plasma which can be obtained from AIF.

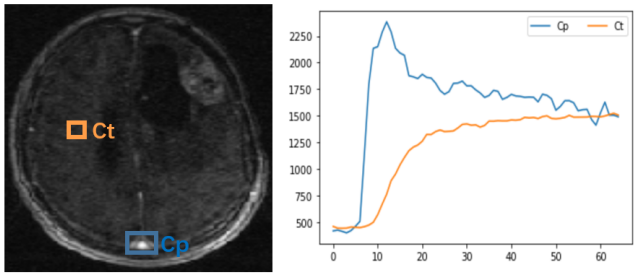

Fig. 2. CA dynamic exchange relationship between plasma and tissue in the brain.

As shown in Figure 2, we can obtain the Ct and Cp curve corresponding to each voxel in DCE-MRI. So the PK parameters can be obtained by using nonlinear least-squares method to fit the Tofts expression(1).

B. DCE-MRI Dataset

We used the publicly available NEURO-RIDER dataset[25] in our research, which contains images of brain tissue from 19 patients, including DCE-MRI obtained by a 1.5 T MR scanner using T1-weighted FLASH sequences. The time resolution is 4.8 seconds, the scan time is 5.2 minutes, and each patient gets 65 frames (or phases). First, we convert the DCE-MR signal intensity-time volume sequence into slice signal intensity-time volume. That is, the signal intensity time volume of each slice has 65 frames. Then, the arterial input function (AIF) curve is obtained by extracting the intensity of the contrast agent in the sagittal sinus region of the brain. Therefore, through the signal intensity-time volume data and the AIF curve of each patient, a typical Tofts model[2,13] can be used to fit the PK parameters. For simplicity, we choose Ktrans, one of the PK parameters that can reflect the permeability of vascular tissue, as the fitting object. The other two PK parameters Kep and Ve can be estimated in the same way. The Ktrans parameter map fitted by a typical Tofts model is regarded as the prediction target of the proposed machine learning method.

C. The framework of the proposed method

In this work, we propose a Frame Channel Spatial Attention network (FCSA-net) in an end-to-end manner to learn the mapping function between the DCE-MRI series and the target Ktrans map generated by the Tofts model. Given the input DCE-MRI series, the proposed FCSA-net can directly output the PK parameter map (Ktrans in Figure 3). As shown in Figure 3, FCSA-net typically consists of two main modules, including the Frame Attention Block (FAB) and the Channel Spatial Attention Block (CSAB). The former FAB module is used to learn which frames of the DCE-MRI sequence are more important and give higher weight to the important frames. The latter CSAB module is used to learn which channels and which parts of the feature map are more important, and provide them with higher weight. In short, the proposed FCSA net uses the attention mechanism to focus on the dominant features in a specific time frame, channel, and spatial region for better estimation. We will illustrate the details of the proposed FAB and CSAB modules in the following sections.

D. FAB module

The structure of the FAB module is shown in Figure 4. The length and width of the DCE-MRI sequence is 32 × 32, and the number of frames is 65. In order to obtain the frame weight vector, we need to reduce the dimensionality of the DCE-MRI sequence in two dimensions of height and width. The dimensionality reduction method is illustrated as follows. The input DCE-MRI sequence passes through two 3 × 3 convolutional layers and two 2 × 2 maximum pooling layers. Each convolutional layer uses relu as the activation function.

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

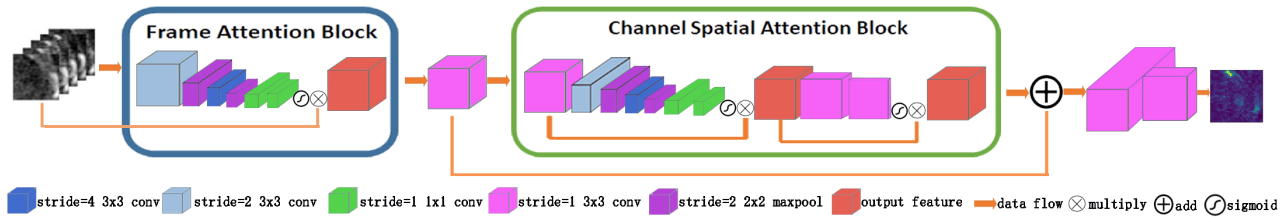

Fig. 3. The architecture of our proposed FCSA-net for direct estimation of Ktrans map.

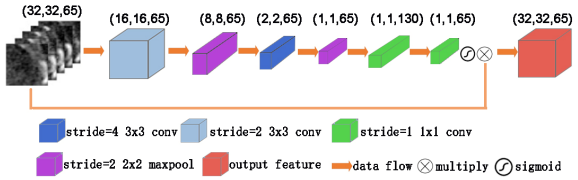

Fig. 4. The structure of the FAB module.

A dimension-reduced feature vector V1 is obtained through the above operations. Followed by two 1 × 1 convolutional layers and Sigmoid activation function, the Frame Weight-Vector(FWV) with a dimension of (1,1,65) is obtained, and the input DCE-MRI sequence is subjected to a dot multiplication operation, and finally the feature map S1 with the length and width 32 × 32 and the channel 65 is generated. The FWV and feature map S1 (red square shown in Figure 4) are defined as

$$V_1 = Pool(Conv_2(Pool(Conv_1(x)))) \tag{2}$$

$$FWV = Sigmoid(FC_2(FC_1(V_1))) \tag{3}$$

$$S_1 = FWV \cdot x \tag{4}$$

where x denotes DCE-MRI input sequence; $Conv_1$ means 3×3 convolution with stride=2/filter=65; $Conv_2$ means 3×3 convolution with stride=4/filter=65; $Pool$ means maximum pooling with stride=2 2×2; FC_1 means 1×1 convolution with stride=1/filter=130; FC_2 means 1×1 convolution with stride=1/filter=65. The above convolutional layers all use relu as the activation function.

E. CSAB module

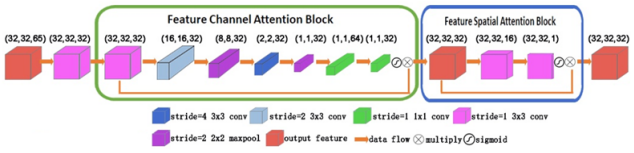

Fig. 5. The structure of the CSAB module.

As shown in Figure 5, in order to obtain the feature channel weight vector, we need to reduce the dimensionality of the feature map in the height and width dimensions. The dimensionality reduction method is illustrated as follows. The input feature map S1 passes through two 3 × 3 convolutional layers to obtain a feature map with a length and width of 32 × 32 and a channel of 32, which is used as the input of the CSAB module. After two 3 × 3 convolutional layers and two 2 × 2 maximum pooling layers, and each convolutional layer uses relu as the activation function, a dimension-reduced feature vector V2 is generated through the above operations. Followed by two 1 × 1 convolutional layers and Sig moid activation function, the Channel-Weight-Vector (CWV) with dimension (1,1,32) is obtained, and the dot multiplication operation is performed with the CSAB input feature map, and finally the feature map S2 with the dimension (32,32,32) is generated. The CWV and feature map S2 (red square shown in the middle of Figure 5) are defined as

$$V_2 = Pool(Conv_5(Pool(Conv_4(Conv_3(Conv_3(S_1)))))) \tag{5}$$

$$CWV = Sigmoid(FC_4(FC_3(V_2))) \tag{6}$$

$$S_2 = CWV \cdot x \tag{7}$$

where $Conv_3$ means 3×3 convolution with stride=1/filter=32; $Conv_4$ means 3×3 convolution with stride=2/filter=32; $Pool$ means 2×2 maximum pooling with stride=2; $Conv_5$ means stride=4/filter=32 3×3 convolution; FC_3 represents 1×1 convolution with stride=1/filter=64; FC_4 represents 1×1 convolution with stride=1/filter=32; Note that the above convolutional layers all use relu as the activation function.

In order to obtain the feature spatial weight matrix, we need to reduce the dimensionality of the feature map in the channel dimension. The dimensionality reduction method is illustrated as follows. The feature map S_2 passes through two 3×3 convolutional layers and the sigmoid activation function to obtain a Spatial Weight-Matrix (SWM) with a dimension of (32,32,1), and performs a dot multiplication operation with the feature map S_2 to obtain a feature map S_3 with the dimension (32,32,32). The SWM and feature map S_3 (red square on the right of Figure 5) are defined as

$$SWM = Sigmoid(Conv_7(Conv_6(S_2))) \tag{8}$$

$$S_3 = SWM \cdot S_2 \tag{9}$$

where $Conv_6$ means 3×3 convolution with stride=1/filter=16 and uses relu as the activation function.$Conv_7$ means 3×3 convolution with stride=1/filter=1.

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

F. FCSA-net

As shown in Figure 3, the proposed FCSA-net is composed of the two FAB and CSAB modules. The output of the overall network is defined as follows:

$$y = Tanh(Conv_9(Conv_8(Conv_3(S_1) + S_3))) \quad (10)$$

where y represents the output of network estimation; $Conv_8$ represents the 3×3 convolution and relu activation function with stride=1/filter=128; $Conv_9$ represents the 3×3 convolution with stride=1/filter=1. The loss function of the network is mean square error (MSE) to reduce the error between the prediction of the network y and the target Ktrans map generated by the Tofts model.

G. The implementation

We used the open-source deep learning framework, Tensorflow, to implement the proposed framework and an RTX2080ti server graphics card. The loss function uses MSE, and we used the Adam optimizer with a learning rate of 5×10^{-4} to minimise the loss function. The maximum number of training iterations is 20k, and the batch-size is 64. The source code for the basic implementation is available at (https://github.com/1057939502/Attention-model-for-PK-estimation).

IV. EXPERIMENTAL RESULTS

A. Train/test dataset and evaluation metrics

The NEURO-RIDER dataset contains DCE-MRI sequences of brain tissues from 19 patients, 14 patients are randomly selected as the training set, and the remaining 5 patients are used as the test set. In the axial view of each patient's 3D brain tissue, we selected the 2D slice with the largest cross-sectional area, up and down 5 slices, with a total of 10 slices. For each slice of each patient, we randomly extracted 4 non overlapping 32 × 32 tissue regions. In this way, each patient can obtain 40 tissue regions with phase = 65 and shape of 32 × 32, that is, 40 signal intensity-time volumes with shape of (32,32,65). The training set requires data augmentation to increase the number of training data, in which each region performs 25 random rotations around the center.

To evaluate the performance of the proposed model for the estimation of Ktrans maps, we adopt two image similarity indexes of Structural similarity index (SSIM) and Root mean square error (RMSE) to evaluate the estimation performance. In addition, Peak Signal to Noise Ratio (PSNR) is also adopted to evaluate the image quality of estimation map. Finally, the consuming time of training for each method are also recorded. In order to reduce the experimental error, each model was repeatedly trained 10 times to take the performance mean and standard deviation. The previously reported CNN-based method[3-4]and the LSTM-based method[5- 6] are also implemented for comparison. In addition, in order to evaluate that our context modeling method with training parameters is better than the one-step method GAP without parameters, we also implement SE-NET and CBAM[21-22], which have similar structure and use GAP to context modeling, for PK parameter estimation for comparison.

TABLE I
CALCULATION COST OF PK PARAMETER MAP OF DIFFERENT METHODS.

Methods	NLS	CNN-based	LSTM-based	Attention-based
Time(msec)	115283	6.2	9.5	7.3

TABLE II
QUANTITATIVE COMPARION OF THE PROPOSED METHOD AND PREVIOUSLY REPORTED METHODS.

Methods	SSIM	PSNR	RMSE
CNN(L)	0.869±0.007	38.65±0.34	0.0129±0.0005
CNN(L+G)[3-4]	0.885±0.013	40.02±0.51	0.0123±0.0009
LSTM[5]	0.892±0.014	40.77±0.59	0.0122±0.0012
AB-LSTM[6]	0.911±0.012	41.52±0.49	0.0115±0.0011
SE-NET[21]	0.903±0.008	41.15±0.41	0.0119±0.0014
CBAM[22]	0.912±0.011	41.88±0.39	0.0106±0.0009
CNN(L)+FAB(ours)	0.918±0.008	41.91±0.34	0.0099±0.0007
CSAB(ours)	0.914±0.008	41.91±0.35	0.0106±0.0008
FCSA-net(ours)	0.921±0.009	42.18±0.46	0.0097±0.0008

B. Quantitative comparison

As shown in Table I, in order to compare the calculation cost between the deep learning model and the nonlinear least squares method for PK parameter estimation, we used a 32×32 DCE-MRI sample to calculate the PK parameter map with different methods. we can see that the calculation cost of NLS(nonlinear least-squares) is very large, while the forward propagation time of the deep learning model is extremely small. Thus, in terms of calculation time, the deep learning model solves the time consumption problem of traditional PK parameter fitting.

As shown in Table II, the CNN-based method with local and global features combination (CNN(L+G)) showed better performance than that of the local features (CNN(L)), and the LSTM showed improved performance than the CNN-based method due to the learning of long-term (temporal) dependence of signals. Note that the recently proposed AB-LSTM obtained further improved performance due to the adoption of Bidirectional LSTM to extract long-term dependence features and the attention layer to weight spatial-temporal features. Comparatively, higher performance was obtained by embedding the proposed FAB module into the CNN based method with local features (CNN(L)+FAB), indicating the superiority of learning frame weights for the prediction. In contrast, the method of AB-LSTM only consider the feature weight instead of the importance of frame weight and did not analyze the influence of attention mechanism on spatial-temporal information in DCE-MRI for PK parameter estimation. It is evident that the proposed CSAB module can also yield higher performance, demonstrating the effectiveness of spatial and channel attention for the prediction. In addition, by comparing

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

FAB with SE-NET, and CSAB with CBAM, we can know that the context modeling method, which extended receptive fields layer by layer and have learning parameters, is better than the one-step, parameterless method GAP. Finally, the proposed FCSA-net that combines the two modules together can obtain the best prediction performance by focusing on dominant spatial and temporal characteristics. The metric of PSNR also shows that the proposed model generates appealing images of estimation maps.

In the viewpoint of time cost for the training as tabulated in Table III, the additional time consuming of the attention mechanism in the proposed model is 3-4 minutes compared with the CNN based method. By comparison, the number of parameters in the LSTM model and the AB-LSTM model is very large, which leads to huge time consuming of training.

C. Visualization comparison

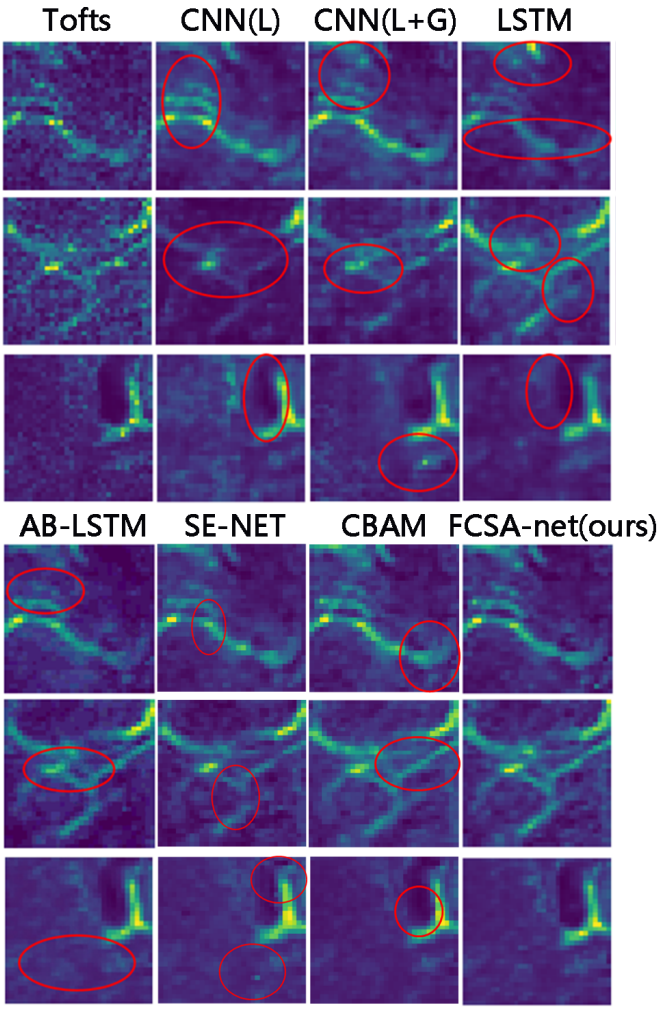

Fig. 6. Visual comparison of Ktrans map estimation with different methods.

To qualitatively assess the performance of Ktrans map estimation with different methods, the red circle shown in Figure 6 is to facilitate the reader to observe the difference area. It can be seen that the quality of the image fitted by the proposed FCSA-net is the best, and the detailed area is fitted to be closest to the Ktrans map generated by the Tofts model. Furthermore, we also visualized the learned weight variables of the attention mechanism. From the Frame-Weight-Vector line charts generated by the FAB module shown in Figure 7(a-c), it can be seen that frames in the early stage of the injection of contrast media have higher weights of contribution, which is consistent with the finding proposed in the clinical interpretation that the rising stage of signal intensity related to the parameter value of Ktrans[2]. The Channel-Weight-Vector line charts shown in Figure 7(d-f) also demonstrate that channels have different contributions to the final prediction. Evidently, the Spatial-Weight-Matrix maps shown in Figure 7(g-i) clearly indicate that brighter areas show greater weights that have more contribu tions to the final prediction in the areas, which further verifies why the proposed method in Figure 6 has better estimation performance in the specific areas. Finally, we adopted the recently proposed visualization method Grad-CAM[26] to reveal which important regions in the image contribute to the improvement of prediction performance. Visualization comparison of the CNN method and the proposed FCSA-net method shown in Figure 8(a-c) and (d-f) by the Grad-CAM method indicates that the attention mechanism in the proposed method can focus on dominant features in specific spatial areas with abundant textures (or edges) and neglect the background areas with noise and no contrast enhancement, resulting in higher prediction performance of Ktrans map in the three cases.

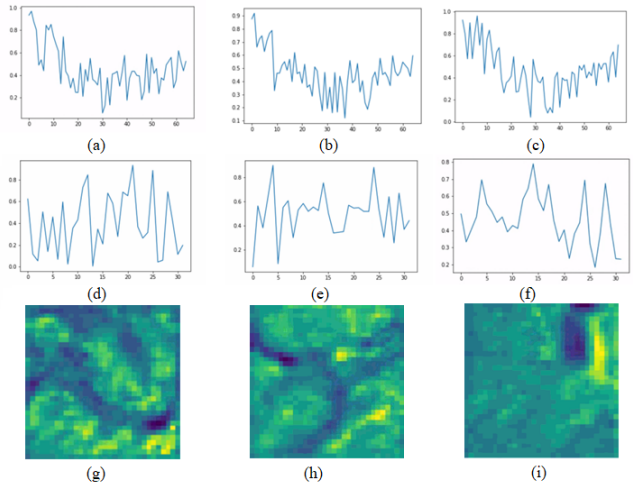

Fig. 7. Visualization of Frame-Weight-Vector (a-c), Channel-Weight-Vector (d-f) and Spatial-Weight-Matrix (g-i) for the three cases.

V. DISCUSSION

In the present study, the PK parameter map calculated on the Tofts model by the nonlinear least-squares method is used as the target of the deep learning model[27]. The deep learning model we designed pursues a smaller computational cost and is closer to the fitting accuracy of the nonlinear

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

TABLE III
TIME CONSUMING FOR TRAINING OF THE PROPOSED METHOD AND PREVIOUSLY REPORTED METHODS.

Methods	CNN(L)	CNN(L+G)[3-4]	LSTM[5]	AB-LSTM[6]	SE-NET[21]	CBAM[22]	CNN(L)+FAB(ours)	CSAB(ours)	FCSA-net(ours)
Time(min)	19	21	232	417	22	23	22	23	25

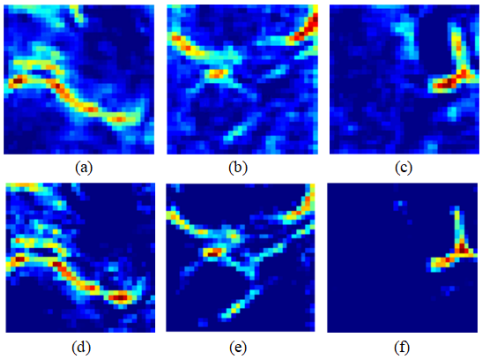

Fig. 8. Visulization results by the Grad-CAM for the CNN method (a-c) and the proposed FCSA-net (d-f) corresponding to the three cases. Color maps are generated by using red weighting to identify areas associated with the prediction.

least-squares method. The experimental results show that the model proposed in this paper is better than the existing deep learning models for PK parameter estimation (Table II). This work is the same as the previous studies[3-6] that choose a common mathematical PK model (Tofts model), and uses the PK parameter map calculated by the nonlinear least-squares method as the target for deep learning model learning.

In this study, we designed an attention-based deep learning network that can effectively extract spatiotemporal information from DCE-MRI for PK parameter fitting. Specifically, we combined the advantages of the deep learning CNN model[3-4] and the LSTM model[5-6] in DCE-MRI, and avoided their respective shortcomings. The FCSA-net we proposed can extract context information from a large receptive field in both time and space dimensions at the same time, and is used to focus on important frames, feature important channels and regions in the DCE-MRI sequence, with less additional training cost to get the best fitting performance.

Inspired by SE-net and CBAM[21-22], we propose two attention modules, FAB and CSAB, to construct an end-to-end FCSA-net. FCSA-net can better extract context information through large receptive fields in space and time dimensions, and focus the model on important frames and regions, thereby maximizing the performance of PK parameter fitting. The existing SE-net and CBAM use GAP in context modeling, which has no self-adaptability and loses a lot of effective information[23,24]. Instead of using GAP in context modeling, we connect the convolutional layers of 3×3 receptive fields and the pooling layers of different scale receptive fields to expand the receptive field and down-sampling layer by layer. This method can make context modeling learnable, retain more effective information from input feature, and obtain more effective time/space weights to focus on more important feature[24]. It can be seen from the two comparative experiments of "SE-net and FAB" and "CBAM and CSAB" in Table II that the context modeling method is more effective than GAP.

At present, previous studies did not use deep learning model to clinically explain the PK parameter fitting of DCE-MRI. A clear understanding of the spatial and temporal relationships in the DCE-MRI sequence is essential for the clinical application of deep networks. In this study, by visualizing the weight vector and matrix of the attention mechanism, we can observe which spatiotemporal information makes a more important contribution to PK parameter fitting. If the importance information can be used in combination with the clinical significance, it can guide the clinicians to formulate the adjustment plan of time and space resolution when the DCE-MRI sequence is collected. For example, it can be seen from the weight vector that the previous frame plays a more important role in PK parameter fitting. Clinicians can appropriately increase the acquisition time resolution in the early stage to obtain a DCE-MRI sequence that better reflects the physiological parameters of PK. In order to save storage and acquisition costs, the acquisition time resolution can be appropriately reduced in the middle and late stages. In addition, the visualization results also verify our hypothesis: not all times, feature channels and regions have the same contribution to PK parameter fitting. We believe that the clinical interpretability of the visualized weight matrix should be of great significance to the diagnosis of clinical diseases in the future.

This study also has some limitations. In order to compare objectively with previous studies on deep network for fitting PK parameters, we also use the PK parameter map calculated by the nonlinear least-squares method as the training target, and train the deep learning network in a supervised manner. However, the dynamic sequence signal of DCE-MRI is susceptible to noise, organ movement, etc., which leads to deviations in nonlinear least-squares fitting[28]. Therefore, the current research only solves the fast calculation problem of PK parameter fitting, which can be used as an alternative to nonlinear least-squares fitting, but the problem of fitting deviation existing in nonlinear least-squares fitting is still unsolved. In the follow-up research work, we will consider research from two aspects: a) we generate an accurate PK parameter map gold standard based on Tofts model in synthetic data to train the attention-based model proposed in this study, and then the domain adaptation technology[29-30] is used to realize the migration between the synthetic data domain and

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.

the real sample data domain, so as to obtain a higher quality PK fitting parameter map on the real data; b) an unsupervised learning deep learning network[31] is constructed based on the Tofts model and the original DCE-MRI image. The network first transforms the input DCE-MRI image into a PK parameter map, and then the PK parameter map is used to reconstruct the DCE-MRI image based on the Tofts model, and the mean square error between the reconstructed DCE-MRI image and the original DCE-MRI image is used as the loss function of the deep network, so that the generation of the PK parameter map without labels can be realized. No matter which machine learning method is used, the model proposed in this paper can be used as the backbone of the deep network to realize the mapping relationship learning from DCE-MRI sequence image to PK parameter map, and obtain better PK parameter fitting effect.

VI. CONCLUSION

In this study, we proposed an attention based deep learning model FCSA-net for the estimation of PK parameters. The open source RIDER-NEURO dataset verified the superiority of the proposed method compared with the previously reported CNN-based method and the LSTM-based method. In addition, visualization of the attention mechanism also indicated the effectiveness of the proposed attention modules for the PK parameters estimation of DCE-MRI.

REFERENCES

[1] GJM. Parker, DL. Buckley,"Tracer Kinetic Modelling for T1-Weighted DCE-MRI", A. Jackson, DL. Buckley, GJM. Parker, editors. Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology. Berlin; Heidelberg: Springer, pp.81-92, 2005.

[2] PS. Tofts, G. Brix, DL. Buckley, et al, "Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols", J Magn Reson Imaging, no.10, pp.223-32, 1999.

[3] C. Ulas, D. Das, MJ. Thrippleton, et al, "Convolutional Neural Networks for Direct Inference of Pharmacokinetic Parameters: Application to Stroke Dynamic Contrast-Enhanced MRI", Frontiers in Neurology, no.9, 1147, 2019.

[4] J. Kettelkamp, SG. Lingala, "Arterial input function and tracer kinetic model-driven network for rapid inference of kinetic maps in Dynamic Contrast-Enhanced MRI(AIF-TK-net)", IEEE 17th International Symposium on Biomedical Imaging (ISBI) April 3-7, Iowa City, Iowa, USA, pp.1450-1453, 2020.

[5] J. Zou, JM. Balter, Y. Cao, "Estimation of Pharmacokinetic Parameters from DCE-MRI by Extracting Long and Short Time-dependent Features Using an LSTM Network", Medical Physics, no.3, 14222, 2020.

[6] J. Zou, J. Balter, Y. Cao, Estimation of Pharmacokinetic Parameters from DCE MRI Using an Attention Bidirectional Long Short-term Memory Neural Network. International Journal of Radiation Oncology Biology Physics, vol.108, no.3, 1 November, pp.S130. 2020.

[7] DJ Collins,AR Padhani, "Dynamic magnetic resonance imaging of tumor perfusion", IEEE Engineering in Medicine and Biology Magazine, 23:65–83, 2004

[8] AJ Farrall, JM Wardlaw, "Blood–brain barrier: ageing and microvascular disease-systematic review and meta-analysis", Neurobiology of Aging, 30:337–52, 2007

[9] JM Wardlaw,C Smith,M Dichgans, "Mechanisms of sporadic cerebral small vessel disease: insights from neuroimaging", Lancet Neurology, 12:483–97, 2013

[10] AK Heye, RD Culling, et al, "Assessment of blood–brain barrier disruption using dynamic contrast-enhanced MRI", A systematic review. Neuroimage, 6:262–74, 2014

[11] H.B.W. Larsson, M. Stubgaard, et al, "Quantitation of blood-brain barrier defect by mag netic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumors.", Magnetic Resonance in Medicine, 16(1), 117–131,1990.

[12] G. Brix, W. Semmler, et al, "Pharma cokinetic parameters in CNS Gd-DTPA enhanced MR imaging.", Journal of Computer Assisted Tomography, 15(4), 621–628, 1991.

[13] P.S. Tofts and A.G. Kermode, "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamen tal concepts.", Magnetic Resonance in Medicine, 17(2), 357–367, 1991.

[14] C.S. Patlak, R.G. Blasberg, and J.D. Fenstermacher, "Graphical evalua tion of blood-to-brain transfer constants from multiple-time uptake data.", Journal of Cerebral Blood Flow and Metabolism, 3(1), 1–7, 1983.

[15] P.S. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging.", Journal of magnetic resonance imaging, 7(1), 91–101, 1997.

[16] GJM Parker, DL Buckley, Tracer Kinetic Modelling for T1-Weighted DCE MRI. In: Jackson A, Buckley DL, Parker GJM, editors. Dynamic Contrast Enhanced Magnetic Resonance Imaging in Oncology. Berlin;Heidelberg: Springer , pp. 81–92, 2005

[17] RM Lebel, J Jones, JC Ferre, M Law, KS Nayak. "Highly accelerated dynamic contrast enhanced imaging.", Magnetic Resonance in Medicine, 71:635–44 2014

[18] GJM Parker, C. Roberts, et al, "Experimentally-derived functional form for a population-averaged high temporal-resolution arterial input function for dynamic contrast-enhanced MRI." Magnetic Resonance in Medicine, 56(5), 993–1000, 2006

[19] D. Gadian, J. Payne, D. Bryant, et al, "Gadolinium-DTPA as a contrast agent in mr imaging-theoretical projec tions and practical observations.", Journal of Computer Assisted Tomography, 9(2), 242–251, 1985

[20] C Debus, R Floca, D Nrenberg, A Abdollahi, M Ingrisch, "Impact of fitting algorithms on errors of parameter estimates in dynamic contrast-enhanced MRI.", Physics in Medicine Biology, 62:9322, 2017

[21] H Jie, S Li, S Gang, et al, "Squeeze-and-Excitation Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, PP.99, 2017.

[22] S Woo, J Park, J Y Lee, et al, "CBAM: Convolutional Block Attention Module." European Conference on Computer Vision(ECCV), Springer, Cham, 2018.

[23] Y Cao, J Xu, S Lin, et al, "GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond", 2019 IEEE/CVF International Conference on Computer Vision Workshop(ICCVW), IEEE, 2020.

[24] J Hu, L Shen, S Albanie, et al, "Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks.", Conference and Workshop on Neural Information Processing Systems (NIPS), 2018

[25] D. Barboriak, "Data From RIDER NEURO MRI. The Cancer Imaging Archive", 2015.[Online]. Available: http://doi.org/10.7937/K9/TCIA.2015.VOSN3HN1.

[26] RR. Selvaraju, M. Cogswell, A. Das, et al, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, pp.618-626, 2016.

[27] SP Sourbron, DL Buckley, "Classic models for dynamic contrast-enhanced MRI.", NMR IN BIOMEDICINE, 26:1004–27, 2013

[28] BM Kelm, BH Menze, O Nix, CM Zechmann, FA Hamprecht, "Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge.", IEEE Transactions on Medical Imaging , 28:1534–47, 2009

[29] J Kamphenkel, PF Jaeger, S Bickelhaupt, et al, "Domain Adaptation for Deviating Acquisition Protocols in CNN-based Lesion Classification on Diffusion-Weighted MR Images" , arXiv.org, 2018.

[30] E Tzeng, J Hoffman, N Zhang, et al, "Deep Domain Confusion: Maximizing for Domain Invariance.", Computer Science, 2014.

[31] S Barbieri, OJ Gurney Hampion, R Klaassen, et al. "Deep learning how to fit an intravoxel incoherent motion model to diffusion-weighted MRI,", Magnetic Resonance in Medicine, 83(2), 2020

Authorized licensed use limited to: Univ of Calif San Francisco. Downloaded on July 23,2024 at 02:47:16 UTC from IEEE Xplore. Restrictions apply.