2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

A Temporal Deep Learning Approach for MR Perfusion Parameter Estimation in Stroke

King Chung Ho[1*], Fabien Scalzo[2*], Karthik V. Sarma[1], Suzie El-Saden[1], Corey W. Arnold[1]
[1] Medical Imaging Informatics Group, Department of Radiological Sciences,
[2] Departments of Neurology and Computer Science,
University of California Los Angeles, Los Angeles, CA, 90024, USA
Email: johnny5550822@g.ucla.edu, {fscalzo, ksarma, sels, carnold}@mednet.ucla.edu
* *These authors contributed equally to this work.*

***Abstract*— Perfusion magnetic resonance (MR) images are often used in the assessment of acute ischemic stroke to distinguish between salvageable tissue and infarcted core. Deconvolution methods such as singular value decomposition have been used to approximate model-based perfusion parameters from these images. However, studies have shown that these existing deconvolution algorithms can introduce distortions that may negatively influence the utility of these parameter maps. There is limited previous work on utilizing machine learning algorithms to estimate perfusion parameters. In this work, we present a novel bi-input convolutional neural network (bi-CNN) to approximate four perfusion parameters without using an explicit deconvolution method. These bi-CNNs produced good approximations for all four parameters, with relative average root-mean-square errors (ARMSEs) ≤ 5% of the maximum values. We further demonstrate the utility of the estimated perfusion maps for quantifying the salvageable tissue volume in stroke, with more than 80% agreement with the ground truth. These results show that deep learning techniques are a promising tool for perfusion parameter estimation without requiring a standard deconvolution process.**

I. INTRODUCTION

Stroke is the second most common cause of death worldwide and remains a leading cause of long-term disability. Recanalization of the occluded vessel is the objective of current therapies and can lead to recovery if it is achieved early enough. However, recanalization is also associated with higher risks of hemorrhagic transformation especially in the context of poor collateral flow and longer time to treatment. While safety time windows have been established based on population studies, a given individual patient may be unnecessarily excluded from a high-impact treatment opportunity. Therefore, there is a need to better utilize available imaging data to evaluate the risks and benefits of an intervention toward a more individualized treatment recommendation.

Magnetic resonance (MR) and computed tomography (CT) perfusion imaging are widely used for quantification of cerebral perfusion in the clinical diagnosis of acute stroke. The benefit of a potential endovascular intervention can be assessed using an estimation of the hypoperfused tissue volume [1] that could be salvaged by a successful recanalization and the resulting reperfusion. Model-based perfusion parameters such as cerebral blood volume (CBV), cerebral blood flow (CBF), time-to-maximum (Tmax), and mean transit time (MTT) can be derived from MR and CT perfusion and used to estimate the volume of salvageable brain tissue. To be able to compute these physiological parameters (CBF, CBV, Tmax, and MTT) from perfusion imaging, the concentration time curve ($CTC(t)$) observed at any voxel in the collected raw data must be processed. Because the injected contrast bolus is not a perfect impulse and varies across acquisitions, it is generally assumed that it follows the Indicator-Dilution Theory [2], i.e., the observed CTC of a voxel is the convolution of the arterial input function, *AIF(t)*, with a residue function, *R(t)*, scaled by CBF:

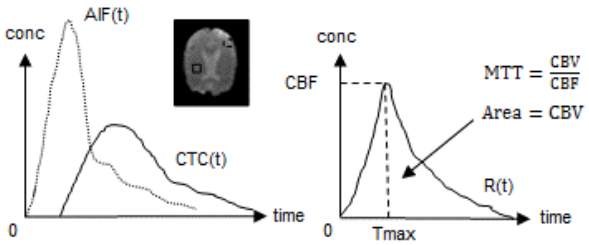

Fig. 1. An illustration of the tissue concentration time curve (CTC), arterial input function (AIF), and residue function (R) corresponding to a voxel. The CTC and AIF are observable from the raw perfusion images, whereas R is obtained via deconvolution of CTC and AIF. The perfusion parameters (CBV, CBF, Tmax, MTT) are normally defined in the R(t).

$$CTC(t) = \mathrm{CBF} \cdot \big(AIF(t) * R(t)\big), \qquad (1)$$

where *R(t)* represents the fraction of observed contrast remaining in the vasculature (within a voxel) at a certain time *t*, and *AIF(t)* describes the contrast input to the vasculature (within a voxel) at a certain time *t*. The perfusion parameters (CBF, CBV, Tmax, MTT) are defined from *R(t)* (Fig. 1), and cannot be directly observed from the imaging data.

The most straightforward techniques to obtain *R(t)* are based on deconvolution of the CTC with the AIF using singular value decomposition (SVD). Because the acquired concentration curves are generally very noisy, this deconvolution may produce residue functions that are not physiologically plausible and subject to distortions that can underestimate the perfusion parameters. Recognizing this limitation, several groups have developed alternative techniques that provide more robust estimates of perfusion parameters. Gaussian Process deconvolution [3] uses Gaussian priors for individual time points of the residue function. This produces a smoother estimate of the residue function. A smooth estimate of the residue function can also be obtained using

Tikhonov regularization where an oscillation penalty is applied in a least squares solution [4] or using Gamma-variate functions. Bayesian estimation of perfusion parameters [5] has also received substantial attention and has been shown to successfully handle higher levels of noise at the cost of longer computation time.

In this work, we present a novel alternative solution to estimate perfusion parameters based on recognizing patterns from data. The proposed model is a bi-input convolutional neural network (bi-CNN) that takes in the signal of interest (i.e., CTC) and the AIF to produce estimated perfusion parameter values. Here, we apply the bi-CNN to estimate four perfusion parameters: CBV, CBF, Tmax, and MTT. Our results show that the bi-CNN estimations are comparable to the existing techniques, demonstrating that CNNs are capable of learning rich temporal feature filters that can extract important patterns from the data to make accurate parameter estimations. To the best of our knowledge, our work is the first to leverage deep learning techniques in the area of perfusion parameter estimation. This work introduces an alternative method that has the potential to improve the current quantitative analysis of perfusion images (e.g., increased robustness to noise), which may ultimately impact the stroke medical decision process and improve patient outcomes.

II. RELATED WORK

A. Deconvolution Methods

Several methods have been developed to estimate perfusion parameters using deconvolution. The standard technique is the singular value decomposition (SVD), which is a deconvolution technique to solve (1) in order to obtain R(t) from which the perfusion parameters are defined. Several variations of SVD have been proposed. Delayed-corrected SVD (dSVD) [6] was developed to perform deconvolution while doing delay correction for contrast delay. Another common delay-insensitive method is the block-circulant SVD (bSVD) [7], which employs a block-circulant decomposition matrix to remove the causality assumption built into standard SVD. Additionally, an oscillation index (OI) can be used as a threshold in an iterative process of repeating bSVD deconvolution to identify the best residue function; this approach is known as oscillation-index SVD (oSVD) [7]. Other less common techniques include Gaussian process deconvolution [3], Tikhonov regularization [4], and Bayesian approximation [5]. Although a number of techniques have been developed to approximate perfusion parameters, none have taken a machine learning approach. Our method departs from prior work in that the estimation relies purely on identifying patterns (features) from the inputs, with feature filters that are learned automatically from the input perfusion data. Our approach is inspired by the many recent successes in deep learning algorithms.

B. Feature Learning in Convolutional Neural Networks

Many state-of-the-art classification records have been achieved by deep convolutional neural networks (CNNs) [8]. The success of these systems is heavily based on the powerful capability of CNNs to learn feature filters from pixels in different settings, such as images [9], [10], and videos [11], [12]. These data-driven features are learned by hierarchical convolutional feature filters, which have been shown to be effective in detecting local characteristics that improve classification [13]. This inspires us to adopt CNNs for perfusion parameter estimation. By generating a large amount of training data from the brain (a brain perfusion image typically contains more than ten thousand voxels), it is possible to train a well-tuned deep CNN for parameter estimations. To the best of our knowledge, this is the first time stroke MR perfusion parameters are estimated by deep neural networks.

III. PROBLEM FORMULATION

In this section, we will briefly describe the biological derivation and definition of the four perfusion parameters of interest (CBV, CBF, MTT, Tmax), and their applications in stroke. We will illustrate the use of standard singular value decomposition (SVD) to obtain the residue function for the perfusion parameters. We will define the estimation task and discuss the proposed approach in the next section.

A. Indicator-Dilution Theory in Tissue Perfusion

In MR perfusion imaging, a bolus of contrast dye is injected intravenously into a patient during continuous imaging, allowing for the concentration of contrast to be measured for each voxel over time as the bolus is disseminated throughout the body. In order to reason with this temporal data, model-based perfusion parameters have been defined and are calculated to create parameter maps of the brain. These parameters are useful for identifying potentially salvageable tissue that can be saved by treatment [1]. Typically, tissue perfusion is modeled by the Indicator-Dilution theory [2], where the measured tissue concentration time curve (CTC) of a voxel is directly proportional to the convolution of the arterial input function (AIF) and the residue function (R), as scaled by CBF in (1). This model follows the principle of the conservation of mass, meaning that the amount of contrast entering the voxel is equal to the sum of the contrast leaving the voxel and the contrast within the voxel. To obtain the perfusion parameters, we need to first derive the residue function (R) using SVD.

B. Singluar Value Decomposition (SVD)

Equation (1) can be expressed in integral form [7] as:

$$CTC(t) = \mathrm{CBF} \int_0^t AIF(t)R(t-\tau)d\tau, \tag{2}$$

In perfusion images, *CTC(t)* and *AIF(t)* can be observed from the raw signals. To obtain R(t) by SVD, (2) is first discretized to:

$$ctc(t_j) = \Delta t \cdot \mathrm{CBF} \cdot \sum_{i=0}^{j} AIF(t_i) \cdot R(t_j - t_i), \tag{3}$$

where Δt is the sampling frequency. Equation (3) may then be formulated as an inverse matrix problem:

$$\begin{bmatrix} ctc(t_0) \\ ctc(t_1) \\ \vdots \\ ctc(t_{N-1}) \end{bmatrix} = \Delta t \begin{bmatrix} AIF(t_0) & 0 & \cdots & 0 \\ AIF(t_1) & AIF(t_0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ AIF(t_{N-1}) & AIF(t_{N-2}) & \cdots & AIF(t_0) \end{bmatrix} \times \begin{bmatrix} R(t_0) \\ R(t_1) \\ \vdots \\ R(t_{N-1}) \end{bmatrix} \cdot \mathrm{CBF}, \tag{4}$$

$$\boldsymbol{c} = \boldsymbol{A} \cdot \boldsymbol{b}, \tag{5}$$

where $\boldsymbol{c}$ represents the *CTC(t)*, $\boldsymbol{A}$ represents the *AIF(t)*, and $\boldsymbol{b}$ represents the *R(t)* (constants are not shown for simplification). Using SVD, we can decompose $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U} \cdot \boldsymbol{S} \cdot \boldsymbol{V^T}, \tag{6}$$

$$\boldsymbol{A^{-1}} = \boldsymbol{V} \cdot \boldsymbol{W} \cdot \boldsymbol{U^T}, \tag{7}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and $\boldsymbol{S}$ is a non-negative square diagonal matrix, $\boldsymbol{W}$=1/$\boldsymbol{S}$ along the diagonals and zero elsewhere. Then, $\boldsymbol{b}$ (*R(t)*) can be obtained as following:

$$\boldsymbol{b} = \boldsymbol{V} \cdot \boldsymbol{W} \cdot \boldsymbol{U^T} \cdot \boldsymbol{c}\,. \tag{8}$$

C. Perfusion Parameter Definition

Four parameters (CBV, CBF, MTT, Tmax) can be defined from *R(t)*. CBV describes the total volume of flowing blood in a given volume of a voxel. It is equal to the area under the curve of *R(t)*. CBF describes the rate of blood delivery to the brain tissue within a volume of a voxel, and is the constant scaling factor of the ratio between the CTC and the convolution of the arterial input function (AIF) and the residue function in (1). It is equal to the maximum value of the residue function. By the Central Volume Theorem, CBV and CBF can be used to derive MTT, which represents the average time it takes the contrast to travel through the tissue volume of a voxel. Tmax is the time point where the *R(t)* reaches its maximum. It approximates the time needed for the bolus to arrive at the voxel. The mathematical expressions of these parameters are listed in the following:

$$\mathrm{CBV} = \int_0^\infty R(t)dt\,. \tag{9}$$

$$\mathrm{CBF} = \max\big(R(t)\big)\,. \tag{10}$$

$$\mathrm{MTT} = \frac{\mathrm{CBV}}{\mathrm{CBF}}. \tag{11}$$

$$\mathrm{Tmax} = \arg\max_t(R(t))\,. \tag{12}$$

These parameters are important to characterize the underlying tissue. A patient with arterial occlusion and ischemic stroke normally has a substantial drop in CBF and CBV, and a higher Tmax in the affected brain volume distal to the blood vessel blockage. At first, the affected brain volume maybe still be salvageable by treatment, but irreversible damage occurs over several hours due to insufficient blood supply. Thresholds have been established for these perfusion parameters that define the volume of dead tissue core and the under-perfused but potentially salvageable tissue [14], [15].

D. The Estimation Task

Our estimation task is to estimate the four perfusion parameters (CBV, CBF, MTT, Tmax) for a voxel, given its CTC and AIF. We propose a pattern recognition model for this task in the form of a novel bi-input convolutional neural network (bi-CNN), which takes the two inputs (CTC, AIF) and generates an estimated perfusion value for a voxel. Separate bi-CNNs were trained to estimate each perfusion parameter. The overall estimation task is defined as:

$$v = f(AIF, CTC), \tag{13}$$

where *v* is the estimated value, and *f(·)* is the bi-CNN with the trained weights. The bi-CNN is trained with thousands of training patches to learn important features from the input data to make an approximation.

IV. Approach

A. Training Data Definition

Each training example (voxel patch) consists of a pair of the CTC and its AIF. A CTC or AIF is a one dimensional vector with a size of 1 x *t,* where *t* is the number of time point (in this work, *t* = 70). As previous work suggests [16], regional information corresponding to a voxel's surroundings can improve prediction in MR images. Therefore, a small region is included in each training voxel, resulting in a size of 3 x 3 x *t* patch (width x height x time; the z-dimension is omitted), where the center of the patch is the voxel of interest for estimation.

B. Bi-input Convolutional Neural Network (bi-CNN)

The bi-CNN (Fig. 2) consists of three components: (1) convolution, (2) maps stacking, and (3) fully-connected. In the convolution, a CTC and its AIF are convolved independently via multiple convolutional layers (i.e., two convolution chains), where temporal filters are learned. Each convolution chain follows a denoising architecture [17] that attempts to remove artifacts (e.g., noise, distortion) that are often seen in the input perfusion signals, which is important to identify fine-grained features from CTC and AIF signals that help estimation. As suggested [17], a simple signal with artifacts model can be defined as:

$$y = x * k, \tag{14}$$

where *y* is the observed 1D signal (instead of a 2D image), *x* is the original artifact-free signal, and *k* is the convolution kernel that is caused by the artifacts. When we apply the Fourier transform operator, $F(\cdot)$, with Tikhonov regularizer, *x* can be expressed as:

$$x = F^{-1}\left(\frac{1}{F(k)}\left\{\frac{|F(k)|^2}{|F(k)|^2+\frac{1}{SNR}}\right\}\right) * y = k^* * y, \tag{15}$$

where SNR is the signal to noise ratio and k^* is the pseudo inverse kernel. The new representation of *x* can be further expanded into a matrix representation by the kernel separability theorem, where k^* is decomposed into $k^* = \boldsymbol{U} \cdot \boldsymbol{S} \cdot \boldsymbol{V}^T$. This leads to a new representation of *x*:

$$x = k^* * y = \textstyle\sum_j s_j \cdot u_j * \left(v_j^T * y\right), \tag{16}$$

where u_j and v_j are the j[th] columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively, and s_j is the j[th] singular value. This new expression shows that the original artifact-free signal, *x*, can be obtained via the weighted sum of separable 1D filters [17]. This leads to the design of a convolution chain where two separated 1D convolutions are performed (L1 to L2, and L2 to L3), with filter size of 1 x 1 x 36 and 1 x 1 x 35 respectively. We add a convolutional layer (L3 to L4) after the denoising architecture to learn filters for detecting the spatial contributions of neighboring voxels. The output feature maps of the convolution chains are then stacked together in the maps stacking layer (L5), resulting in a matrix with a size of 64 x 2 x 2 x 1. It is then connected to two fully-connected layers where hierarchical features are learned to

correlate the AIF and CTC derived features. The output of the network (L8) is the estimated parameter value. The training optimization of the network is to obtain network weights, Θ, that minimize the mean squared loss between the true value, V, and the estimated value, $\hat{V}(\Theta)$, across the samples with size n:

$$\arg min_{\Theta}\, loss = \frac{1}{n}\sum_{i=1}^{n}(V_i - \hat{V}_i(\Theta))^2, \qquad (17)$$

C. Architecture for each Perfusion Parameter

It is important to observe that the previously mentioned bi-CNN does not contain any max-pooling layers. This architecture worked well for CBV, MTT, and Tmax estimation, but it did not perform well for CBF estimation, which is a physiological parameter representing the maximum cerebral blood flow. The biological nature of CBF inspires us to apply a max-pooling layer (with a max operator), which helps identifying maximum values. The max-pooling layer is inserted into L3 to replace the second convolutional layer in each convolutional chain for bi-CNNs of CBF. The size of the max-pooling layer is set to 1 x 1 x 35 to maintain the size consistency across the rest of the network. This change significantly improves the CBF estimation.

V. Experiments

A. Dataset

MR perfusion data was collected retrospectively for a set of 11 patients treated for acute ischemic stroke at UCLA. The ground truth perfusion maps (CBV, CBF, MTT, Tmax) and AIFs were generated using bSVD in the sparse perfusion deconvolution toolbox [18] and the ASIST-Japan perfusion mismatch analyzer [19] respectively. All the perfusion images were interpolated to have a consistent 70s time interval for bi-CNNs. The ranges of CBV, CBF, MTT, and Tmax values are between 0-201 ml/100g, 0-1600 ml/100g/min, 0-25.0 s, and 0-69 s (Tmax was clipped at 11s because there were too few examples beyond this value) respectively. During experiments, we observed that unequal sampling of the training data can lead to biased prediction [20]. This motivated us to group each perfusion parameter value into ten bins, and draw equal sized training samples from each bin. This resulted in four sets of training data (CBV, CBF, MTT, Tmax), with sizes of 91,950, 97,110, 87,080, and 74,850 respectively.

B. CNN Configuration and Implementation

The overview of the bi-CNN is shown in Fig. 2. A training example consists of a pair of inputs: CTC and its AIF, with a size of 3 x 3 x 70. Each convolution chain consists of three convolutional layers where 32 maps are learned (with zero-padding and a stride of 1). A non-linear rectified linear unit (ReLU) layer is attached to every convolutional layer and fully-connected layer (except for the max-pooling layer). We note that there are two changes that are important to optimize the performance of the model, which are different from standard CNN configurations [21]. First, dropout was not included in the fully-connected layers because we observed that it decreased performance during validation. We suspect that this may be due to the nature of the problem of parameter estimation (i.e., estimating a continuous value versus predicting a categorical label), where every output unit may contribute (to some degree) to the estimated value. Second, the initial learning rates are different for different parameter estimations. We observed that the training losses can easily explode when the learning rate is too high, especially for perfusion parameters with high maximum values (e.g., max(CBF) = 1600). Therefore, the initial learning rates for CBV, CBF, MTT, and Tmax were 0.0005, 0.00005, 0.005, 0.005 respectively, with a learning rate decay of 1e-8, 1e-9, 1e-7, 1e-7 respectively.

The bi-CNN was trained with batch gradient descent (batch size: 50; epochs: 10) and backpropagation. A momentum of 0.9 were used. A heuristic was applied to improve the learning of deep CNN weights [21], where the learning rate was divided by 10 when the validation error rate stopped improving with the current learning rate. This heuristic was repeated three times. The deep CNN was implemented in Torch7, and the training was done on a NVIDIA Tesla K40 GPU.

C. Evaluation

The performance of the bi-CNN estimators was evaluated by leave-one-patient-out cross-validation (i.e., training was performed excluding data from one patient and then evaluating the results on that held-out patient). The average root-mean-square error (ARMSE) of validations was calculated using following definition:

$$ARMSE = \frac{1}{n_T}\sum_{j=1}^{n_T}\sqrt{\frac{1}{s_j}\sum_{i=1}^{s_j}(V_{i,j} - \hat{V}_{i,j})^2}\,, \qquad (18)$$

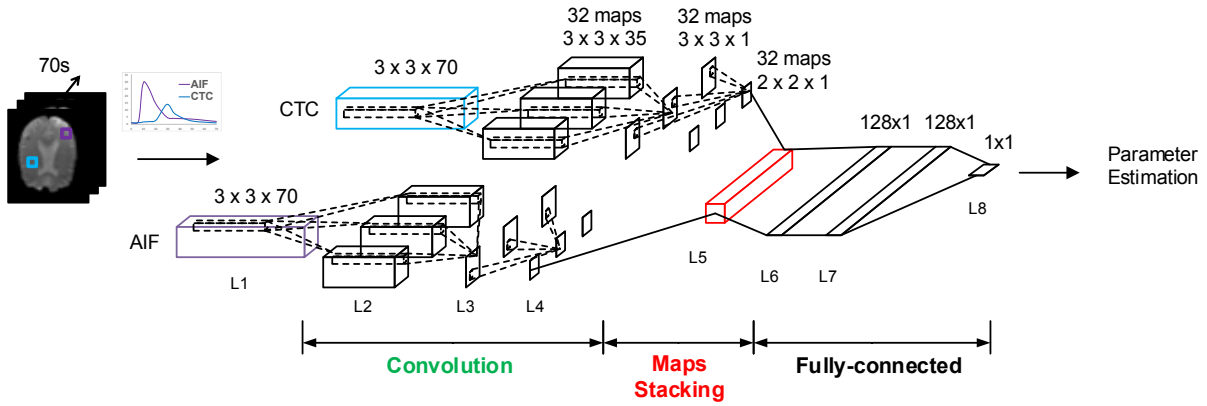

Fig. 2. The proposed bi-CNN. It consists of three components: (1) convolution, (2) maps stacking, and (3) fully-connected. Feature maps are first learned separately for a CTC and its AIF in the convolution chains which follow the denoising architecture [17]. The feature maps are then stacked together in the maps stacking component, followed by two fully-connected layers to learn a combined feature representation for parameter estimation. The size of the outputs from each layer operation (e.g., convolution) and partial connections between layers are shown.

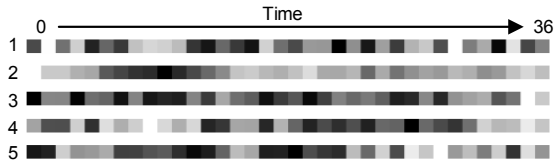

Fig. 3. Examples of learned temporal filters (1 x 1 x 36) in the first layer of the CTC convolution chain. Each row represents a temporal filter; each column represents a unit filter at a time point. These filters capture signal changes along the time dimension for parameter estimation.

where n_T is the total number of patients, V is the ground truth value, $\hat{V}$ is the estimated value, and s_j is the number of samples. We also demonstrated the utility of the bi-CNN by comparing the salvageable tissue binary masks generated from the bi-CNN and the ground truth perfusion maps. Published CBF and Tmax thresholds [14], [15] were used to define the salvageable tissue binary masks. The similarity between these masks (the ground truth mask, A, and the estimated mask, B) was calculated using the Dice coefficient [22]:

$$\text{Dice}(A,B) = 2\frac{|A\cap B|}{|A|+|B|}. \quad (19)$$

A value of 0 indicates no overlap, and a value of 1 indicates perfect similarity (i.e., B=A). A good overlap between masks is generally considered to have occurred when the Dice coefficient is larger than 0.7.

D. Results and Discussion

Fig. 3 shows some examples of learned convolutional filters from the first layer of the CTC convolution chain. Each row represents a 1 x 1 x 36 temporal filter and each column is a unit filter at a time point. As can be seen, these filters capture high signals (white) and low signals (black) at different time points, which helps the fine-grained temporal feature detections from the source signals. This is important to identify features for accurate parameter estimation. Using these learned temporal filters, the bi-CNNs achieved an ARMSE of 4.80 ml/100g, 27.4 ml/100g/min, 1.18 s, 1.33 s for CBV, CBF, MTT, and Tmax respectively, which are equivalent to 2.39%, 1.71%, 4.72%, and 1.19% of the individual perfusion parameter's maximum value. The small ARMSE results show that the bi-CNNs are capable of learning feature filters to approximate perfusion parameters from CTCs and AIFs without using standard deconvolution.

Examples of estimated perfusion maps are shown in Fig. 4. All of the estimated perfusion maps (CBV, CBF, MTT, and Tmax) showed good alignment with the ground truth and hypoperfusion (i.e. less blood flow or delayed Tmax) could be observed visually from some of the estimated maps (red boxes). The differences between the estimated maps and the ground truth were minimal. To further verify the usability of the estimated perfusion maps, a CBF cutoff of 50.2 ml/100g/min [14] and a Tmax cutoff of 4s [15] were used to generate the salvageable tissue masks from the ground truth and the estimated perfusion maps (Fig. 5). The average Dice coefficients for the CBF and Tmax masks were 0.830±0.109 and 0.811±0.071 respectively, showing good overlap between the ground truth masks and the estimated masks. This result shows that the bi-CNN can generate useful masks for salvageable tissue approximation.

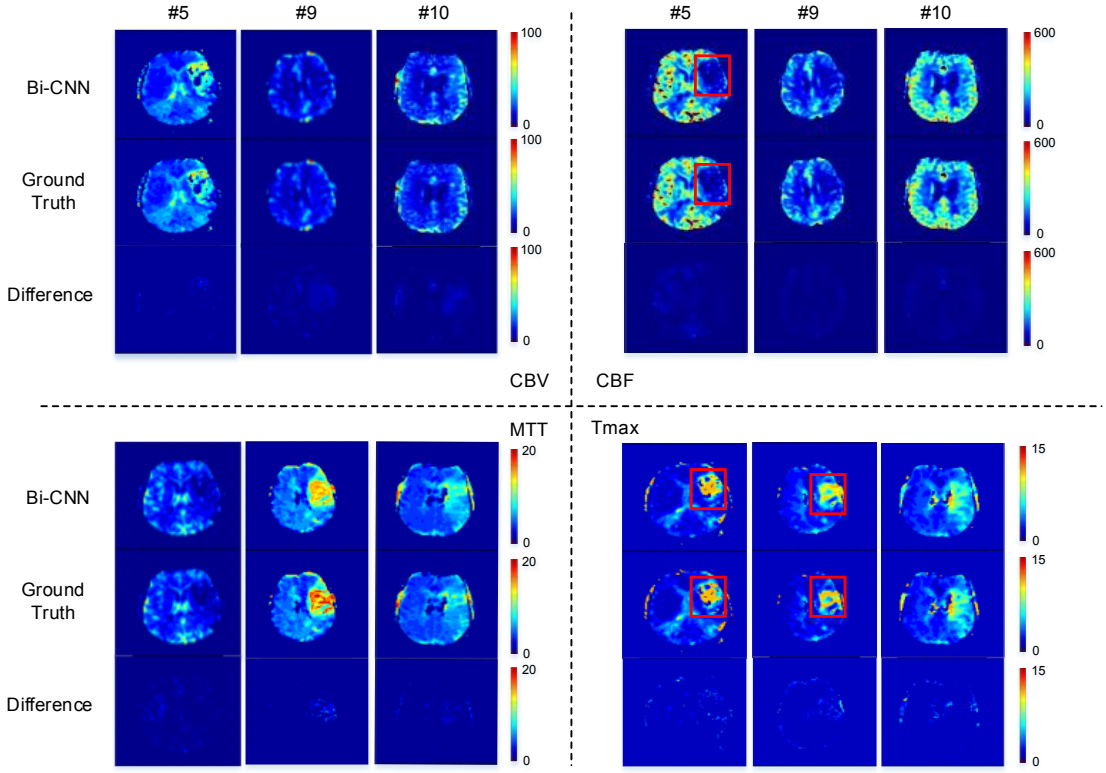

Fig. 4. Examples of estimated perfusion maps (CBV, CBF, MTT, Tmax) generated by the bi-CNNs. Top row: bi-CNN maps; middle row: ground truth; bottom row: the difference between the bi-CNN maps and the ground truth. The estimated perfusion maps show good alignment with the ground truth. Perfusion abnormalities (i.e., hypoperfusion) can be visually detected in some estimated perfusion maps (red boxes).

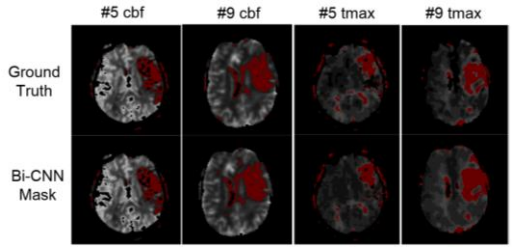

Fig. 5. Salvageable tissue masks (red) defined on the CBF and Tmax maps from the bi-CNNs and the ground truths. The bi-CNN generated masks have good alignments with the ground truth masks, showing their usability to detect the salvageable tissue. Note that the difference in contrast grayscale scale is caused by different range of perfusion parameter values.

E. Limitations

There are a few limitations in our work. First, the performance of the bi-CNNs, which is a machine learning approach different from standard deconvolution, is limited by the amount of available training data. With more cases, we can train larger networks with more epochs to learn the variability embodied by additional patients, which could potentially improve the performance. Second, we did not evaluate the bi-CNNs using digital phantoms [19], which is a more accurate source of ground truth. We plan to perform this evaluation with comparison to standard deconvolution techniques and other pattern recognition algorithms in future work. Third, we did not investigate the optimal patch size for the parameter estimation. More spatial context information may boost the performance of the voxel-wise estimation. Finally, using the current implementation of bi-CNNs to generate an estimated perfusion map requires more computational time than standard deconvolution (~5x slower). Batch and multi-GPU processing will be implemented to shorten the map generation time so that it is practical to apply the models clinically.

VI. Conclusions

In this work, we propose a novel approach for perfusion parameter estimation using a bi-input convolutional neural network. Our results show that the patch-based bi-CNN model is capable of estimating four perfusion parameters in stroke patients without using a standard deconvolution method (e.g., SVD). The estimated perfusion maps can be used to generate binary masks that are representative of the salvageable tissue. This model can potentially be extended to other disease domains in which perfusion imaging is used, such as cancer. Future work includes experimenting on a larger dataset and comparing bi-CNNs with other parameter estimation methods.

Acknowledgments

This research was supported by National Institutes of Health (NIH) grant R01 NS076534 and an NVIDIA Academic Hardware Grant. KVS acknowledges support from NIH NIGMS GM08042 and the UCLA-Caltech Medical Scientist Training Program.

References

[1] J. G. Merino and S. Warach, "Imaging of acute stroke," *Nat. Rev. Neurol.*, vol. 6, no. 10, pp. 560–571, 2010.

[2] A. Fieselmann, M. Kowarschik, A. Ganguly, J. Hornegger, and R. Fahrig, "Deconvolution-Based CT and MR Brain Perfusion Measurement : Theoretical Model Revisited and Practical Implementation Details," vol. 2011, 2011.

[3] I. K. Andersen, A. Szymkowiak, C. E. Rasmussen, L. G. Hanson, J. R. Marstrand, H. B. W. Larsson, and L. K. Hansen, "Perfusion quantification using Gaussian process deconvolution," *Magn. Reson. Med.*, vol. 48, no. 2, pp. 351–361, 2002.

[4] F. Calamante, D. G. Gadian, and A. Connelly, "Quantification of Bolus-Tracking MRI: Improved Characterization of the Tissue Residue Function Using Tikhonov Regularization," *Magn. Reson. Med.*, vol. 50, no. 6, pp. 1237–1247, 2003.

[5] T. Boutelier, K. Kudo, F. Pautot, and M. Sasaki, "Bayesian hemodynamic parameter estimation by bolus tracking perfusion weighted imaging," *IEEE Trans. Med. Imaging*, vol. 31, no. 7, pp. 1381–1395, 2012.

[6] M. Ibaraki, E. Shimosegawa, H. Toyoshima, K. Takahashi, S. Miura, and I. Kanno, "Tracer delay correction of cerebral blood flow with dynamic susceptibility contrast-enhanced MRI.," *J. Cereb. Blood Flow Metab.*, vol. 25, no. 3, pp. 378–90, 2005.

[7] O. Wu, L. Østergaard, R. M. Weisskoff, T. Benner, B. R. Rosen, and A. G. Sorensen, "Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix," *Magn. Reson. Med.*, vol. 50, no. 1, pp. 164–174, 2003.

[8] L. Y., B. Y., and H. G., "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] N. Stier, N. Vincent, D. Liebeskind, and F. Scalzo, "Deep learning of tissue fate features in acute ischemic stroke," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, 2015, pp. 1316–1321.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3361–3368, 2011.

[12] A. Karpathy and T. Leung, "Large-scale Video Classification with Convolutional Neural Networks," *Proc. 2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1725–1732, 2014.

[13] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *Comput. Vision–ECCV 2014*, vol. 8689, pp. 818–833, 2014.

[14] A. M. Smith, C. B. Grandin, T. Duprez, F. Mataigne, and G. Cosnard, "Whole brain quantitative CBF, CBV, and MTT measurements using MRI bolus tracking: implementation and application to data acquired from hyperacute stroke patients," *J. Magn. Reson. Imaging*, vol. 12, no. 3, pp. 400–410, 2000.

[15] J. M. Olivot, M. Mlynash, V. N. Thijs, S. Kemp, M. G. Lansberg, L. Wechsler, R. Bammer, M. P. Marks, and G. W. Albers, "Optimal tmax threshold for predicting penumbral tissue in acute stroke," *Stroke*, vol. 40, no. 2, pp. 469–475, 2009.

[16] F. Scalzo, Q. Hao, J. R. Alger, X. Hu, and D. S. Liebeskind, "Regional prediction of tissue fate in acute ischemic stroke," *Ann. Biomed. Eng.*, vol. 40, no. 10, pp. 2177–2187, 2012.

[17] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Advances in Neural Information Processing Systems*, 2014, pp. 1790–1798.

[18] R. Fang, T. Chen, and P. C. Sanelli, "Towards robust deconvolution of low-dose perfusion CT : Sparse perfusion deconvolution using online dictionary learning," *Med. Image Anal.*, vol. 17, no. 4, pp. 417–428, 2013.

[19] K. Kudo, S. Christensen, M. Sasaki, L. Østergaard, H. Shirato, K. Ogasawara, M. Wintermark, and S. Warach, "Accuracy and reliability assessment of CT and MR perfusion analysis software using a digital phantom," *Radiology*, vol. 267, no. 1, pp. 201–211, 2013.

[20] K. C. Ho, W. Speier, S. El-Saden, D. S. Liebeskind, J. L. Saver, A. A. T. Bui, and C. W. Arnold, "Predicting Discharge Mortality after Acute Ischemic Stroke Using Balanced Data," in *AMIA Annual Symposium Proceedings*, 2014, vol. 2014, p. 1787.

[21] A. Krizhevsky, Ii. Sulskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[22] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004.