



ELSEVIER

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Challenge report

AIFNet: Automatic vascular function estimation for perfusion analysis using deep learning

Ezequiel de la Rosa^{a,b,*}, Diana M. Sima^a, Bjoern Menze^b, Jan S. Kirschke^c, David Robben^{a,d,e}^a *icomatrix, Leuven, Belgium*^b *Department of Computer Science, Technical University of Munich, Munich, Germany*^c *Neuroradiology, School of Medicine, Technical University of Munich, Munich, Germany*^d *Medical Imaging Research Center (MIRC), KU Leuven, Leuven, Belgium*^e *Medical Image Computing (MIC), ESAT-PSI, Department of Electrical Engineering, KU Leuven, Leuven, Belgium*

ARTICLE INFO

Article history:

Received 23 September 2020

Revised 25 June 2021

Accepted 4 August 2021

Available online 6 August 2021

Keywords:

Ischemic stroke

Perfusion imaging

Arterial input function

Deep learning

ABSTRACT

Perfusion imaging is crucial in acute ischemic stroke for quantifying the salvageable *penumbra* and irreversibly damaged *core* lesions. As such, it helps clinicians to decide on the optimal reperfusion treatment. In perfusion CT imaging, deconvolution methods are used to obtain clinically interpretable perfusion parameters that allow identifying brain tissue abnormalities. Deconvolution methods require the selection of two reference vascular functions as inputs to the model: the arterial input function (AIF) and the venous output function, with the AIF as the most critical model input. When manually performed, the vascular function selection is time demanding, suffers from poor reproducibility and is subject to the professionals' experience. This leads to potentially unreliable quantification of the penumbra and core lesions and, hence, might harm the treatment decision process. In this work we automatize the perfusion analysis with AIFNet, a fully automatic and end-to-end trainable deep learning approach for estimating the vascular functions. Unlike previous methods using clustering or segmentation techniques to select vascular voxels, AIFNet is directly optimized at the vascular function estimation, which allows to better recognise the time-curve profiles. Validation on the public ISLES18 stroke database shows that AIFNet almost reaches inter-rater performance for the vascular function estimation and, subsequently, for the parameter maps and core lesion quantification obtained through deconvolution. We conclude that AIFNet has potential for clinical transfer and could be incorporated in perfusion deconvolution software.

© 2021 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Stroke is currently the second leading cause of mortality and the third leading cause of disability worldwide (Stroke Unit Trialists Collaboration, 2013). In physio-pathological terms, it is defined as an 'acute neurologic dysfunction of vascular origin with sudden (within seconds) or at least rapid (within hours) occurrence of symptoms and signs corresponding to involvement of focal areas in the brain' (Force, 1989). Two main types of the disease can be recognised: ischemic and hemorrhagic, representing 85% and 15% of total cases respectively (Hinkle and Guanci, 2007). We focus on the ischemic case, where there is a shortage in the blood supply to the brain tissue, cutting the provision of oxygen and glucose. During the ischemic event, brain tissue might become necrotic (i.e., cells

are dead and the tissue is irreversibly damaged, known as *core*) or in a hypo-perfused but salvageable state (i.e., tissue is at risk but could return to a healthy condition, known as *penumbra*).

1.1. Perfusion CT in acute ischemic stroke

Acute ischemic stroke therapies rely on reperfusion techniques, where the main goal is to reestablish the blood flow supply in the affected territories by thrombolysis or thrombectomy. Identifying which patients might benefit from these treatments is critical for clinical decision making (Campbell and Parsons, 2018; Albers et al., 2016). To this end, assessment and quantification of the core and penumbra tissues are required. In the acute scenario, computer tomography (CT) is the most widely used imaging technique, where perfusion CT (CTP) enables the determination of the core and penumbra areas. An iodinated contrast agent is intra-venously injected in the patient for 7–10 s, and continuous CT acquisition

* Corresponding author.

E-mail address: ezequiel.delarosa@icomatrix.com (E. de la Rosa).

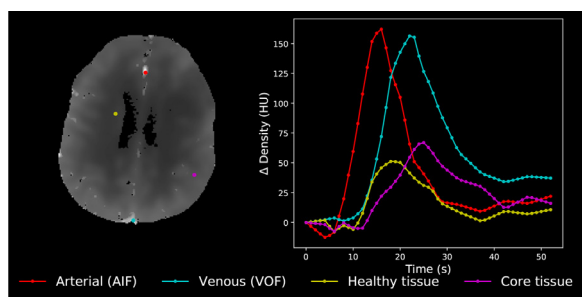


Fig. 1. Contrast enhancement curves for different brain tissues. Left: a perfusion CT example from ISLES18. Right: Corresponding time curves at the indicated locations. Healthy and diseased brain areas have been identified through diffusion weighted imaging. The healthy and core time-curves are scaled by a factor of six for visualization. HU: Hounsfield units.

is followed for around 50 s (Fieselmann et al., 2011). As such, 4D data is generated, resulting in a brain volume imaged during the agent passage through the brain vessels and parenchyma. The process for evaluating brain tissue status is performed by firstly obtaining parameter maps from the CTP time series and by later applying a tissue discrimination rule (mainly, thresholding). Typical maps include cerebral blood flow (CBF), cerebral blood volume (CBV), time to peak (TTP) and time to the maximum of the residue function (Tmax). It is worth saying that there is no *gold* standard for quantifying perfusion metrics (Lorenz et al., 2006), and all methods found in literature provide merely non-exact solutions. Experimental studies have shown that CBV and CBF discriminate ischemic and oligemic tissue with 90.6% and 93.3% sensitivity and specificity, respectively, when using histological measurements as ground truth (Murphy et al., 2007). The most widely used methods for CTP parameter map estimation are based on deconvolution (Konstas et al., 2009), which provides a solution to the indicator dilution theory described by:

$$c_{tissue}(t) = c_{art}(t) \otimes h(t) \quad (1)$$

where $c_{tissue}(t)$ represents the CTP contrast enhancement in a voxel of tissue, $c_{art}(t)$ is the contrast enhancement in the arteries (known as arterial input function, from now on ‘AIF’), $h(t)$ is the flow-scaled residue function and \otimes represents the convolution operator. The delay-invariant singular value decomposition deconvolution is the preferred technique for algebraically solving Eq. (1) and it is widely implemented in software packages (Fieselmann et al., 2011; Konstas et al., 2009; Kosior and Frayne, 2007; Kudo et al., 2010; Vagal et al., 2019). The method has been extensively validated in clinical practice, showing better performance compared to similar techniques (Konstas et al., 2009; Fieselmann et al., 2011) like the maximum slope approach (Konstas et al., 2009; Klotz and König, 1999), non-delay invariant deconvolution (Østergaard et al., 1996a,b), etc. Deconvolution methods require as input to the algorithm the CTP series and two vascular functions: the AIF and the venous output function (VOF). These vascular functions are reference time-curves representing the contrast concentration inlet and outlet to the tissue under consideration $c_{tissue}(t)$. Fig. 1 shows an example case of vascular functions (i.e. AIF and VOF) and contrast enhancement curves for healthy and core tissue areas. In clinical practice the AIF and VOF are generally selected by a radiologist, a time demanding and highly variable process that implies selecting in the CTP series the optimal candidate voxels. Frequently, a single voxel per vascular function is selected, which leads to low SNR curves. Voxel selection is, moreover, subject to the professionals’ training and experience, which not only introduces human bias (Lorenz et al., 2006) but it may also affect CBF maps depending which side of the brain the AIF is chosen from (Wu et al., 2003; Thijs et al., 2004). The AIF is so

critical for generating accurate maps that very small changes in its shape and/or location may produce a profound effect over the generated maps (Mlynash et al., 2005; Mouridsen et al., 2006). Besides, given the acute context of the disease, a fast voxel selection has to be performed. It has been shown that for every 30-minute delay in reperfusion, the probability of good outcome decreases by 20% (Khatri et al., 2014). Given these limitations, automatic, fast and reproducible core and penumbra quantification are highly desired.

1.2. Automatic core and penumbra segmentation

Automatic machine and deep learning approaches for core and penumbra quantification have been explored in two ways: 1) by direct parameter maps estimation and 2) by direct lesions segmentation. On one hand, automatic parameters maps estimation (i.e., bypassing deconvolution) was explored in (McKinley et al., 2018; Meier et al., 2019; Robben and Suetens, 2018; Ulas et al., 2018a,b). However, the main drawback of these methods is the fact that *silver* standard maps obtained through deconvolution or other methods (e.g. compartmental models in the case of perfusion MRI) are approximated. Note that in these approaches there is also an AIF assumption behind the parameter maps ground truth. As such, these methods do not improve the perfusion gold standard, but aim to reproduce it with a different model. On the other hand, direct lesion segmentation approaches use native CTP data with or without perfusion maps as model inputs. Thus, the neural networks are used for finding a non-linear transformation from CTP and/or CBF, CBV, MTT and Tmax that estimates brain lesions. For instance, in (Bertels et al., 2018) and (Robben et al., 2020) direct lesion segmentation is conducted by only using CTP images. While the former work exploits contralateral brain information into a U-Net based architecture, the latter work includes metadata and vascular functions into a multiresolution DeepMedic-based (Kamnitsas et al., 2017) architecture. Other works include parameter maps obtained through deconvolution as inputs to the model (Clèrigues et al., 2019; Abulnaga and Rubin, 2018; Song and Huang, 2018; Wang et al., 2020). Similarly as in Bertels et al. (2018), Clèrigues et al. (2019) exploit brain symmetry information with U-nets. Song and Huang (2018) and Wang et al. (2020) propose, instead, to synthesize pseudo diffusion weighted imaging (DWI) data to improve core lesion segmentation. While deep learning based approaches showed good overall performance, their main limitation is the poor model’s explainability and lack of quality control. Since these fully ‘black-box’ methods do not allow AIF or perfusion maps inspection, they preclude physicians to recompute the parameter maps with a manually corrected AIF in clinically or technically challenging cases. As such, the clinical transferability potential of these models is limited. In this work we aim to automatize, instead, the well validated deconvolution process by the automatic selection of vascular functions. In this way, we avoid approximating parameters that can be directly estimated through a physical model while also preserving explainability and quality control in clinical settings.

1.3. Automatic vascular function selection

Automatic vascular function selection has been explored for perfusion MRI in (Murase et al., 2001; Mouridsen et al., 2006; Peruzzo et al., 2011; Shi et al., 2014; Shi and Malik, 2000; Yin et al., 2015; Fan et al., 2019; Winder et al., 2020). These methods mainly rely on clustering techniques, where fuzzy c-means (Murase et al., 2001), K-means (Mouridsen et al., 2006), hierarchical clustering (Peruzzo et al., 2011), gamma-variates based clustering (Rausch et al., 2000) and affine propagation clustering (Shi et al., 2014) were explored. Heuristic approaches have also

been traditionally used, where some rules are defined for finding the best-matching curve, such as in (Mlynash et al., 2005; Rempp et al., 1994). Other techniques use normalized cuts (Shi and Malik, 2000; Yin et al., 2015) and independent component analysis (Calamante et al., 2004). Moreover, vascular function estimation using deep neural networks can be conducted through segmentation approaches aiming to detect arterial voxels candidates. Fan et al. (2019) proposed a deep learning segmentation approach for delineating AIF candidates in perfusion MRI. The method uses two independently optimized 3D CNNs for conducting arterial tissue segmentation: one extracting spatial information in the $x - y - z$ axis, and another one extracting temporal-information in the $x - y - t$ axis (with t representing the temporal domain). Afterwards, the networks' results are merged using a late-fusion support vector machine. More recently, Winder et al. (2020) proposed a binary output CNN for classifying arterial vs non-arterial voxels in CTP and perfusion MRI. The AIF is then estimated by geometrically averaging the most probable arterial voxels. Though segmentation or classification methods can identify potential good curves, they have some limitations: i) They require complete manual annotation of all "good-looking" voxel curves, which is very time demanding and ii) They could not always guarantee optimal AIF curve selection (for a possible definition of optimal AIF selection, see Methods 3.1.2) since the algorithms are mainly optimized to perform selection based on spatial information rather than on time profiles. Segmentation methods may lead, for instance, to the undesired selection of noise-corrupted, low contrast enhanced or time delayed AIFs, which introduce errors in the deconvolution algorithms. In CTP imaging, however, vascular function selection is under-explored. Excepting the work of Winder et al. (2020) the few existing methods are mostly private and patented. Besides, most of the methods developed for perfusion MRI have not been validated for CTP. Despite perfusion CT and perfusion MRI having common working points, there are still technical differences that may affect the automatic selection of CTP vascular functions (such as lower tissue-density contrasts and lower SNR of CT compared to MRI). Moreover, additional challenges in CTP include overlapping density distribution of bone, artifacts and calcifications with the iodine contrast.

In this work we propose AIFNet, an end-to-end supervised convolutional neural network devised for estimating vascular functions (i.e. AIF and VOF) in perfusion imaging. The model is easy to train and deploy given the minimal data annotation required, which can be as little as a single voxel per vascular function. AIFNet receives 4D CTP series as input and generates as output i) the estimated AIF and VOF curves and ii) a voxel-wise, interpretable probability map representing the voxelwise contribution to the estimated vascular signal. Unlike other approaches, AIFNet is optimized at a vascular function level, which helps the network to better learn the time-curve profiles. The method preserves clinical interpretability and also enables quality control of the selected AIF/VOF brain vasculature, thus enhancing its clinical transferability potential. Through an extensive analysis at signal, parameter maps and lesion quantification levels, we show that our method performs almost as good as manual raters on the open ISLES18 acute stroke database.

2. Methods

2.1. Function estimation with deep learning

AIFNet is a fully end-to-end deep learning approach for vascular function estimation. It works by estimating a 3D probabilistic volume that represents the voxelwise contribution to the vascular signal. The advantage of having an averaged curve using multiple voxels lies on the higher function's SNR as well as on the method

robustness. The network receives as input the 4D perfusion series $x(t)$ and outputs the predicted arterial and venous functions as $\hat{y}(t) = AIF_{Net}(x(t))$, being $x(t) = \{x_t; t = 1, 2, \dots, T\}$ with x_t representing the sampled time point volumes of dimension $M \times N \times Q$. We want to find for the considered volume, its corresponding vascular functions (AIF and VOF, for simplicity not differentiated in the notation) represented by $\hat{y}(t) = \{\hat{y}_t; t = 1, 2, \dots, T\}$, where \hat{y}_t is the estimated signal at time t (in Hounsfield units). For finding $\hat{y}(t)$, we represent each time point \hat{y}_t as a weighted average of all voxels of the volume x_t at that t time point as:

$$\hat{y}_t = \sum_{q=1}^Q \sum_{n=1}^N \sum_{m=1}^M x_t(m, n, q) * P_{vol}(m, n, q) \quad (2)$$

where P_{vol} is the 3D probabilistic volume containing the voxelwise contribution to the vascular function and fulfilling:

$$\sum_{q=1}^Q \sum_{n=1}^N \sum_{m=1}^M P_{vol}(m, n, q) = 1 \quad (3)$$

Our problem is hence confined to finding P_{vol} . With this aim, AIFNet receives as input native CTP series, and generates as outputs P_{vol} and its associated vascular function. To find \hat{y}_t it is important to optimize the similarity of the shape rather than the amplitude. This is due to two facts: 1) the absolute contrast values of the AIF can be disregarded, since given the high partial volume effect in the arteries, the AIF is later recalibrated with the VOF (Fiesemann et al., 2011) (also see Section 2.4.1) and 2) a suboptimal deconvolution might occur by selecting delayed input functions. The penalty in the time domain is introduced by using Pearson's correlation as loss function as follows:

$$\mathcal{L}(y(t), \hat{y}(t)) = - \frac{\sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^T (y_t - \bar{y})^2} \sqrt{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2}} \quad (4)$$

where $y(t)$ and $\hat{y}(t)$ are the ground truth and predicted vascular functions with respective mean values $\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$ and $\bar{\hat{y}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$.

2.2. Architecture

AIFNet architecture is shown in Fig. 2. It uses 3D convolutional layers for volumetric feature extraction, which are finally translated into a probabilistic volume through a 3D softmax operation. After finding P_{vol} , a voxelwise multiplication and 3D average pooling blocks are used for obtaining $\hat{y}(t)$, by means of Eq. (2). Each convolutional layer $L_k = \{k = 1, 2, \dots, K\}$ has 2^{3+k} filters with a $3 \times 3 \times 3$ kernel with exception of L_1 , which uses a $3 \times 3 \times 1$ one with the aim of compensating the lower image resolution along the z -axis. The CTP time points are incorporated as channel information into the network. A fixed number of T time points are used for all scans. In our experiments we use a T equal to the smallest number of time points found among all scans. Rectified linear units are used as activation functions (Krizhevsky et al., 2012). For mapping the convolutional layers to a single probabilistic volume, we add an extra convolution block (L_{out}) with only one filter in between L_K and the softmax operator.

2.3. Training phase

The network is optimized using stochastic gradient descent with momentum. A batch size of one sample is used. Regularization of the model is reached using a perfusion-specific data augmentation approach.

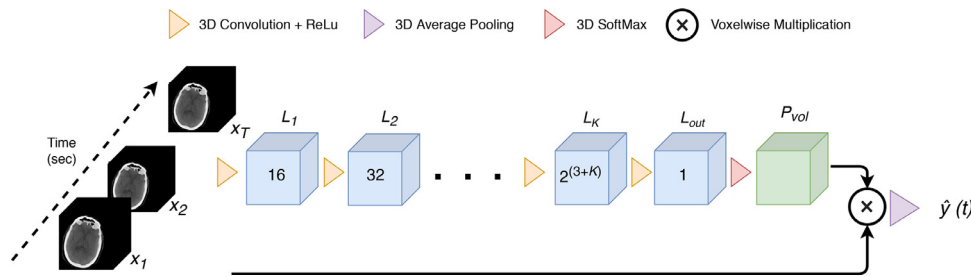


Fig. 2. AIFNet architecture. The CTP time points x_t ($t = 1, 2, \dots, T$) are incorporated as channels in the network. All convolutional layers use $3 \times 3 \times 3$ kernels except L_1 , which uses $3 \times 3 \times 1$. L_k is the k -th convolutional layer (with $k = 1, 2, \dots, K$). Inside each feature block the number of channels used is indicated. P_{vol} is the probabilistic volume. The 3D average pooling block averages the volumetric information along the $x - y - z$ axes, such that the predicted vascular function $\hat{y}(t)$ is a 1D vector of length T .

2.3.1. Perfusion specific data augmentation

We adapt the data augmentation strategy proposed in (Robben and Suetens, 2018) for working at an image level. Two perfusion specific phenomena are modelled: *i*) the variability of the contrast bolus arrival, which depends on the injection protocol and the patient's cardiovascular system and *ii*) the variability of the curve's peak-to-baseline (PTB) values, which depends on the iodine concentration in the contrast agent. Bolus arrival changes are simulated by randomly shifting the time attenuation curves, for which the first or last CT volumes are replicated (late or early simulated arrivals respectively). On the other hand, curve PTB changes are simulated in a three-step approach. Firstly, the pre-contrast averaged volume is subtracted from the perfusion series. Secondly, a random scaling is applied. Thirdly, the pre-contrast volume is re-added to the perfusion series. Uniform distributions are used for simulating the random time shifts and the random PTB scaling.

2.4. Testing phase

In the testing scenario, vascular function predictions are obtained by feeding the parametrized AIFNet model with the unseen CTP scans. The voxelwise multiplication and 3D average pooling blocks of AIFNet are performed over the full-length CTP perfusion series, with the aim of obtaining vascular function predictions that preserve the same number of time points as the native CTP scan. For VOF a signal recalibration step is also applied, as detailed below.

2.4.1. VOF signal recalibration

Our multiple signal averaging approach has the disadvantage of underestimating the VOF peaks. Since the VOF's role in deconvolution-based perfusion analysis is to compensate for partial volume effect in the AIF by its recalibration, it is important that its PTB matches the same amplitude as single CTP candidate voxels. Ideally, a suitable VOF curve has the highest PTB value among all venous voxel candidates. Therefore, we use a probabilistic volume that encodes voxelwise contribution to the function estimation. Firstly, we generate a 3D volume encoding the voxelwise PTB values. Secondly, we scale this volume with P_{vol} in order to obtain probabilistic-weighted PTB values. The VOF is finally recalibrated with the maximal value found in the weighted PTB distribution. We prefer using weighted PTB instead of only considering P_{vol} , since the highest probability voxel of P_{vol} might have a low PTB, thus leading to an underestimation of the VOF PTB value.

3. Experiments

3.1. Data

3.1.1. ISLES18

The large public multi-center and multi-scanner ISLES18 dataset is used for our experiments (Maier et al., 2017; Kistler et al., 2013; Cereda et al., 2016). It consists of 156 CTP acquisitions acquired from 103 acute stroke patients from three US centers and one Australian center. In the ISLES challenge, data is split into a train (94 CTP volumes scanned from 63 patients) and a test (62 CTP volumes scanned from 40 patients) sets. The mismatch between patients and scans is due to the limited field of view of some scanners, which leads to two independent CTP acquisitions from different brain regions in some cases. We have directly accessed the clean and preprocessed data through the ISLES challenge site (<http://www.isles-challenge.org/>). For each acquisition, CTP and DWI data were performed within 3 hours of each other. The open database provides CTP scans for the whole dataset and infarct core lesion masks (delineated in DWI images) for the training set only. Subjects having more than 50% of the DWI lesion with normal perfusion at the moment of the CTP acquisition were excluded, as well as those subjects with bad quality of the baseline CTP data and/or with inappropriate image coregistration due to distortions (Cereda et al., 2016). CTP volumes have been motion corrected and coregistered for matching the DWI lesion masks. Finally, scans have been spatio-temporally resampled (with a 256×256 dimension matrix and with a temporal resolution of one volume per second). For a more detailed description of this database the reader is referred to (Cereda et al., 2016).

3.1.2. Vascular function annotation

All training and testing scans are in-house annotated by two independent raters (DR & EdIR). A single *global* AIF and VOF per scan is selected (i.e., functions are measured from a major artery/vein and used as global inputs for the tissue in the whole brain (Calamante, 2013)), where the following AIF time attenuation curves are preferred: *i*) contralateral voxels to the affected area (rather than ipsilateral ones) (Kealey et al., 2004; Calamante, 2013), *ii*) Early bolus arrival AIF curves with a large and narrow peak enhancement (Calamante, 2013) *iii*) Curves with high contrast-to-noise ratio and, ideally, less affected by partial volume effect (qualitatively assessed) (Calamante, 2013). The *best* voxel candidate (following the just mentioned criterion) among the anterior cerebral arteries, middle cerebral arteries, internal carotid arteries or the basilar artery are chosen as AIF. On the other hand, VOF curves are located in the superior sagittal, transverse or sigmoid sinuses, which are large vessels less affected by partial volume effect than other vessels. All vascular function annotations are provided as supplementary material.

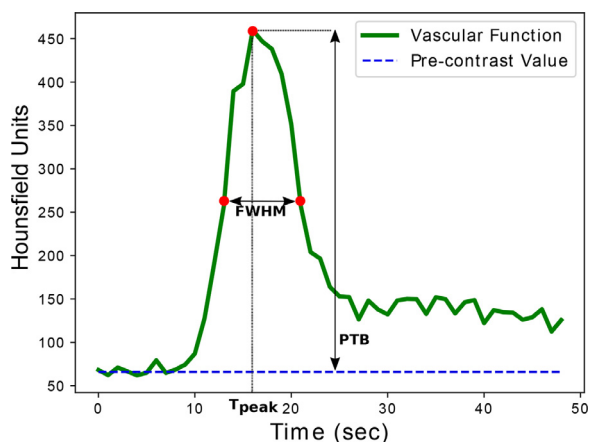


Fig. 3. Vascular function metrics. FWHM: Full width at half maximum; PTB: Peak to baseline; T_{peak} : Time at which the curve peak occurs.

3.2. Performance assessment

In order to evaluate the performance of AIFNet, we conduct a 5-fold (train 70%, validation 10%, test 20%) cross-validation experiment using the annotations of rater #1. All training and testing cases of the ISLES18 database are used in this experiment. In each fold, the train set is used to parametrize the network, the validation set to apply an early-stopping criterion (with the aim of avoiding overfitting) and the test set is independently used for predicting unseen cases. For an in-depth evaluation of the proposed method, results are assessed at a signal, parametric map and lesion quantification level.

3.2.1. Vascular function

Since there is no ground truth for the vascular functions we compare the predictions $\hat{y}(t)$ against the manual annotations $y(t)$ of both the raters (from now on, we specifically refer to our method as $\hat{y}_{\text{AIFNet}}(t)$, to rater #1 as $y_{r1}(t)$ and to rater #2 as $y_{r2}(t)$). The agreement between $y(t)$ and $\hat{y}_{\text{AIFNet}}(t)$ is only computed over the time domain since the AIF absolute contrast values rely on the VOF rescaling (see sections 2.3 and 2.4.1). To this end, we measure the time at which the curve peak occurs (namely T_{peak}), which should indicate potential time shifts of the predictions with respect to ground truth. Moreover, as a measure of the function's width, we quantify the full-width at half-maximum (FWHM) interval. FWHM points are preferred to the curve's onset/offset since these are more difficult to measure and in many cases the offset point is missing. Besides, for evaluating the VOF recalibration strategy, we assess the whole signal agreement using mean squared error and the measured PTB values. All signal metrics are illustrated in Fig. 3.

3.2.2. Parameter maps and lesion quantification

Parameter maps (CBF, CBV, MTT, Tmax) are computed using the well known time-delay invariant singular value decomposition deconvolution. The method is the most commonly found in (clinical) software. Since deconvolution is a mathematically ill-conditioned problem, regularization techniques are necessary. We use Tikhonov regularization over the singular values under a Volterra discretization scheme (Sourbron et al., 2007). Absolute and relative parameter maps are computed, where the relative ones are calculated by the voxelwise normalization of the absolute ones with the mean control tissue region value. Control tissue is defined as the region with normal perfusion (i.e., $T_{\text{max}} < 6$ s (Lin et al., 2016)).

To understand the impact of the vascular functions on the perfusion metrics we compare the parameter maps (obtained through

deconvolution) between the automatically and manually annotated vascular functions. The same deconvolution strategy is adopted in both cases. In this way we are sure that variations are only due to the vascular functions. The assessment of the method in terms of lesion quantification is conducted by comparing i) the hypoperfused and core masks obtained from CTP with automatically annotated vascular functions against the ones obtained using the curves labeled by the two experts and ii) by comparing the obtained core masks against the ISLES18 DWI masks. Note that this latter comparison is only done for the ISLES18 training set, since DWI lesion masks are not available for the test set. Hypoperfused tissue was defined as brain tissue with $T_{\text{max}} > 6$ s (Lin et al., 2016) and core tissue was defined within the hypoperfused area as rCBF < 38% (a cutoff previously found to be optimal by Cereda et al. (2016)).

3.2.3. Comparison with other methods

Besides the comparison with the two manual raters, our method is compared with an unsupervised AIF selection approach (Mouridsen et al., 2006) and with two similar CNNs: a regression and a segmentation network, both of them modified versions of AIFNet. For having comparable deep learning approaches, we keep the network's architecture and configuration as close as possible to AIFNet. The same perfusion-specific data augmentation of Section 2.3.1 is used for both the segmentation and regression networks and a unitary batch size is used. The networks are tested following the same 5-fold cross-validation experiments used for AIFNet by assuring that for all the models the same train-validation-test splits are preserved.

Unsupervised clustering AIF selection We use an in-house reimplementation of the unsupervised clustering approach of (Mouridsen et al., 2006). Firstly, non-arterial voxels are discarded using the area under the curve (with a threshold set at the 90th percentile of candidates) and the roughness (defined as $\int_0^T [C''(t)]^2 dt$, with a threshold set at the 25% most irregular candidates) of each voxels' time curve $C(t)$. Secondly, a two-steps K-means (with 5 clusters) is performed for separating other tissues and venous voxels from the arterial cluster. In each K-means iteration, the arterial cluster is selected as the one having lowest first moment of its mean curve. The final AIF is the mean curve with lowest first moment obtained after applying K-means twice.

Regression AIFNet The regression CNN has been introduced by our group in (de la Rosa et al., 2020). It consists of a 3D + 2D neural network equipped with six convolutional layers with average pooling and with two fully connected layers at the end. The last fully connected layer is a 1D vector with same number of neurons as time points in the perfusion CTP and represents the vascular function prediction $\hat{y}_{\text{Reg}}(t)$. The 3D to 2D data transformation in the network is conducted by squeezing the z-axis information by means of average pooling. Homogenizing the z-axis dimension is required for dealing with the variable CTP coverage (varying between 2 and 16 slices per scan). The 3D convolution kernels have dimension $3 \times 3 \times 3$ and the 2D convolution ones have dimension 3×3 . Unlike the original work where optimization was carried out with a segmentation loss for the core tissue, in this work the Pearson's correlation coefficient is preserved as loss function. The network is optimized with RMSprop gradient descent (Hinton et al., 2012).

Segmentation AIFNet This network is fed with AIF binary masks as ground truth. It is similar to AIFNet by preserving the whole architecture except the voxelwise multiplication and 3D average pooling blocks (see Fig. 2). Besides, the last convolutional block (L_{out}) has two kernels followed by a softmax operation for conducting background and foreground segmentation. For compensating the large class imbalance, this network is trained using weighted categorical cross-entropy as loss function and is optimized using stochastic gradient descent with momentum. The AIF is then esti-

Table 1

AIF agreement among methods and rater #1 for the AIF signals as a whole, as well as for their T_{peak} and FWHM parameters. Pearson's correlation coefficients are computed between pairs of AIF signals; the mean (standard deviation) and 5th - 95th percentile interval are provided. For the T_{peak} and FWHM parameters, Pearson's correlation coefficients and errors (in seconds) are reported across all scans as mean (standard deviation). $y_{r1}(t)$, $y_{r2}(t)$: AIF annotated by raters #1 and #2; $\hat{y}_{Kmeans}(t)$, $\hat{y}_{Seg}(t)$, $\hat{y}_{Reg}(t)$, $\hat{y}_{AIFNet}(t)$: AIF estimated with the K-means approach (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively; r : Pearson's correlation coefficient; T_{peak} : time at which the peak of the curve occurs; FWHM: full-width at half-maximum. P5th: 5th percentile; P95th: 95th percentile. Paired significance tests are performed between AIFNet and the other approaches. *: p-value < 0.05; †: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

		Signal		T_{peak}		FWHM	
		r	r (P5 th , P95 th)	r	Error [s]	r	Error [s]
Inter-rater		0.971 (0.075)†	(0.883, 1)	0.964	-0.14 (1.29)†	0.902	-0.08 (1.74)†
$y_{r1}(t)$ vs	$\hat{y}_{Kmeans}(t)$	0.610 (0.315)†	(-0.101, 0.955)	0.678	-5.37 (5.55)†	0.369	-1.95 (6.27)†
	$\hat{y}_{Seg}(t)$	0.677 (0.15)†	(0.393, 0.897)	0.851	-3.95 (3.43)†	0.587	-4.43 (5.03)†
	$\hat{y}_{Reg}(t)$	0.837 (0.260)†	(0.563, 0.986)	0.740	-0.30 (3.30)	0.419	-2.86 (3.65)†
	$\hat{y}_{AIFNet}(t)$	0.965 (0.05)	(0.838, 0.997)	0.940	-0.55 (1.75)	0.854	-0.89 (2.14)

mated as the average function among the top ranked voxels, such that the AIF Pearson's correlation is maximized.

3.2.4. Statistical analysis

For evaluating the vascular signals, Pearson's correlation coefficients are computed between pairs of signals and across all scans for the considered metrics (i.e. T_{peak} , FWHM and PTB). Mean, standard deviation and (5th, 95th) percentiles are provided. Additionally, to assess a potential bias of the different metrics, we compute the mean and standard deviation of the errors. The assessment of the parameter maps is performed using Pearson's correlation coefficients computed per scan within the brain masks (excluding background, skull, ventricles and vessels). Hypoperfused and core tissue segmentations are evaluated by comparing the CTP masks obtained by the different methods with the CTP masks obtained by the experts. Additionally, for the core tissue we compare the experts and the different methods CTP masks with the ground truth DWI masks from ISLES18. The mean volume error and the mean absolute volume error are used for evaluating lesion volumetric agreement with the ground truth. The Dice coefficient is used as a general segmentation performance metric. In all cases, paired t-tests are performed after visual inspection of the data distributions. Under the presence of non-normal distributions, outliers, or heteroskedasticity, a paired Wilcoxon-test is preferred. The significance level is set in all cases to $\alpha = 0.05$.

4. Results and discussion

All models are trained on a machine with a Tesla K80 Nvidia GPU (12 Gb dedicated), with 64 gb RAM and an Intel Xeon E5-2686 v4 multiprocessor. The training stage takes ~11 hours for an AIF/VOF model. Manual annotations take between 2 and 4 minutes for both functions per scan, depending on the number of slices of the volume. On the other hand, predictions take ~6 seconds per each vascular function per scan.

4.1. Signal agreement

4.1.1. AIF

Table 1 shows a summary of the different methods' performance compared to rater #1. Likewise, the comparison with rater #2 is shown in Table S1 (supplementary materials). The automatic predictions of AIFNet obtain high agreement with both raters in all the metrics considered. There is an overall better agreement with $y_{r2}(t)$, even when the network is trained using the $y_{r1}(t)$ annotations, suggesting good generalization at inter-rater level. When the entire vascular signal is evaluated, the method obtains Pearson's r values reaching the raters range. A slightly lower 5th percentile is observed in the agreement between $\hat{y}_{AIFNet}(t)$ and $y_{r1}(t)$

when compared with the inter-rater agreement. This discordance is, however, not found when comparing $\hat{y}_{AIFNet}(t)$ with $y_{r2}(t)$, which obtains fully overlapping ranges with the inter-rater performance. The 95th percentile obtained between AIFNet and the raters is, as expected, close to $r = 1$ but never reaching perfect agreement, due to the weighted multivoxel selection strategy proposed.

When the method performance is assessed in terms of T_{peak} , a high correlation with the manual annotations is found. It can be observed from the inter-rater comparison that the T_{peak} annotations of $y_{r2}(t)$ are slightly delayed when compared with the ones of $y_{r1}(t)$. The AIF functions that AIFNet selects are on average ~ 0.5 seconds delayed when compared with the raters. This temporal trend toward delayed events explains the slight overall lower agreement between AIFNet and both raters. Similarly, the agreement that is obtained for the FWHM between AIFNet and the raters is slightly lower than the inter-raters level. The predicted FWHM windows are on average ~ 1 second longer than the manual ones. These time differences found in T_{peak} and FWHM with our method are below the temporal CTP resolution (one frame, the minimal possible). The main reason behind these differences is the flip side of the coin of the multivoxel selection strategy. Thus, vascular function estimation based on multiple voxels could not always provide the *earliest* bolus arrival with the *highest* and *narrowest* curves, but averaged values over the activated voxels. Selecting vascular functions with these characteristics is, hence, not always fully possible with our strategy, since generally a single or just a few voxels fulfill these requirements for AIF.

The comparison of the different methods shows that AIFNet has a much better agreement with the raters than the other approaches. While the segmentation CNN slightly outperforms the K-means method, the regression CNN outperforms both the segmentation CNN and the K-means approach (Table 1). The regression CNN not only correlates better with the raters at signal level but also localized with less delay T_{peak} than these two other methods. There are no statistically significant differences in the T_{peak} errors of the regression CNN and the ones of AIFNet. An explanation to this observation could be in the optimized loss function: the regression network, likewise AIFNet, is optimized at the predicted time-curve level instead of at the image spatial level (which is the case for the segmentation CNN). Another observation is that the segmentation CNN (which comprises almost the same architecture as AIFNet but optimized with a segmentation loss) obtained a much worse performance than our proposal. The segmentation CNN fails in localizing properly the AIF peaks, and provides delayed and much wider curves. These results suggest that the used Pearson loss function enhances the task performance by i) mostly activating arterial voxels with good AIF curves and by ii) selectively discarding suboptimal arterial voxels whose AIFs are delayed, highly noise-corrupted or with poor contrast enhancement.

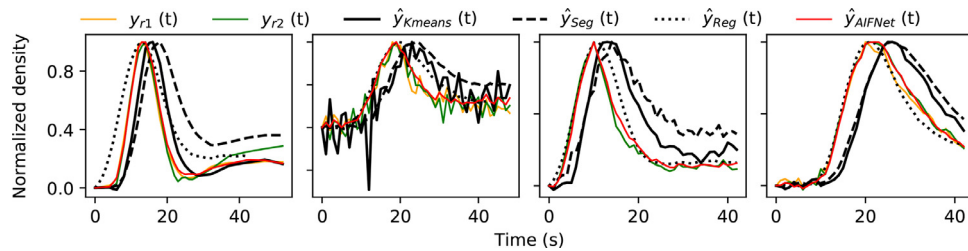


Fig. 4. Predicted arterial input functions (AIF) on diverse quality ISLES18 scans. First image: a noise-free scan. Second image: a head-motion corrupted scan. Third image: an early bolus arrival scan (the pre-contrast increase signal is missing). Fourth image: a truncated perfusion scan (the curve tale is missing). $y_{r1}(t)$, $y_{r2}(t)$: AIF annotated by raters 1 and 2; $\hat{y}_{Kmeans}(t)$, $\hat{y}_{Seg}(t)$, $\hat{y}_{Reg}(t)$, $\hat{y}_{AIFNet}(t)$: AIF estimated with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively.

Table 2

VOF agreement between AIFNet and both the raters for the VOF signals as a whole, as well as for their T_{peak} , FWHM and PTB parameters. Pearson's correlation coefficients are computed between pairs of VOF signals; the mean (standard deviation) and 5th - 95th percentile interval are provided. For the T_{peak} , FWHM and PTB parameters, Pearson's correlation coefficients and errors (in seconds for T_{peak} and FWHM and in Hounsfield units for PTB) are reported across all scans as mean (standard deviation). $y_{r1}(t) - y_{r2}(t)$: annotated VOF signals by raters #1 and #2; $\hat{y}_{AIFNet}(t)$: predicted VOF signals with AIFNet; r: Pearson's correlation coefficient; MSE: Mean squared error; HU: Hounsfield units; T_{peak} : time at which the peak of the curve occurs; FWHM: full-width at half-maximum; PTB: Peak-to-baseline; P5th: 5th percentile; P95th: 95th percentile. A paired significance test is performed between the inter-rater results and the AIFNet-raters ones. *: p-value < 0.05; †: p-value < 0.01.

		Inter-rater	$y_{r1}(t)$ vs $\hat{y}_{AIFNet}(t)$	$y_{r2}(t)$ vs $\hat{y}_{AIFNet}(t)$
Signal	r	0.985 (0.047)	0.981 (0.069)	0.983 (0.051)
	r (P5 th , P95 th)	(0.944, 1)	(0.914, 0.999)	(0.925, 0.999)
	MSE [HU]	1424 (3622)	1235 (2623)	1558 (3740)
	MSE (P5 th , P95 th)	(0, 7148)	(17, 7024)	(25, 8213)
T_{peak}	r	0.980	0.955	0.963
	Error [s]	0.27 (1.14)	-0.07 (1.69) [†]	-0.33 (1.51) [†]
FWHM	r	0.829	0.827	0.911
	Error [s]	0.12 (2.28)	-0.04 (2.42)	-0.15 (1.74)
PTB	r	0.921	0.953	0.919
	Error [HU]	11 (55)	9 (44)	-2 (58)

In Fig. 4, AIF predictions with the different methods over different quality scans are shown. As it can be seen, AIFNet closely follows the manual rater annotations even under challenging scenarios.

4.1.2. VOF

In Table 2 a summary of the performance of our method for VOF estimation is shown. A high agreement with the manual annotations is obtained, which is better than the performance obtained for the AIF estimation. These results can be expected since VOF compared to AIF is less affected by partial volume effect, has higher SNR and hence provides lower inter-rater variability (Table 2).

For all the considered metrics excepting T_{peak} there are no statistically significant differences between the inter-rater agreement and the AIFNet vs raters agreement. When the entire VOF signals are considered, a high correlation with the manual annotations is achieved, reaching inter-rater variability ranges. In terms of T_{peak} , a good performance is obtained though the same delaying effect previously described for AIF is found. In this case, however, the delays are within the inter-rater range. For FWHM, the agreement between our method and rater #2 is much higher than among raters. Unlike the AIF analysis, it is worth noticing that there is no flattening or widening of the VOF curves predicted with AIFNet. The evaluation of the recalibration strategy using the curve mean squared error and the PTB metric shows a high agreement and high correlation between AIFNet and the manual annotations reaching inter-rater ranges. In the assessment of the mean squared error, the inter-rater's 5th percentile is zero, which implies that the raters have sometimes selected the exact same voxel. The evaluation of the PTB errors shows no clear trend of our method towards

under/over-estimation of the VOF signals, suggesting a good overall performance of the recalibration strategy.

4.1.3. Arterial localization

The anatomical localization that AIFNet conducts can be assessed from the voxelwise activation encoded in P_{vol} . Unlike most AIF selection approaches selecting only few candidates, AIFNet allows multiple voxel contribution for building the vascular functions.

In Fig. 5 the best and worst AIF (in correlations terms) among all predictions are shown. While the prediction with higher agreement achieves a Pearson's $r = 0.999$ (left-side of the figure), the case with poorest agreement achieves an $r = 0.674$ (right-side of the figure). Both raters have chosen the same AIF voxel in the best performance scenario. In the top-left part of Fig. 5 it can be seen that just a few voxels are activated in the displayed CT slice, having high activation values. The AIF voxel selected by the raters ($y_{r1,r2}(t)$) is also being activated by AIFNet, being the second highest value of P_{vol} . Mainly voxels belonging to the anterior cerebral artery are chosen. Besides, the AIF that our method predicts follows closely the raters' function, with no observable delays and with almost no differences in the curves' shape. On the other hand, localization results from the worst Pearson's correlation case shows a different behaviour. Several voxels belonging to different arteries are enhanced by the network with a homogeneous activation distribution. The anterior cerebral artery and middle cerebral arteries are mainly selected. When assessing $y_{r1}(t)$ and $\hat{y}_{AIFNet}(t)$, it is noticeable that the low Pearson's r is driven by the time shift between the functions (which is 4 seconds measured at the curve peaks). In this case, AIFNet outperforms rater #1 by estimating a vascular function with high agreement in morphology, which oc-

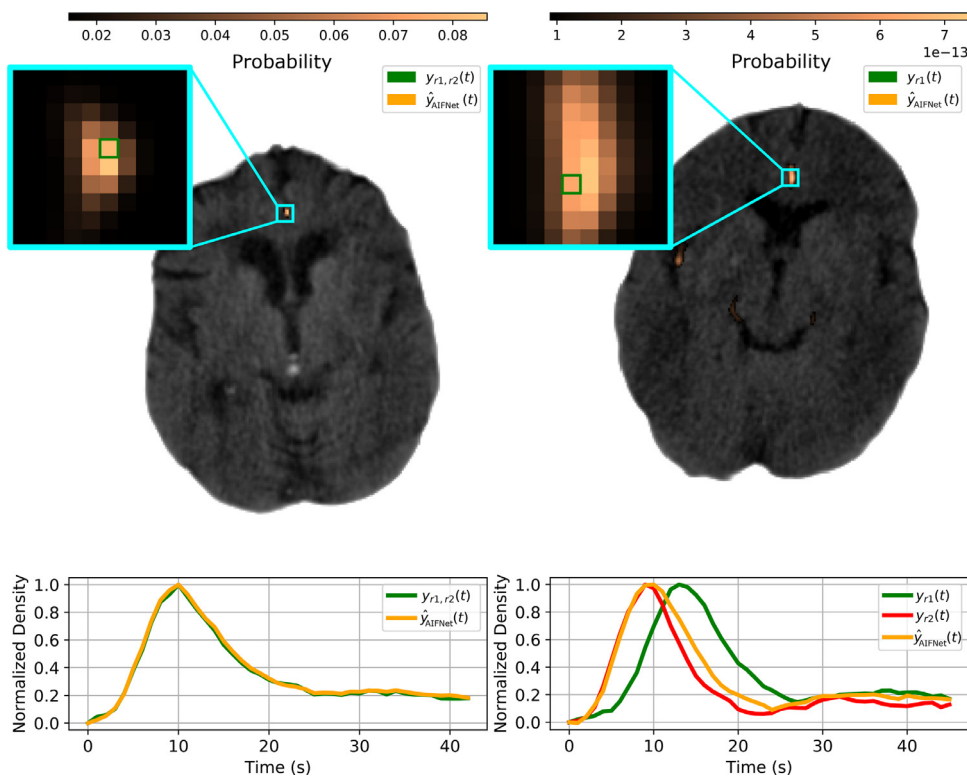


Fig. 5. Best (left) and worst (right) prediction performance in terms of Pearson's correlation between AIF functions. Above, the voxels selected by rater #1 ($y_{r1}(t)$), rater #2 ($y_{r2}(t)$) and AIFNet ($\hat{y}_{AIFNet}(t)$) as arterial input functions. Below, their corresponding vascular function. Note that in the best performance case, both raters have chosen the exact same voxel as AIF. In the worst performance case, the selected voxel location for $y_{r2}(t)$ is not shown since was annotated in a different volume slice.

Table 3

Parameter maps agreement among methods in terms of Pearson's correlation. Mean (standard deviation) values are provided. Correlation has been computed per scan using all the voxels within the brain tissue (excluding background, skull, ventricles and vessels). y_{r1} : Parameter maps obtained after deconvolving the images with the AIF of rater 1; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Parameter maps obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. CBF: cerebral blood flow; CBV: cerebral blood volume; MTT: mean transit time; T_{max} : time to the maximum of the residue function. The values in bold indicate the outperforming method (in terms of Pearson's r) for each parameter map.

		Pearson's r coefficient			
		CBF	CBV	T_{max}	MTT
Inter-rater		0.998 (0.016)	0.987 (0.157)	0.944 (0.086)	0.927 (0.204)
y_{r1} vs	\hat{y}_{Kmeans}	0.967 (0.098)	0.960 (0.274)	0.786 (0.217)	0.772 (0.291)
	\hat{y}_{Seg}	0.974 (0.118)	0.947 (0.315)	0.781 (0.237)	0.758 (0.310)
	\hat{y}_{Reg}	0.990 (0.027)	0.972 (0.160)	0.809 (0.179)	0.791 (0.253)
	\hat{y}_{AIFNet}	0.998 (0.007)	1.000 (0.003)	0.921 (0.094)	0.908 (0.188)

curs much earlier than the manually selected one. We consider the annotation of rater #1 suboptimal, probably because the voxel was chosen from an artery branch already affected by the occlusion. However, our prediction follows more closely $y_{r2}(t)$ (Pearson's $r = 0.980$). There is no observable function delay between $y_{r2}(t)$ and $\hat{y}_{AIFNet}(t)$, though a slightly wider FWHM can be appreciated for $\hat{y}_{AIFNet}(t)$.

4.2. Parameter maps and lesion quantification

The parameter maps correlation between rater #1 and the different approaches is shown in Table 3. Similar results are obtained when comparing the parameter maps with rater #2 (Table S2). Correlation values are computed for each scan within the brain tissue (excluding background, skull, ventricles and vessels). Among all the compared methods AIFNet obtains the best agreement with the rater for each of the parameter maps, showing consistency with the experts' results. An outstanding agreement is observed between the raters and AIFNet for CBF and CBV, reaching inter-

rater performance. For T_{max} and MTT, however, the agreement is still high but marginally under the inter-rater performance. We hypothesize that the found lower consistency in these parameter maps could be driven by the ~ 0.5 seconds delay in T_{peak} and by the ~ 1 second wider FWHM of AIFNet predictions. It is worth to notice that the other automatic AIF selection methods also obtained better performance for CBF and CBV than for T_{max} and MTT. Parameter maps and lesion masks obtained with all methods are shown in Fig. 6 for the scan with median AIF Pearson correlation (y_{r1} vs \hat{y}_{AIFNet} comparison). There is a high qualitative correspondence between raters and AIFNet at all levels. It is also seen here that there is a better correspondence between rater #1 and all automatic AIF selection methods for estimating rCBF than estimating T_{max} .

In Fig. 7 the hypoperfused and core volumes between rater #1 and all the methods are shown. Likewise, the methods' agreement with rater #2 follows a similar pattern (please see the Supplementary material, Fig. S1). In Table 4 (Table S3), the lesion volumes quantification performance is shown for all the methods when

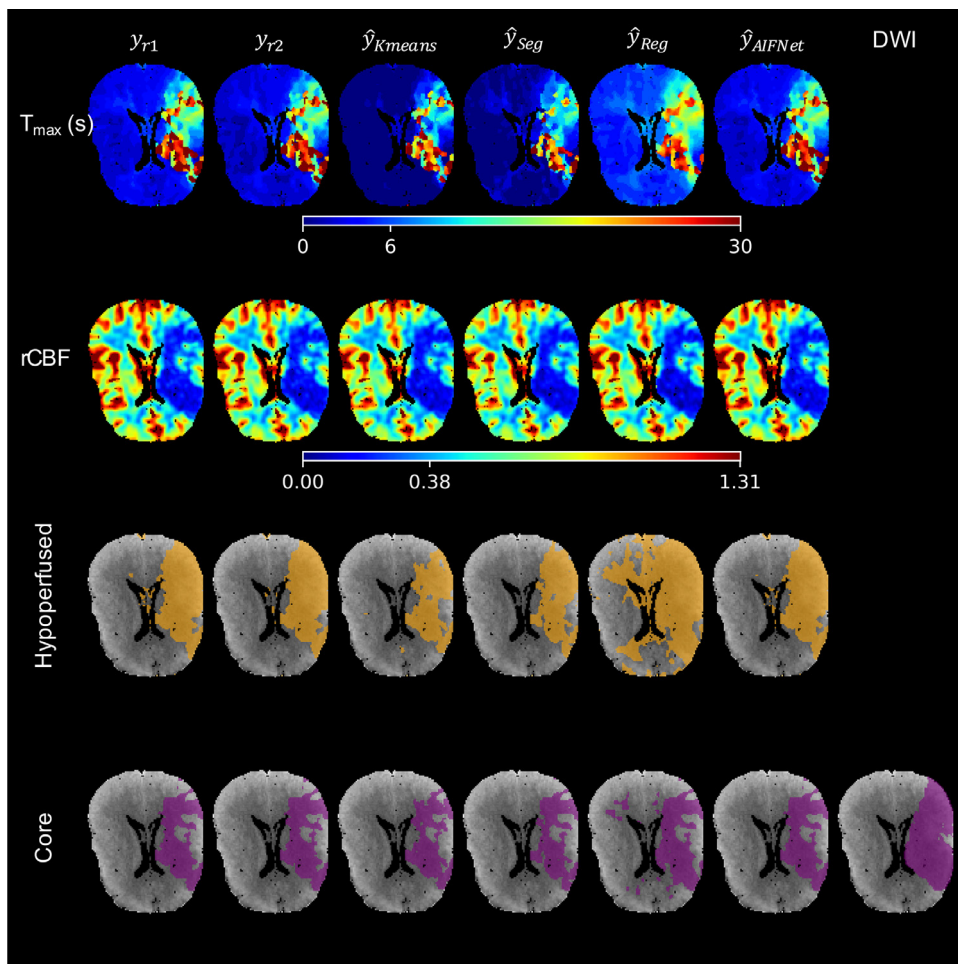


Fig. 6. Estimated parameter maps and brain lesions for all the methods, obtained after deconvolving the CTP images with the annotated or predicted vascular functions. The shown example is the scan with median AIF Pearson correlation (y_{r1} vs \hat{y}_{AIFNet} comparison). Hypoperfused tissue is obtained after thresholding $T_{max} < 6$ s. Core tissue is obtained after thresholding the rCBF map at 38% over the entire hypoperfused region. y_{r1} , y_{r2} : Results obtained after deconvolving the images with the AIF of rater 1 and rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Results obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. rCBF: relative cerebral blood flow map; T_{max} : time to the maximum of the residue function map. DWI: Ground truth delineated over DWI in ISLES18.

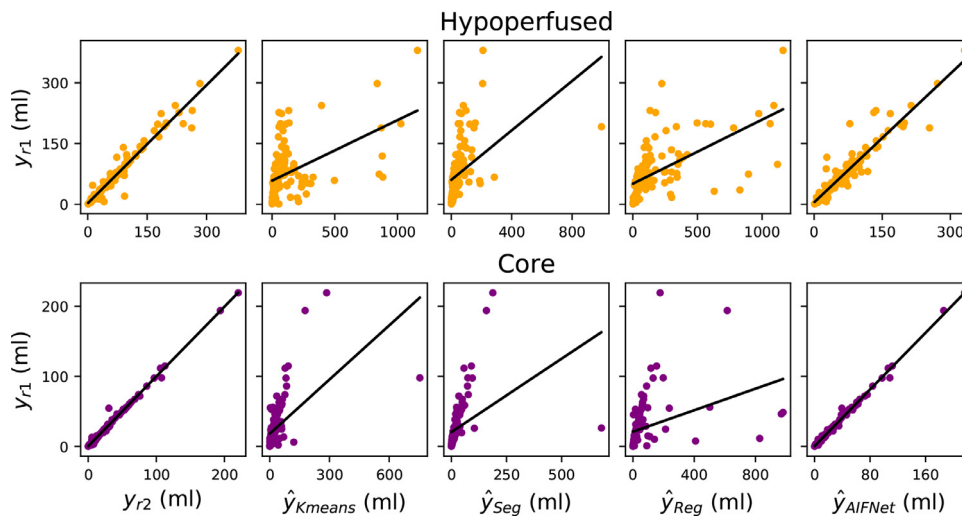


Fig. 7. Lesion volume agreement among methods. y_{r1} , y_{r2} : Lesion volumes obtained after deconvolving the images with the AIF of rater 1 and rater 2; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Lesion volumes obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively.

Table 4

Brain lesion quantification performance for all the methods, obtained after deconvolving the CTP images with the manual and automatic vascular functions. Mean (standard deviation) values are provided. Hypoperfused tissue is obtained after thresholding $T_{\max} < 6$ s. Core tissue is obtained after thresholding the rCBF map at 38% over the entire hypoperfused region. y_{r1} : Lesion volumes obtained after deconvolving the images with the AIF of rater 1; \hat{y}_{Kmeans} , \hat{y}_{Seg} , \hat{y}_{Reg} , \hat{y}_{AIFNet} : Lesion volumes obtained after deconvolving the images with the AIF predicted with K-means (Mouridsen et al., 2006), with the regression CNN (de la Rosa et al., 2020), with the segmentation CNN and with AIFNet, respectively. DWI: Agreement obtained when comparing the different CTP approaches with the diffusion weighted imaging ground truth provided in ISLES18. VE: Volume error; AVE: Absolute volume error. A paired significance test is performed between AIFNet and the other approaches. *: p-value < 0.05; †: p-value < 0.01. The values in bold indicate the outperforming method for the metric under consideration.

		Hypoperfused			Core		
		Dice [%]	VE [ml]	AVE [ml]	Dice [%]	VE [ml]	AVE [ml]
Inter-rater		91.7 (13.8)†	0.5 (13.5)†	6.3 (12.0)†	91.8 (14.1)†	0.1 (2.5)†	0.9 (2.3)†
y_{r1} vs	\hat{y}_{Kmeans}	48.3 (28.2)†	-36.9 (174.3)	82.7 (157.8)†	58.3 (27.0)†	-2.6 (55.2)†	13.3 (53.6)†
	\hat{y}_{Seg}	51.3 (19.4)†	26.5 (80.8)†	45.5 (71.8)†	61.9 (20.7)†	1.7 (53.9)†	12.0 (52.6)†
	\hat{y}_{Reg}	70.1 (23.4)†	-78.5 (198.2)†	94.2 (191.3)†	72.3 (23.3)†	-32.0 (136.4)	35.7 (135.5)†
	\hat{y}_{AIFNet}	87.3 (13.0)	8.8 (20.3)	12.7 (18.2)	88.3 (13.6)	0.5 (2.3)	1.3 (1.9)
y_{r1}					38.3 (19.4)	6.8 (20.0)	14.3 (15.6)
y_{r2}					38.3 (19.5)	6.6 (20.2)†	14.2 (15.7)
DWI vs	\hat{y}_{Kmeans}				31.5 (19.9)†	1.2 (72.9)†	25.1 (68.4)†
	\hat{y}_{Seg}				32.7 (19.0)†	12.7 (22.2)†	16.4 (19.6)*
	\hat{y}_{Reg}				35.1 (19.8)*	-34.2 (154.3)†	52.6 (149.0)†
	\hat{y}_{AIFNet}				38.1 (19.5)	7.2 (20.5)	14.5 (16.2)

compared with rater #1 (#2). AIFNet consistently outperforms the other approaches (i.e., with statistically significant comparisons for almost all the core and hypoperfused metrics) and reaches a high agreement with the experts slightly below the inter-rater performance (there are statistically significant differences between raters and AIFNet as well). While the agreement with the raters for the core volumes is very high, a slight bias in the hypoperfused volumes can be appreciated, suggesting that AIFNet tends to underestimate these tissue areas. This bias is a consequence of the AIF differences obtained with AIFNet (described in Section 4.1) that impact over T_{\max} .

Furthermore, Table 4 compares the predicted core volumes with the ground truth DWI core masks. In this comparison, AIFNet closely follows the manual raters performance: except the volume error comparison between AIFNet and rater #2, there are no statistically significant differences with the raters in the quantified metrics. Our results suggest that AIFNet estimates the acute brain infarcts as good as manual raters do. The core volume correlation between the different CTP methods and the DWI ground truth is depicted in Figure S2. When comparing the core CTP volumes of all the investigated approaches (raters and automatic algorithms) with the DWI ones, there are statistically significant differences for all the methods (p-values < 0.01, Wilcoxon test). Nonetheless, this cross-modality (i.e. CTP-DWI) performance evaluation should be carefully judged, since external source of errors are being introduced (such as brain perfusion changes due to time acquisition differences, modality co-registration errors, etc). These limitations in cross-modality correspondences explain the low overall Dice coefficients even found for the raters in the CTP-DWI comparison.

Overall, the parameter maps and lesion quantification analysis show that the small AIF differences found with AIFNet (mainly in T_{peak} and FWHM) do not produce a large impact over the deconvolution process. The main observation related to these AIF differences is the bias found in the hypoperfused volumes, showing a trend in AIFNet to slightly underestimate this tissue. There are no important differences in the core predictions of AIFNet in comparison to the ones of the manual raters. There are neither differences in the CTP-DWI core agreement: the raters and AIFNet are both equally consistent compared to the ISLES18 DWI ground truth. Overall, our analysis of the vascular functions, of the parameters maps and lesion volumes obtained after CTP deconvolution show that AIFNet obtains state-of-the art performance in automatic AIF selection. The method almost behaves as a manual expert in all

the considered deconvolution stages and its performance is close to the inter-rater one.

4.3. Comparison with other methods

The comparison of different AIF selection methods shows that our innovative CNN is the most suitable approach among the investigated ones for performing the task. Unlike AIFNet, the other automatic algorithms predict AIFs that consistently compromise the parameter maps quality and, hence, the estimation of the brain lesions. We also observe that our method can robustly work under challenging perfusion cases, as shown for different quality scans in Fig. 4.

The K-means method (Mouridsen et al., 2006) shows several limitations and provides a large performance variability among the computed AIFs (large standard deviations of the AIF metrics). These results could be expected since the method has been originally devised for perfusion MRI, which accounts with higher signal-to-noise ratio and smoother perfusion curves than CTP. As such, the AIF unsupervised clustering might better perform in perfusion MRI than in CTP. The results from the segmentation CNN show similar performance to the K-means method and, as such, worse than the AIFNet performance. Our approach not only shows to be quantitatively better than the segmentation CNN, but also provides the advantage of requiring minimal data labels: while classification or segmentation approaches would ideally annotate most of the target class samples for their training, AIFNet can be trained with a single voxel annotation. It is worth reminding the reader that the only difference between the segmentation CNN and AIFNet is the optimized loss function. The segmentation network works at a spatial level by recognising similar anatomical/shape areas to the target class, but it fails in selecting good AIF candidate voxels. We show that the AIF selection task is consistently improved when the network learns from the predicted AIF (as happens with the proposed Pearson correlation loss) rather than from the selected voxels only, as happens when using segmentation loss functions. While arterial voxels can be extremely similar in terms of density values, localization and/or anatomical context, their perfusion curves can differ significantly. As such, segmentation-based methods can misleadingly activate arterial voxels whose perfusion curves are suboptimal. These findings are also supported by the better results obtained with the regression CNN (whose training loss is also the Pearson correlation, as for AIFNet) compared to

the segmentation CNN. Nonetheless, the regression CNN results are much worse than the ones of our method. Moreover, fully AIF regression CNNs as earlier proposed in (de la Rosa et al., 2020) have some drawbacks. First, from a qualitative point of view, the model does not provide voxelwise arterial localization, making it less interpretable than other approaches and less transferable to clinical settings. Second, the model requires that the scans' duration is homogeneous. As such, truncation artifacts might be introduced during this process (Kasasbeh et al., 2016). It is interesting to observe that AIFNet overcomes the segmentation and regression CNN limitations by working as a hybrid segmentation-regression network. Its architecture allows voxelwise activation and selection like most segmentation approaches, but its optimization is performed like a regression CNN, which shows to enhance the performance. In other words, AIFNet explores the advantages of segmentation and regression models for improving the AIF estimation.

Another successful deep learning approach for AIF selection has been presented by Winder et al. (2020). Similar to AIFNet, the approach achieves close agreement with the manual raters in the selection of AIF and in the estimated perfusion lesions. The approach differs from AIFNet in several ways. First, the CNN is a 1D model receiving as input single voxel perfusion curves (i.e., there is no spatial/contextual information considered but only temporal information). Second, the method is a binary classifier CNN (AIF vs all-the-rest) and hence it is optimized with a classification loss function. Third, the method estimates the AIF by means of a sophisticated geometric averaging approach. It is worth noting that both the methods (AIFNet and (Winder et al., 2020)) avoid delayed AIF curves in different ways. On one hand, (Winder et al., 2020) isolates potential AIF candidates with the classification CNN and then corrects the time curves with the geometric averaging technique. On the other hand, our proposal discards delayed AIF candidates by restricting the CNN learning through a suitable loss function. Some advantages on the usage of AIFNet over the work of Winder et al. (2020) are *i*) the number of manual annotations required for the model training (around < 20 times labeled data), *ii*) almost no pre-processing required, *iii*) the automatic selection of the number of voxels to average (which is a parameter to set in (Winder et al., 2020)) and *iv*) its full end-to-end framework, which makes our approach easier and faster to deploy.

4.4. Ablation analysis

We ablate our network for finding the optimal architecture and training strategy for computing $\hat{y}_{AIFNet}(t)$. The ablation is conducted for the AIF since it is the most critical input to the deconvolution model and it is much more difficult to estimate than the VOF. These experiments are performed using the original train-test data split of the ISLES18 challenge ($n_{train} = 94$, $n_{test} = 62$). For training purposes we randomly exclude 10% of the training data and use it as validation set, assuring that in all the ablation experiments the same train-validation-test sets are used.

Results for the ablation analysis are shown in Table 5. Our experiments show that $K = 5$ convolutional layers are optimal for AIF prediction. The usage of less convolutional blocks leads not only to lower mean performance but also to higher variability. Besides, results do not improve when considering more than $K = 5$ convolutional layers. It is worth to point out the considerable improvement in robustness when problem-specific data augmentation is considered for training the models. Overall, a much higher 5th percentile is obtained with rather than without data augmentation, showing better generalization over challenging cases. For VOF prediction, $K = 2$ convolutional layers are enough to estimate the function at inter-rater performance. Thus, less features are required for finding good VOF voxel candidates. These results are expected given the higher task difficulty for selecting AIF over VOF, as shown in

Table 5

AIFNet ablation performance. K : Number of convolutional layers in the CNN. Given GPU memory constrains, the AIFNet experiment with $K = 6$ is conducted with 2^{2+k} kernels per layer instead of 2^{3+k} (such as the first layer has 8 kernels and the sixth one 256). DA: Data augmentation; std: Standard deviation; Perc: Percentile. The values in bold indicate the outperforming approach for the metric under consideration.

AIFNet #Layers (K)					Pearson's r	
3	4	5	6	DA	Mean (std)	(5th, 95th Perc)
x					0.943 (0.133)	(0.661, 0.999)
	x				0.947 (0.107)	(0.669, 0.999)
		x			0.950 (0.088)	(0.694, 0.999)
			x		0.946 (0.094)	(0.682, 0.999)
				x	0.957 (0.057)	(0.870, 0.999)

Tables 1 and 2, where a better agreement between raters is shown for VOF than for AIF.

4.5. Limitations and future perspectives

A limitation of this work is the lesion ground truth used, since currently there is no gold standard for the penumbra and for the ischemic core. We use, as provided in ISLES18, the core masks delineated in DWI. However, the acquisition delay between CTP and DWI imaging may introduce ischemic core modifications. Another source of mismatch between the imaging modalities might be introduced by the reperfusion therapy, since reversal of the DWI lesion may happen after reperfusion (Campbell et al., 2012). Even more, mismatch errors could also appear during the cross-modality image registration. Consequently, a full correspondence between CTP and DWI core lesions is unlikely to happen. In our experiments, this mismatch could explain the statistically significant differences found when comparing all CTP volumetric core predictions (from raters and AIFNet) against the delineated DWI ground truth. For a better understanding of the different methods' performance, we include a CTP-CTP analysis by considering as hypoperfused and core ground truth the volumes obtained by using the manual rater annotations in the CTP deconvolution. In such a way, the aforementioned cross-modality limitations are no longer affecting the analysis.

As future perspectives for this work we could consider the validation of AIFNet over a larger database, as well as over other imaging modalities, such as perfusion MRI and PET images. Given the challenging task behind vascular estimation over CTP, we expect the method to be easy to adapt to images of better quality (such as MRIs). Exploring whether the technique is generalizable to other organs and pathologies where perfusion analysis is used (such as in brain tumors, myocardial infarction, etc.) also constitutes potential research lines.

5. Conclusions

We have presented AIFNet, a new automatic method for vascular function estimation in brain perfusion imaging. It is developed and validated over the public ISLES18 database, which consists of stroke perfusion CT cases. To our knowledge, this is one of the first automatic methods described in literature fully developed and validated over perfusion CT data. Most of the approaches previously described have been devised and tested over perfusion MRI instead. For tackling the problem, we make use of a fully end-to-end trainable CNN, that is optimized for the prediction of vascular functions. We exploit prior knowledge by performing modality-specific data augmentation during the training stage. Our approach consistently differs from the previous ones,

which mainly rely on clustering or statistical techniques. Additionally, most of these techniques require the definition of a decision rule (mainly a cutoff) for selecting the optimal voxels, a strategy that might be dataset-dependent and, hence, requires parameters tuning. Unlike these methods, we present a non-heuristic function estimation strategy that combines information from multiple voxels by means of a 3D probabilistic volume. AIFNet allows arterial localization and, hence, clinical interpretability. The method is easier to train and deploy compared to other approaches due to its architecture and due to the minimal voxel annotations required as ground truth (one single voxel per vascular function and per scan is enough to parametrize the network). As a consequence, the database labeling process is very fast. This is a clear advantage of AIFNet when compared against segmentation approaches, since the latter are more time consuming by requiring a vessel region annotation and multiple vascular functions checks. We show, as well, that using a suitable loss function enhances the task performance. After validating AIFNet in the ISLES18 dataset, the method outperformed existing methods and achieved results close to the inter-rater agreement, being able to make predictions of vascular functions, parameter maps and perfusion lesions with similar performance as human experts. Besides, the approach shows to be robust under poor quality-scan scenarios. Our results suggests that AIFNet could be implemented in clinical scenarios and, hence, could potentially be included in future brain perfusion deconvolution software. For better reproducibility and direct comparison against future methods, we provide both raters' vascular annotations as supplementary material.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

EdIR and DR are co-inventors in technology related to this research; a patent application has been submitted and is pending. EdIR, DMS and DR are employees of **icometrix**.

CRediT authorship contribution statement

Ezequiel de la Rosa: Data curation, Investigation, Methodology, Software, Writing – original draft. **Diana M. Sima:** Supervision, Writing – review & editing, Funding acquisition. **Bjoern Menze:** Conceptualization, Methodology, Writing – review & editing. **Jan S. Kirschke:** Conceptualization, Writing – review & editing. **David Robben:** Data curation, Formal analysis, Supervision, Conceptualization, Writing – review & editing.

Acknowledgement

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement TRABIT No 765148. DR is supported by an innovation mandate of Flanders Innovation & Entrepreneurship (VLAIO). BM received funding from the project "Stroke treatment goes personalized: Gaining added diagnostic yield by computer-assisted treatment selection" (Deutsche Forschungsgemeinschaft (DFG) - Projekt number 326824585).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.media.2021.102211](https://doi.org/10.1016/j.media.2021.102211)

References

- Abulnaga, S.M., Rubin, J., 2018. Ischemic stroke lesion segmentation in CT perfusion scans using pyramid pooling and focal loss. In: International MICCAI Brainlesion Workshop. Springer, pp. 352–363.
- Albers, G.W., Goyal, M., Jahan, R., Bonafe, A., Diener, H.-C., Levy, E.I., Pereira, V.M., Cognard, C., Cohen, D.J., Hacke, W., et al., 2016. Ischemic core and hypoperfusion volumes predict infarct size in SWIFT PRIME. *Ann. Neurol.* 79 (1), 76–89.
- Bertels, J., Robben, D., Vandermeulen, D., Suetens, P., 2018. Contra-lateral information CNN for core lesion segmentation based on native CTP in acute stroke. In: International MICCAI Brainlesion Workshop. Springer, pp. 263–270.
- Calamante, F., 2013. Arterial input function in perfusion MRI: a comprehensive review. *Prog Nucl Magn Reson Spectrosc* 74, 1–32.
- Calamante, F., Mørup, M., Hansen, L.K., 2004. Defining a local arterial input function for perfusion MRI using independent component analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 52 (4), 789–797.
- Campbell, B.C.V., Parsons, M.W., 2018. Imaging selection for acute stroke intervention. *International Journal of Stroke* 13 (6), 554–567.
- Campbell, B.C.V., Purushotham, A., Christensen, S., Desmond, P.M., Nagakane, Y., Parsons, M.W., Lansberg, M.G., Mlynash, M., Straka, M., De Silva, D.A., et al., 2012. The infarct core is well represented by the acute diffusion lesion: sustained reversal is infrequent. *Journal of Cerebral Blood Flow & Metabolism* 32 (1), 50–56.
- Cereda, C.W., Christensen, S., Campbell, B.C.V., Mishra, N.K., Mlynash, M., Levi, C., Straka, M., Wintermark, M., Bammer, R., Albers, G.W., et al., 2016. A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a DWI standard. *Journal of Cerebral Blood Flow & Metabolism* 36 (10), 1780–1789.
- Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., Lladó, X., 2019. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* 115, 103487.
- Fan, S., Bian, Y., Wang, E., Wang, D.J.J., Yang, Q., Ji, X., Kang, Y., 2019. An automatic estimation of arterial input function based on multi-stream 3D CNN. *Front Neuroinform* 13, 49.
- Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R., 2011. Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. *Journal of Biomedical Imaging* 2011, 14.
- Force, W.T., 1989. Stroke-1989. Recommendations on stroke prevention, diagnosis, and therapy. Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders. *Stroke* 20 (10), 1407–1431.
- Hinkle, J.L., Guanci, M.M., 2007. Acute ischemic stroke review. *Journal of neuroscience nursing* 39 (5), 285–293.
- Hinton, G., Srivastava, N., Swersky, K., 2012. Neural networks for machine learning—Lecture 6a: Overview of mini-batch gradient descent. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf/, Accessed: May 5, 2021 [Online].
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 36, 61–78.
- Kasasbeh, A.S., Christensen, S., Straka, M., Mishra, N., Mlynash, M., Bammer, R., Albers, G.W., Lansberg, M.G., 2016. Optimal computed tomographic perfusion scan duration for assessment of acute stroke lesion volumes. *Stroke* 47 (12), 2966–2971.
- Kealey, S.M., Loving, V.A., Delong, D.M., Eastwood, J.D., 2004. User-defined vascular input function curves: influence on mean perfusion parameter values and signal-to-noise ratio. *Radiology* 231 (2), 587–593.
- Khatri, P., Yeatts, S.D., Mazighi, M., Broderick, J.P., Liebeskind, D.S., Demchuk, A.M., Amarenco, P., Carrozella, J., Spilker, J., Foster, L.D., et al., 2014. Time to angiographic reperfusion and clinical outcome after acute ischaemic stroke: an analysis of data from the Interventional Management of Stroke (IMS III) phase 3 trial. *The Lancet Neurology* 13 (6), 567–574.
- Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P., 2013. The virtual skeleton database: an open access repository for biomedical research and collaboration. *J. Med. Internet Res.* 15 (11), e245.
- Klotz, E., König, M., 1999. Perfusion measurements of the brain: using dynamic CT for the quantitative assessment of cerebral ischemia in acute stroke. *Eur J Radiol* 30 (3), 170–184.
- Konstas, A.A., Goldmakher, G.V., Lee, T.-Y., Lev, M.H., 2009. Theoretic basis and technical implementations of CT perfusion in acute ischemic stroke, part 1: theoretic basis. *American Journal of Neuroradiology* 30 (4), 662–668.
- Kosior, J.C., Frayne, R., 2007. PerfTool: a software platform for investigating bolus-tracking perfusion imaging quantification strategies. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 25 (3), 653–659.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105.
- Kudo, K., Sasaki, M., Yamada, K., Momoshima, S., Utsunomiya, H., Shirato, H., Ogasawara, K., 2010. Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients. *Radiology* 254 (1), 200–209.
- Lin, L., Bivard, A., Krishnamurthy, V., Levi, C.R., Parsons, M.W., 2016. Whole-brain CT perfusion to quantify acute ischemic penumbra and core. *Radiology* 279 (3), 876–887.
- Lorenz, C., Benner, T., Chen, P.J., Lopez, C.J., Ay, H., Zhu, M.W., Menezes, N.M., Aro-

- nen, H., Karonen, J., Liu, Y., et al., 2006. Automated perfusion-weighted MRI using localized arterial input functions. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 24 (5), 1133–1139.
- Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med Image Anal* 35, 250–269.
- McKinley, R., Hung, F., Wiest, R., Liebeskind, D.S., Scalzo, F., 2018. A machine learning approach to perfusion imaging with dynamic susceptibility contrast MR. *Front Neurol* 9, 717.
- Meier, R., Lux, P., Jung, S., Fischer, U., Gralla, J., Reyes, M., Wiest, R., McKinley, R., Kaesmacher, J., 2019. Neural network-derived perfusion maps for the assessment of lesions in patients with acute ischemic stroke. *Radiology: artificial intelligence* 1 (5), e190019.
- Mlynash, M., Eyngorn, I., Bammer, R., Moseley, M., Tong, D.C., 2005. Automated method for generating the arterial input function on perfusion-weighted MR imaging: validation in patients with stroke. *American Journal of Neuroradiology* 26 (6), 1479–1486.
- Mouridsen, K., Christensen, S., Gyldensted, L., Østergaard, L., 2006. Automatic selection of arterial input function using cluster analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 55 (3), 524–531.
- Murase, K., Kikuchi, K., Miki, H., Shimizu, T., Ikezoe, J., 2001. Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced MR imaging. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 13 (5), 797–806.
- Murphy, B.D., Chen, X., Lee, T.-Y., 2007. Serial changes in CT cerebral blood volume and flow after 4 hours of middle cerebral occlusion in an animal model of embolic cerebral ischemia. *American Journal of Neuroradiology* 28 (4), 743–749.
- Østergaard, L., Sorensen, A.G., Kwong, K.K., Weisskoff, R.M., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part II: Experimental comparison and preliminary results. *Magn Reson Med* 36 (5), 726–736.
- Østergaard, L., Weisskoff, R.M., Chesler, D.A., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Magn Reson Med* 36 (5), 715–725.
- Peruzzo, D., Bertoldo, A., Zanderigo, F., Cobelli, C., 2011. Automatic selection of arterial input function on dynamic contrast-enhanced MR images. *Comput Methods Programs Biomed* 104 (3), e148–e157.
- Rausch, M., Scheffler, K., Rudin, M., Radü, E.W., 2000. Analysis of input functions from different arterial branches with gamma variate functions and cluster analysis for quantitative blood volume measurements. *Magn Reson Imaging* 18 (10), 1235–1243.
- Rempp, K.A., Brix, G., Wenz, F., Becker, C.R., Gückel, F., Lorenz, W.J., 1994. Quantification of regional cerebral blood flow and volume with dynamic susceptibility contrast-enhanced MR imaging. *Radiology* 193 (3), 637–641.
- Robben, D., Boers, A.M.M., Marquering, H.A., Langezaal, L.L., Roos, Y.B., van Oostenbrugge, R.J., van Zwam, W.H., Dippel, D.W.J., Majoie, C.B., van der Lugt, A., et al., 2020. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Med Image Anal* 59, 101589.
- Robben, D., Suetens, P., 2018. Perfusion parameter estimation using neural networks and data augmentation. In: *International MICCAI Brainlesion Workshop*. Springer, pp. 439–446.
- de la Rosa, E., Robben, D., Sima, D.M., Kirschke, J.S., Menze, B., 2020. Differentiable deconvolution for improved stroke perfusion analysis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 593–602.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Departmental Papers (CIS)* 107.
- Shi, L., Wang, D., Liu, W., Fang, K., Wang, Y.-X.J., Huang, W., King, A.D., Heng, P.A., Ahuja, A.T., 2014. Automatic detection of arterial input function in dynamic contrast enhanced MRI based on affinity propagation clustering. *J. Magn. Reson. Imaging* 39 (5), 1327–1337.
- Song, T., Huang, N., 2018. Integrated extractor, generator and segmentor for ischemic stroke lesion segmentation. In: *International MICCAI Brainlesion Workshop*. Springer, pp. 310–318.
- Sourbron, S., Luyypaert, R., Morhard, D., Seelos, K., Reiser, M., Peller, M., 2007. Deconvolution of bolus-tracking data: a comparison of discretization methods. *Physics in Medicine & Biology* 52 (22), 6761.
- Stroke Unit Trialists Collaboration, 2013. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev* 9 (9).
- Thijs, V.N., Somford, D.M., Bammer, R., Robberecht, W., Moseley, M.E., Albers, G.W., 2004. Influence of arterial input function on hypoperfusion volumes measured with perfusion-weighted imaging. *Stroke* 35 (1), 94–98.
- Ulas, C., Das, D., Thrippleton, M.J., del C. Valdés, H.M., Armitage, P.A., Makin, S.D., Wardlaw, J.M., Menze, B.H., 2018. Convolutional neural networks for direct inference of pharmacokinetic parameters: Application to stroke dynamic contrast-enhanced MRI. *Front Neurol* 9.
- Ulas, C., Tetteh, G., Thrippleton, M.J., Armitage, P.A., Makin, S.D., Wardlaw, J.M., Davies, M.E., Menze, B.H., 2018. Direct estimation of pharmacokinetic parameters from DCE-MRI using deep CNN with forward physical model loss. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 39–47.
- Vagal, A., Wintermark, M., Nael, K., Bivard, A., Parsons, M., Grossman, A.W., Khatri, P., 2019. Automated CT perfusion imaging for acute ischemic stroke: pearls and pitfalls for real-world use. *Neurology* 93 (20), 888–898.
- Wang, G., Song, T., Dong, Q., Cui, M., Huang, N., Zhang, S., 2020. Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks. *Med Image Anal* 101787.
- Winder, A., d'Esterre, C.D., Menon, B.K., Fiehler, J., Forkert, N.D., 2020. Automatic arterial input function selection in CT and MR perfusion datasets using deep convolutional neural networks. *Med Phys*.
- Wu, O., Østergaard, L., Weisskoff, R.M., Benner, T., Rosen, B.R., Sorensen, A.G., 2003. Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 50 (1), 164–174.
- Yin, J., Sun, H., Yang, J., Guo, Q., 2015. Automated detection of the arterial input function using normalized cut clustering to determine cerebral perfusion by dynamic susceptibility contrast-magnetic resonance imaging. *J. Magn. Reson. Imaging* 41 (4), 1071–1078.