PAPER

Deep feature-domain matching for cardiac-related component separation from a chest electrical impedance tomography image series: proof-of-concept study

IOP Publishing Physiol. Meas. 43 (2022) 125005 https://doi.org/10.1088/1361-6579/ac9c44

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Deep feature-domain matching for cardiac-related component separation from a chest electrical impedance tomography image series: proof-of-concept study

RECEIVED
19 January 2022

REVISED
5 October 2022

ACCEPTED FOR PUBLICATION
20 October 2022

PUBLISHED
28 December 2022

Ke Zhang[1], **Maokun Li**[1], **Haiqing Liang**[2], **Juan Wang**[3], **Fan Yang**[1], **Shenheng Xu**[1] **and Aria Abubakar**[4]

[1] Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology (BNRist), Institute for Precision Medicine, Tsinghua University, Beijing 100084, People's Republic of China
[2] TEDA International Cardiovascular Hospital, Tianjin 300457, People's Republic of China
[3] National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
[4] Schlumberger, Houston, TX 77056, United States of America

E-mail: maokunli@tsinghua.edu.cn

Keywords: cardiac-related signal, deep metric learning, electrical impedance tomography, feature-domain matching, pulmonary perfusion imaging, signal separation

Abstract

Objectives. The cardiac-related component in chest electrical impedance tomography (EIT) measurement is of potential value to pulmonary perfusion monitoring and cardiac function measurement. In a spontaneous breathing case, cardiac-related signals experience serious interference from ventilation-related signals. Traditional cardiac-related signal-separation methods are usually based on certain features of signals. To further improve the separation accuracy, more comprehensive features of the signals should be exploited. *Approach.* We propose an unsupervised deep-learning method called deep feature-domain matching (DFDM), which exploits the feature-domain similarity of the desired signals and the breath-holding signals. This method is characterized by two sub-steps. In the first step, a novel Siamese network is designed and trained to learn common features of breath-holding signals; in the second step, the Siamese network is used as a feature-matching constraint between the separated signals and the breath-holding signals. *Main results.* The method is first tested using synthetic data, and the results show satisfactory separation accuracy. The method is then tested using the data of three patients with pulmonary embolism, and the consistency between the separated images and the radionuclide perfusion scanning images is checked qualitatively. *Significance.* The method uses a lightweight convolutional neural network for fast network training and inference. It is a potential method for dynamic cardiac-related signal separation in clinical settings.

1. Introduction

Electrical impedance tomography (EIT) is a rapidly developing medical imaging modality with a broadening spectrum of applications (Cheney *et al* 1999, Holder 2004, Bayford 2006, Adler and Boyle 2017, de Castro Martins *et al* 2019). One of the most promising applications of EIT is chest imaging. The measurement process is as follows: first, a number of electrodes are attached around a human chest. Then a small alternating current is injected through a pair of electrodes and the induced voltages on other electrodes are measured. The above process is repeated as the injection position is switched around the chest. Finally, a number of voltage measurements are obtained and used for image reconstruction.

EIT is a potential technique for pulmonary perfusion imaging. One of the methods for perfusion imaging using EIT is based on a bolus injection of hypertonic saline (Frerichs *et al* 2002, Borges *et al* 2012, Kircher *et al* 2020, Hentze *et al* 2021). In this work, we study pulmonary perfusion imaging using EIT without the bolus injection of hypertonic saline. This method requires the separation of the cardiac-related component from the

spontaneous breathing image series, which is usually simultaneously influenced by respiratory and cardiac-related activities. However, this task is very challenging for the following reasons: (1) the cardiac-related component is often weak, and is one or two orders of magnitude weaker in amplitude than the ventilation-related component; (2) cardiac- and ventilation-related components may couple with each other; and (3) cardiac- and ventilation-related components may overlap in the frequency spectrum.

In the past 30 years, several methods have been developed for EIT cardiac-related signal separation. The electrocardiogram (ECG)-gating method (Eyuboglu and Brown 1988, Kunst *et al* 1998, Noordegraaf *et al* 1998) was one of the earliest introduced. This method takes an average of the raw EIT signal over a number of heart cycles determined by the synchronously measured ECG signal. The ventilation-related component, which appears to be random with respect to the cardiac-related component, can be effectively reduced after the averaging. The main drawback of the ECG-gating method is that it gives satisfactory results only after a long period (e.g. 200 heart cycles) of averaging and thus lacks real-time capability. Another fundamental method is based on frequency-domain filtering (Zadehkoochak *et al* 1992, Frerichs *et al* 2009, Carlisle *et al* 2010, Grant *et al* 2011). This method has several drawbacks: (1) the cut-off frequency of the filter needs to be adjusted dynamically according to the heart rate; (2) it cannot handle the case where cardiac- and ventilation-related components overlap or are too close in the frequency spectrum; and (3) it has poor noise-suppression ability.

In 2008, a principal component analysis (PCA)-based separation method (Deibele *et al* 2008) was proposed. This method first analyzes EIT signals using PCA, then generates a few temporal templates by projecting the signals on principal bases, and finally filters the signals using the temporal templates. The PCA-based method can find common temporal patterns in EIT signals and has good noise-suppression ability. However, the eigen decomposition required by PCA is computationally expensive and calls for considerable computing resources for real-time applications. This problem can be alleviated by performing PCA at set intervals if the distribution of the signal changes slowly with time (Deibele *et al* 2008).

The above three separation methods are compared in Pikkemaat and Leonhardt (2010). Other methods include those based on independent component analysis (Rahman *et al* 2013) and empirical mode decomposition (Sun *et al* 2017, Cheng *et al* 2022). A decomposition method based on an orthonormal basis (Krivoshei *et al* 2008) was proposed in the context of bio-impedance signals. For more detailed reviews of EIT signal-separation methods, readers are referred to Leonhardt and Lachmann (2012) and Nguyen *et al* (2012).

In summary, the above separation methods are based on different separation criteria. For example, the frequency-domain filtering method is based on the frequency-domain separability of cardiac- and ventilation-related components, and the PCA-based method exploits the linear uncorrelation of cardiac- and ventilation-related components. In this proof-of-concept study, we introduce a criterion that is based on the matching degree of the separated signal and the breath-holding signals in the feature domain. Breath-holding signals (or apnea signals) are used because they are not influenced by respiratory activity, and have been used by researchers to study pulmonary perfusion (Isaacson *et al* 2006, Fagerberg *et al* 2009a, 2009b).

The contributions of this paper are as follows: (1) the idea of feature-domain matching is introduced to separate the cardiac-related component from a spontaneous breathing EIT image series with the guidance of breath-holding signals; (2) a novel Siamese network with auto-encoding regularization and feature partition is designed; and (3) the idea of deep metric learning is used to learn useful features of breath-holding signals.

The rest of the paper is organized as follows. In section 2, the proposed separation algorithm is introduced in detail. Section 3 describes the procedure for training and test data preparation. Section 4 and 5 present the training and test results. Discussion is made in section 6, and conclusions are drawn in section 7.

2. Separation algorithm

Notation in this paper is as follows. Lowercase bold Latin letters are used to represent column vectors. Uppercase bold Latin letters are used to represent matrices. Script letters are used to represent functions or operators. Some terms used in this paper are explained as follows.

- Siamese network: a network containing two neural networks with identical architectures and shared weights.
- Metric learning: a machine-learning methodology, the goal of which is to learn a representation function that maps the input data into a feature space in which similar data are close to each other and dissimilar data are far away from each other.
- Feature domain: a space of feature vectors that are high-level representations of the original data.

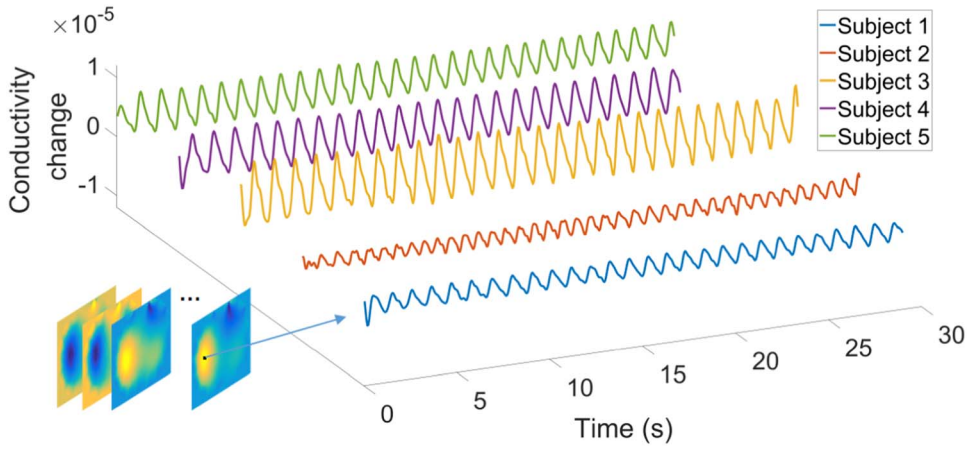

Figure 1. Temporal signals of lung pixels selected from the breath-holding image series of five human subjects.

2.1. Motivation

Assume $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N]$ is the measured EIT boundary voltage matrix, where $\mathbf{d}_i (i = 1, 2, \ldots, N)$ is the voltage frame measured at the discrete time point i, and N is the total number of frames. By using a linearized difference image reconstruction algorithm, a series of difference images can be obtained as (Adler *et al* 2009)

$$\mathbf{a}_i = \mathbf{R} \cdot (\mathbf{d}_i - \mathbf{d}_{ref}), \quad i = 1, 2, \ldots, N, \tag{1}$$

where $\mathbf{d}_{ref}$ is the reference voltage frame and $\mathbf{R}$ is the reconstruction matrix. The reference $\mathbf{d}_{ref}$ can be selected as an arbitrary column of $\mathbf{D}$. In our experiments, a reconstruction matrix with Tikhonov regularization (Tikhonov 1963) is used for $\mathbf{R}$, i.e. $\mathbf{R} = (\mathbf{J}^T\mathbf{J} + \alpha\mathbf{I})^{-1}\mathbf{J}^T$, where $\mathbf{J}$ is the Jacobian matrix of the EIT forward model, $\mathbf{I}$ is the identity matrix, and α is a weighting parameter. Reconstruction matrices of other forms (Cheney *et al* 1990, Adler *et al* 2009, Liu *et al* 2019) can also be used. It is worth noting that $\mathbf{a}_i (i = 1, 2, \ldots, N)$ are vectorized conductivity difference images with vector elements corresponding to the pixel values in the reconstruction. The positive, negative, and zero values in $\mathbf{a}_i$ represent conductivity increase, conductivity decrease, and no conductivity change relative to the reference time point, respectively.

The difference image series can be arranged into a matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$. Each column of $\mathbf{A}$ is the difference image at a discrete time point, and each row of $\mathbf{A}$ is the temporal signal of a pixel. Assume voltages are measured on chests of subjects during spontaneous breathing (SB) and breath-holding (BH). We can reconstruct two image series $\mathbf{A}_{SB}$ and $\mathbf{A}_{BH}$ using the spontaneous breathing and the breath-holding voltages, respectively. In a breath-holding image series $\mathbf{A}_{BH}$, the temporal variation of pixels can be regarded as mainly due to the cardiac activity. In a spontaneous breathing image series $\mathbf{A}_{SB}$, the temporal variation of pixels is influenced by respiratory and cardiac activities simultaneously.

Figure 1 shows the temporal signals of lung pixels selected from the breath-holding image series of five human subjects (aged 26–62 years). It can be observed that signals from different subjects show common features, such as the behavior of the slope change and the asymmetry of the wave form within a cardiac cycle. We assume that the cardiac-related components in the spontaneous breathing signals share similar features with the breath-holding signals, at least for the same subject. Based on this assumption, it is possible to use the breath-holding signals to guide the separation of the cardiac-related components from the spontaneous breathing signals.

The features extracted by sight, such as the slope and the asymmetry, are intuitive, but are limited and difficult to reflect the whole picture of a signal. In view of this, convolutional neural networks (CNNs) are used to automatically learn common features from the breath-holding signals. In the separation, the separated signals are forced to match the breath-holding signals in the learned feature domain.

2.2. Network architecture

2.2.1. High-level architecture

The high-level architecture of the proposed separation network is illustrated in figure 2. Overall, it contains two building blocks: a separation CNN and a CNN-based metric network. The separation CNN takes the spontaneous breathing signal as the input and outputs the separated signal. The metric network takes the separated signal and the breath-holding signal as the input and outputs their similarity score. The architecture of the metric network is based on a Siamese network and a cosine-similarity layer. The Siamese network contains two identical feature-extraction CNNs with shared weights.

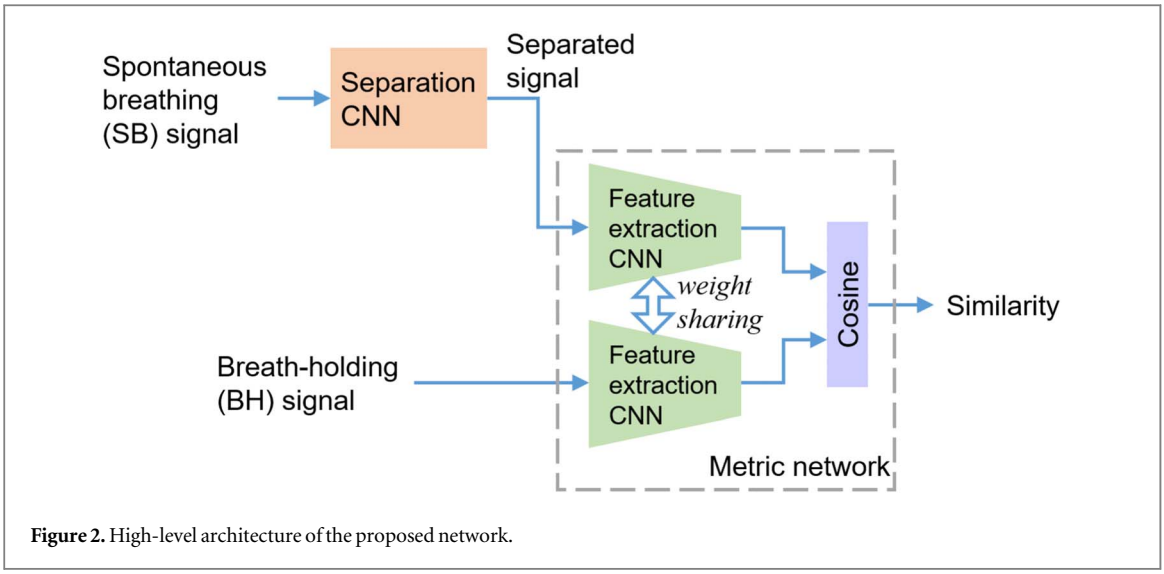

Figure 2. High-level architecture of the proposed network.

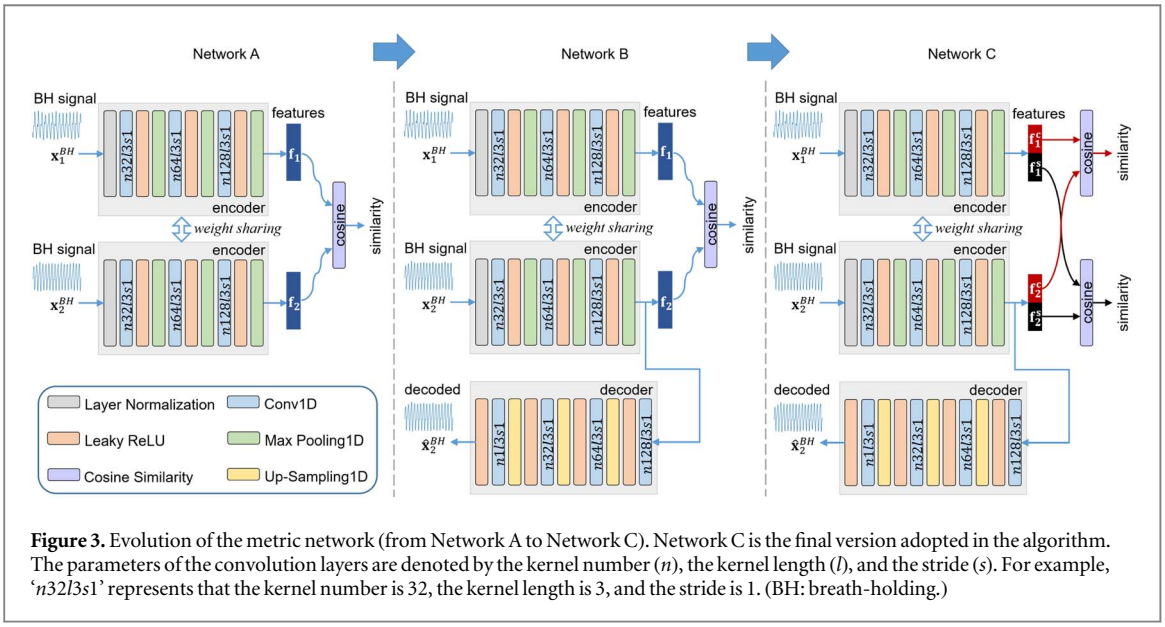

Figure 3. Evolution of the metric network (from Network A to Network C). Network C is the final version adopted in the algorithm. The parameters of the convolution layers are denoted by the kernel number (n), the kernel length (l), and the stride (s). For example, '$n32l3s1$' represents that the kernel number is 32, the kernel length is 3, and the stride is 1. (BH: breath-holding.)

In the network shown in figure 2, the weights of the metric network are pre-trained and fixed before the training of the separation CNN. In the training of the separation CNN, the separated signal is forced to match the breath-holding signal in the learned feature domain of the Siamese network through maximizing the similarity score.

2.2.2. Metric network

The metric network has two modes: the feature-learning mode and the feature-matching mode. In the feature-learning mode, the metric network tries to exploit the common features from the breath-holding signals. Its design process is illustrated in figure 3. First of all, the idea of deep metric learning is used, and a Siamese network (Network A, figure 3) (Yi *et al* 2014) is built. The Siamese network consists of two CNNs with identical structures and shared weights, followed by a cosine layer that calculates the cosine-similarity of the features output by the two CNNs. Assume the two input (breath-holding) temporal signals to the Siamese network are $\mathbf{x}_1^{BH}$ and $\mathbf{x}_2^{BH}$ (transpose of rows of $\mathbf{A}_{BH}$), respectively, which can come from the reconstructed difference image series of the same subject or different subjects. The signals are first Z-Score normalized by the layer-normalization layer (see figure 3 and appendix A for more details).

The output features of CNNs can be written as $\mathbf{f}_1 = \mathcal{F}_1(\mathbf{x}_1^{BH})$ and $\mathbf{f}_2 = \mathcal{F}_2(\mathbf{x}_2^{BH})$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ are operators representing the two CNNs. The values of $\mathbf{f}_1$ and $\mathbf{f}_2$ can be either positive or negative as the layers of the CNNs allow both positive and negative outputs. Then the cosine similarity of $\mathbf{f}_1$ and $\mathbf{f}_2$ are calculated by the cosine-similarity layer (see figure 3 and appendix A for more details). The cosine similarity has a range of values

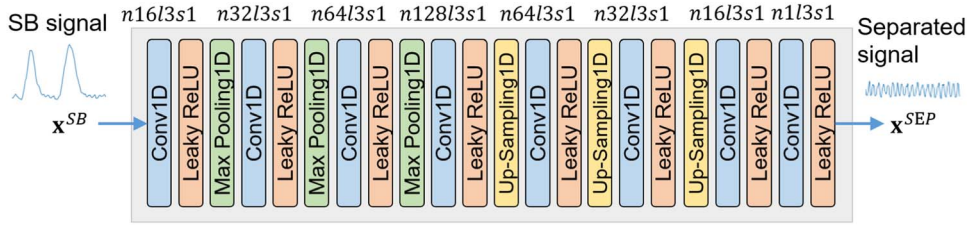

Figure 4. Detailed architecture of the separation network. (SB: spontaneous breathing.)

from −1 to 1, where 1 represents that the vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ have the same direction ($\mathbf{f}_1$ and $\mathbf{f}_2$ are the most similar) and −1 represents that $\mathbf{f}_1$ and $\mathbf{f}_2$ have opposite directions ($\mathbf{f}_1$ and $\mathbf{f}_2$ are the most dissimilar).

In deep metric learning problems, multi-class training data are available and the goal is to map these data into an embedded space in which the embeddings are close for intra-class data and far apart for inter-class data (Yi *et al* 2014, Hoffer and Ailon 2015, Sohn 2016, Song *et al* 2016, Wang *et al* 2019). However, there are no clear classes present in our data. If solely the breath-holding signals are used for the training and the similarity is maximized, the CNNs will simply learn meaningless identical features for all the training data.

How can we ensure that the Siamese network is learning useful features when trained using only one-class breath-holding data? If the features learned by CNNs are decoded by a decoder and the original inputs ($\mathbf{x}_1^{BH}$ and $\mathbf{x}_2^{BH}$) can be reconstructed well, the learned features may be useful because they are input-specific and contain all the information of the inputs. For this reason, a modified network (Network B, figure 3) is built by adding a decoder to the output of one of the CNNs. Accordingly, the optimization goal becomes maximizing the similarity of features and minimizing the decoding error simultaneously.

With the above modification, the feature corresponding to a certain breath-holding signal reserves all its information. However, this means not all parts of this feature are 'common' for breath-holding signals. To give an intuitive example, the behavior of the slope change and the asymmetry within a cardiac cycle are common features for breath-holding signals, whereas the frequency and the initial phase are not.

Based on the above thought, a further modification (Network C, figure 3) is made to the network. Specifically, the output features ($\mathbf{f}_i$, $i = 1, 2$) of the CNNs are divided into two parts: the common part ($\mathbf{f}_i^c$, $i = 1, 2$) and the specific part ($\mathbf{f}_i^s$, $i = 1, 2$), i.e.

$$\begin{aligned} \mathbf{f}_i &= \begin{bmatrix} \mathbf{f}_i^c \\ \mathbf{f}_i^s \end{bmatrix} = \begin{bmatrix} \mathbf{M}_c \cdot \mathbf{f}_i \\ \mathbf{M}_s \cdot \mathbf{f}_i \end{bmatrix}, \quad i = 1, 2 \\ \mathbf{M}_c &= (\mathbf{I}_m, \mathbf{0}_{m\times(l-m)}) \\ \mathbf{M}_s &= (\mathbf{0}_{(l-m)\times m}, \mathbf{I}_{l-m}), \end{aligned} \tag{2}$$

where l is the length of $\mathbf{f}_i$, m is the length of $\mathbf{f}_i^c$, $\mathbf{I}$ is the identity matrix, and $\mathbf{0}$ is the null matrix. For the common part, the optimization goal is to maximize the similarity $\mathcal{S}_{cos}(\mathbf{f}_1^c, \mathbf{f}_2^c)$; for the specific part, the optimization goal is to minimize the similarity $\mathcal{S}_{cos}(\mathbf{f}_1^s, \mathbf{f}_2^s)$. In this way, the specific features are forced into $\mathbf{f}_i^s$, whereas the common features are forced into $\mathbf{f}_i^c$.

Network C in figure 3 is the final network adopted in the algorithm. The number of trainable parameters is 111 233 (31 040 in the encoder and 80 193 in the decoder).

2.2.3. Separation network

The separation network is also based on the CNN, as shown in figure 4. The network adopts an encoder–decoder architecture as both the input and the output are time series with equal lengths. The number of trainable parameters in the network is 64 961.

2.3. Training of the networks

The training contains two stages: (1) training of the metric network; and (2) training of the separation network.

In the training of the metric network (Network C in figure 3), the metric network works in the feature-learning mode. There are three optimization goals: (1) maximizing the similarity of common features; (2) minimizing the similarity of specific features; and (3) minimizing the reconstruction error of the decoder. Accordingly, we design the following loss function

$$\begin{aligned}\mathcal{L}_{metric} = &\beta\mathcal{J}_{dev}(\mathcal{S}_{cos}(\mathbf{M}_c \cdot \mathcal{F}_1(\mathbf{x}_1^{BH}),\ \mathbf{M}_c \cdot \mathcal{F}_2(\mathbf{x}_2^{BH})),\ 1)\\ &+ \gamma\mathcal{J}_{dev}(\mathcal{S}_{cos}(\mathbf{M}_s \cdot \mathcal{F}_1(\mathbf{x}_1^{BH}),\ \mathbf{M}_s \cdot \mathcal{F}_2(\mathbf{x}_2^{BH})),\ -1)\\ &+ \|\mathbf{x}_2^{BH} - \hat{\mathbf{x}}_2^{BH}\|^2,\end{aligned} \tag{3}$$

where $\hat{\mathbf{x}}_2^{BH}$ is the decoded signal by the decoder, $\mathcal{J}_{dev}$ is the binomial deviance having the following form (Hastie *et al* 2005, Yi *et al* 2014):

$$\mathcal{J}_{dev}(x,\ y) = \ln(e^{-2xy} + 1), \tag{4}$$

β and γ are hyper-parameters controlling the trade-off between the three terms, and $\|\cdot\|$ is the l_2 norm. Here, we adopt the binomial deviance loss as it is proved to be robust to outliers.

After the training of the metric network, the decoder and the cosine-similarity layer for specific features are removed. The remaining network, having exactly the structure shown in the dotted box in figure 2, is combined with the separation network. Here, the weights of the metric network are fixed, and the metric network works in the feature-matching mode. The optimization goal for the separation network is to maximize the similarity between the features of the separated signals and the features of the breath-holding signals. To this end, we design the following loss function

$$\mathcal{L}_{sim} = \mathcal{J}_{dev}(\mathcal{S}_{cos}(\mathbf{M}_c \cdot \mathcal{F}_1(\mathcal{G}(\mathbf{x}^{SB})),\ \mathbf{M}_c \cdot \mathcal{F}_2(\mathbf{x}^{BH})),\ 1), \tag{5}$$

where $\mathcal{G}$ is an operator representing the separation network.

However, the above constraint cannot guarantee that the output of the separation network is synchronous with the cardiac-related component of the input spontaneous breathing signal. To address this problem, we design the *spectral inner product* loss:

$$\mathcal{L}_{sip} = \frac{\langle \mathrm{PSD}((\mathbf{x}^{SB} - \mathcal{G}(\mathbf{x}^{SB}))'),\ \mathrm{PSD}((\mathcal{G}(\mathbf{x}^{SB}))')\rangle}{\sum \mathrm{PSD}((\mathcal{G}(\mathbf{x}^{SB}))') + \epsilon}, \tag{6}$$

where PSD$(\cdot)$ is the power spectral density (PSD) of a given temporal signal, $(\cdot)'$ is the time derivative of a time-domain signal, $\langle\cdot,\cdot\rangle$ is the inner product operator (for two column vectors $\mathbf{u}$ and $\mathbf{v}$, their inner product is calculated as $\langle\mathbf{u},\mathbf{v}\rangle = \mathbf{u}^T\mathbf{v}$), and ϵ is a small positive number for numerical stability. More details about the design of equation 6 can be found in appendix B.

In summary, the loss for the separation network can be written as:

$$\mathcal{L}_{sep} = \mathcal{L}_{sim} + \lambda\mathcal{L}_{sip}, \tag{7}$$

where λ is a hyper-parameter controlling the trade-off between the two terms.

3. Data preparation

3.1. Data from healthy subjects

An EIT measurement system (Infivision 1900, Beijing Huarui Boshi Medical Imaging Technology Co., Ltd., Beijing, China) working at a frequency of 20 kHz was used in the data acquisition. The injected alternating current has a root mean square value of 2 mA. The skip-2 excitation-measurement protocol (Silva *et al* 2017) was utilized, without measurement on the excitation electrodes. Besides, only half of the voltages were measured due to the reciprocity. For a 16-electrode configuration, this leads to 104 ($(16 \times 13)/2$) voltage measurements in each frame of data. The data-acquisition rate was 20 frames per second.

Thirty-seven subjects (23 males and 14 females, aged 50 ± 16 years, indexed as Subject 1–37) were recruited for EIT data acquisition. The experiment was approved by the Institution Review Board of Tsinghua University (approval no.: 20 200 044), and informed consent was obtained from the subjects. The experimental protocol is as follows.

1. **Preparation**: attach 16 conductive silica gel electrodes around the chest of a subject at the level of the xiphoid process. Connect the cables of the EIT system to the electrodes. The subject takes the supine position;
2. **Data acquisition**: the subject performs spontaneous breathing for 120 s, end-inspiration breath-holding for 30 s, spontaneous breathing for 60 s, and end-expiration breath-holding for 30 s, successively.

Before image reconstruction, the measured breath-holding voltages are preprocessed for two steps: (1) direct-current (DC) component and baseline wander removal; and (2) noise reduction. For the first step, a band-pass finite impulse response (FIR) filter of order 50 is applied for the data of each subject. The lower cut-off frequency of the filter is set lower than the heart rate, and the upper cut-off frequency is set slightly higher than three times the heart rate. For the second step, we apply the PCA noise reduction as detailed in appendix C.

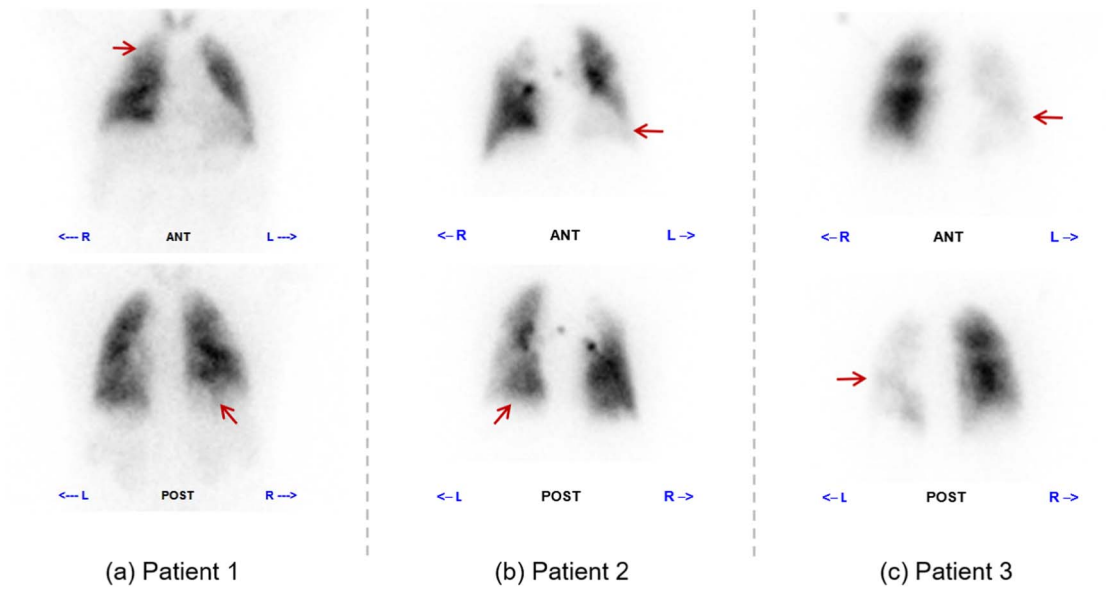

Figure 5. Radionuclide perfusion scanning images (coronal view) of the patients.

For the image reconstruction, a reconstruction matrix $\mathbf{R}$ with Tikhonov regularization (see section 2.1) is used. A chest-shaped finite-element model with 16 547 nodes and 77 322 tetrahedrons is used to calculate the Jacobian matrix. The hyper-parameter α is chosen empirically. The calculated tetrahedral mesh-based reconstruction matrix is then interpolated onto a 120×80 two-dimensional (2D) grid in the electrode plane.

In the reconstruction of the breath-holding image series $\mathbf{A}_{BH}$ of each subject, the reference time point is selected as a peak point of the mean over all voltages. It is worth noting that the reference time points can be arbitrarily selected because different reference time points will only alter the DC component of the signal, which will be removed by the first layer of the Siamese network (see figure 3). The breath-holding signals of pixels, i.e. rows of $\mathbf{A}_{BH}$, are then cropped into 15 s segments. The length of each segment is 300 because the sampling frequency is 20 Hz.

In the generated breath-holding signal segments above, not all of them belong to the lung and heart regions. Therefore, a further step is taken to select all the segments belonging to the lung and heart regions. Firstly, the standard deviation of the time-domain signal of each pixel in $\mathbf{A}_{BH}$ is calculated and a standard deviation image is obtained. Secondly, the standard deviation images of all the subjects are averaged. Finally, a threshold is set and the pixels whose values are above the threshold are segmented as a region of interest (ROI). Because both the lung and the heart have positive conductivity changes in the average standard deviation image, the segmented ROI contains both the heart and the lung. A total number of 195 200 segments are selected from the lung and heart regions. We refer to this dataset as the *Breath-Holding Dataset*.

In the reconstruction of the spontaneous breathing image series $\mathbf{A}_{SB}$ of each subject, the reference time point is selected as a valley point of the mean over all voltages. Because the valley points are usually close to the end of expiration, the reconstructed $\mathbf{A}_{SB}$ can be interpreted as the conductivity decrease of the lung with respect to the end of expiration. However, other arbitrary reference time points can be used as they do not alter the cardiac-related component in the signals. The spontaneous breathing signals of pixels, i.e. rows of $\mathbf{A}_{SB}$, are also cropped into 15 s segments. The total number of generated segments is 307 200. We refer to this dataset as the *Spontaneous Breathing Dataset*.

3.2. Data from patients

Three patients (indexed as Patient 1–3) were diagnosed with pulmonary embolism by radionuclide perfusion scanning. The perfusion scanning images of the patients are shown in figure 5. Patient 1 was diagnosed to have multiple embolisms in the superior and the inferior lobes of the right lung. Patient 2 was diagnosed to have an embolism in the inferior lobe of the left lung. Patient 3 was diagnosed to have an embolism in the left pulmonary artery.

Immediately after the perfusion scanning, EIT data were collected from the patients according to the following protocol.

1. Attach 16 conductive silica gel electrodes at the level of the xiphoid process of the patient;

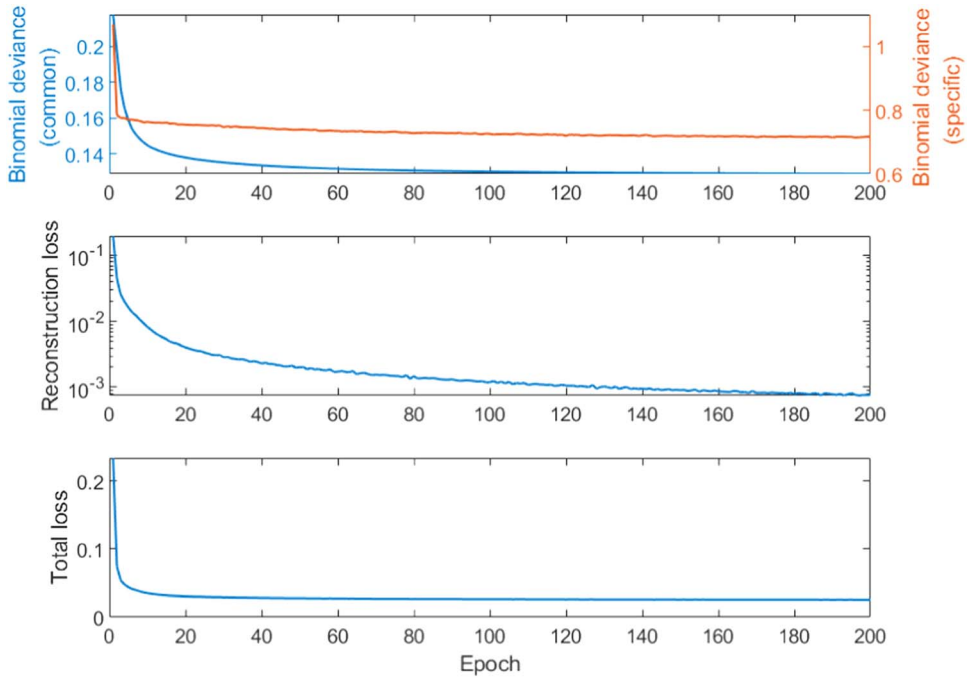

Figure 6. Training losses of the metric network.

2. Collect EIT data from the patient (supine position) during spontaneous breathing for 10 minutes. Perform ECG monitoring at the same time.

The above experiment was approved by the Institution Review Board of TEDA International Cardiovascular Hospital, Tianjin, China (approval no.: (2021)-082 6-6), and informed consent was obtained from the patients. For each patient, a series of spontaneous breathing images were computed using the measured EIT voltages according to (1) (the reference time point is chosen as a valley point of the mean over all voltages).

4. Results of the metric network

We used Keras (Chollet 2015) with Tensorflow (Abadi *et al* 2016) backend to implement our algorithm. One NVIDIA Tesla V100-PCIE-32 GB GPU card was used in the training.

4.1. Training

The data of Subjects 1–30 (183 000 segments in total) were selected from the *Breath-Holding Dataset* to train the metric network. The input dimension is 300×1. According to the layer parameters shown in figure 3, the dimension of the encoded features is 38×128. Then the encoded features are reshaped into a column vector having the dimension 4864×1. According to (2), the reshaped features are then divided into the common part and the specific part. The length of the common part is set as $m = 2432$. Therefore, both the common part and the specific part have the dimension 2432×1.

The training settings are as follows. The hyper-parameters β and γ in (3) are both set as 0.03. An Adam optimizer is used with a learning rate of 0.001. The batch size is set as 1024. In each mini-batch iteration, two batches of data are randomly chosen from the training dataset and fed into the metric network for weights update. The network is trained for 200 epochs. The training losses are shown in figure 6.

4.2. Test

Figure 7 shows some randomly selected signals from the *Breath-Holding Dataset* and their corresponding encoded features. Here, the encoded features are shown as 38×64 dimensional matrices for better visualization. The first column shows the original signals. Although they are all breath-holding signals, their patterns can show significant differences. This may reflect subject-specific cardiac-related activities. The second column shows the learned common features ($\mathbf{f}^c$) of the encoder. Overall, these features are similar to each other while showing some differences in details. These differences are expected because the features are regularized by the

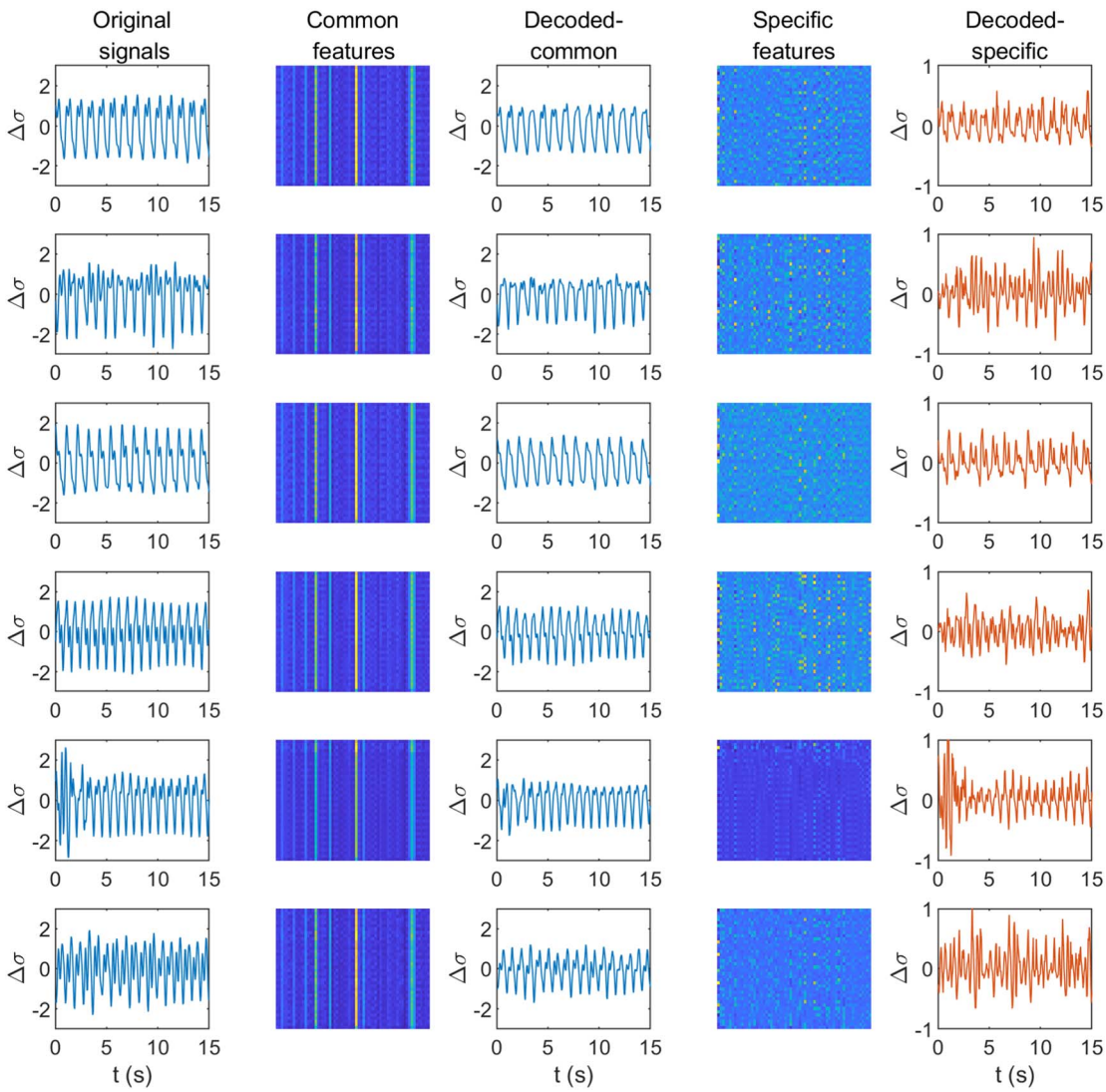

Figure 7. Visualization of the features of the samples randomly chosen from the *Breath-Holding Dataset*. ($\Delta\sigma$: conductivity change.)

reconstruction accuracy of the decoder in the training process. The fourth column shows the learned specific features ($\hat{\mathbf{f}}^s$) of the encoder. It is obvious that these features have significant differences.

To see how common features contribute to the original signals, the learned features are modified by keeping the common part while setting the specific part as zero, and then the modified features are decoded by the decoder, as shown in the third column of figure 7. It can be observed that the decoded signals of the common features are more similar to each other compared with the original signals. Similarly, the contribution of specific features can be studied by keeping the specific part while setting the common part as zero. The decoded signals are shown in the last column of figure 7. It can be observed that the decoded signals of the specific features show significantly larger diversity compared with the decoded signals of the common features. This indicates that the specific features contain distinctive information of the original signals.

The common features of the samples from the *Spontaneous Breathing Dataset* are also visualized, and figure 8 shows some randomly selected results. It is obvious that the features show significant differences to those of the breath-holding signals. The principle of the proposed method is to find the component from a spontaneous breathing signal that matches the features of the breath-holding signals.

5. Results of the separation network

5.1. Training

The training settings of the separation network are as follows. The hyper-parameter λ in the loss (7) is set as 0.001. An Adam optimizer is used with a learning rate of 0.001. The batch size is 128. In each mini-batch iteration, one batch of data is randomly chosen from the *Spontaneous Breathing Dataset* (Subject 1–30) as the

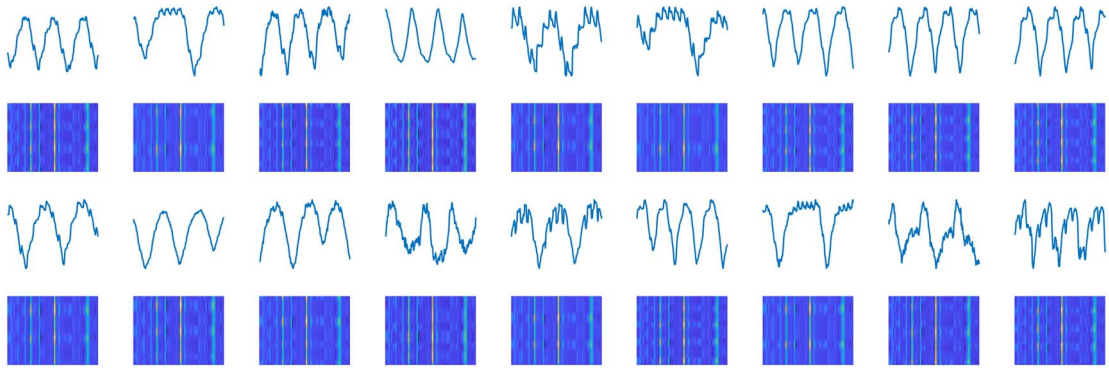

Figure 8. Visualization of the common features of the samples randomly chosen from the *Spontaneous Breathing Dataset*.

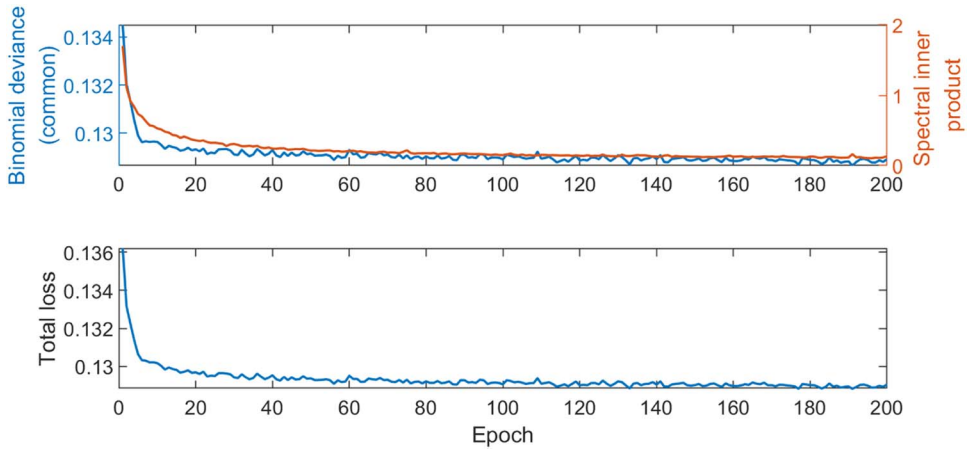

Figure 9. Training losses of the separation network.

Table 1. Settings of the band-pass filters (BPFs) in the baseline methods.

Test data		Heart rate, min^{-1}	Breathing rate, min^{-1}	BPF lower cut-off frequency, min^{-1}	BPF upper cut-off frequency, min^{-1}
Synthetic	no. 1	56	10	50	175
	no. 2	57	18	51	179
	no. 3	81	17	75	255
	no. 4	78	23	72	246
	no. 5	88	14	82	277
	no. 6	84	22	78	265
	no. 7	88	19	82	277
Patient	no. 1	62	19	56	196
	no. 2	78	23	72	252
	no. 3	61	18	55	192

input of the separation network, and another batch of data is randomly chosen from the *Breath-Holding Dataset* (Subject 1–30) as the lower input of the metric network (see figure 2). The network is trained for 200 epochs. The training losses are shown in figure 9.

5.2. Test on synthetic data

In order to quantify the accuracy of separation, the data of Subjects 31–37 (which are not used in the training) are used to generate synthetic data following the flowchart shown in figure 10. Then the synthetic data are tested using the deep feature-domain matching method (DFDM) and the PCA-based method (PCA). Considering that

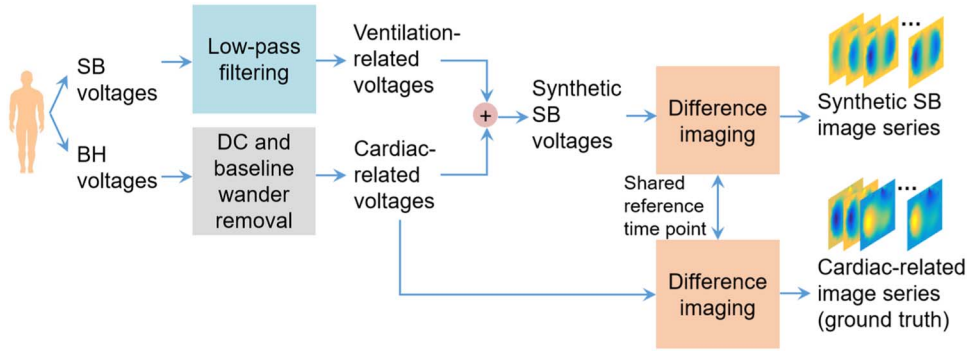

Figure 10. Flowchart of the synthetic data generation. (SB: spontaneous breathing, BH: breath-holding.)

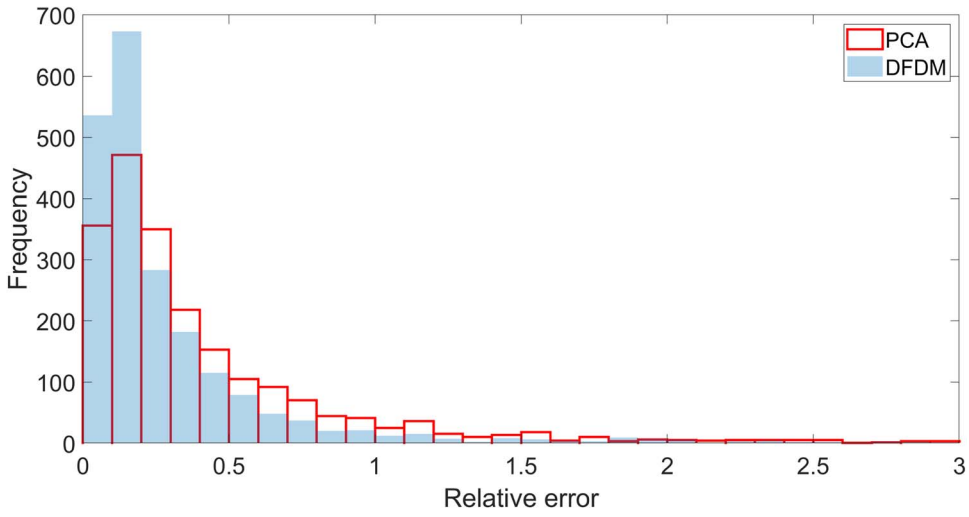

Figure 11. Histograms of the relative separation error of the synthetic data from seven subjects using the PCA-based method (PCA) and the deep feature-domain matching method (DFDM).

Table 2. Mean relative separation errors of the synthetic data using the PCA-based method (PCA) and the deep feature-domain matching method (DFDM).

	no. 1	no. 2	no. 3	no. 4	no. 5	no. 6	no. 7
PCA	0.673	0.647	0.577	0.377	0.309	0.270	0.457
DFDM	0.378	0.343	0.361	0.368	0.211	0.207	0.374

the synthetic data are tailored for the frequency-domain filtering method as they are generated by the direct addition of two components without spectral overlap, we do not use the frequency-domain filtering method for comparison. The PCA-based method is implemented using the configurations described in the original paper (Deibele *et al* 2008). A FIR band-pass filter (BPF) of order 200 is used in this method. The lower cut-off frequency of the BPF is set slightly lower than the heart rate, whereas the upper cut-off frequency is set higher than three times the heart rate. The detailed settings are shown in table 1. In the test, 300 frames of synthetic spontaneous breathing images are used for separation each time. The separation accuracy is quantified by the relative error: $\|\mathbf{a}^{SEP} - \mathbf{a}_*^{SEP}\|/\|\mathbf{a}_*^{SEP}\|$, where $\mathbf{a}^{SEP}$ is the separated image and $\mathbf{a}_*^{SEP}$ is the ground-truth image.

The mean relative errors of the synthetic data are shown in table 2. Compared with the PCA-based method, the proposed method achieves better or comparable (no. 4) separation accuracy. The relative separation errors of the synthetic test samples from all the seven subjects are then combined together and their histograms are shown in figure 11. It can be seen that the proposed method has more image frames, the relative errors of which are quite small (e.g. < 0.2) compared with the PCA-based method. However, both methods have image frames,

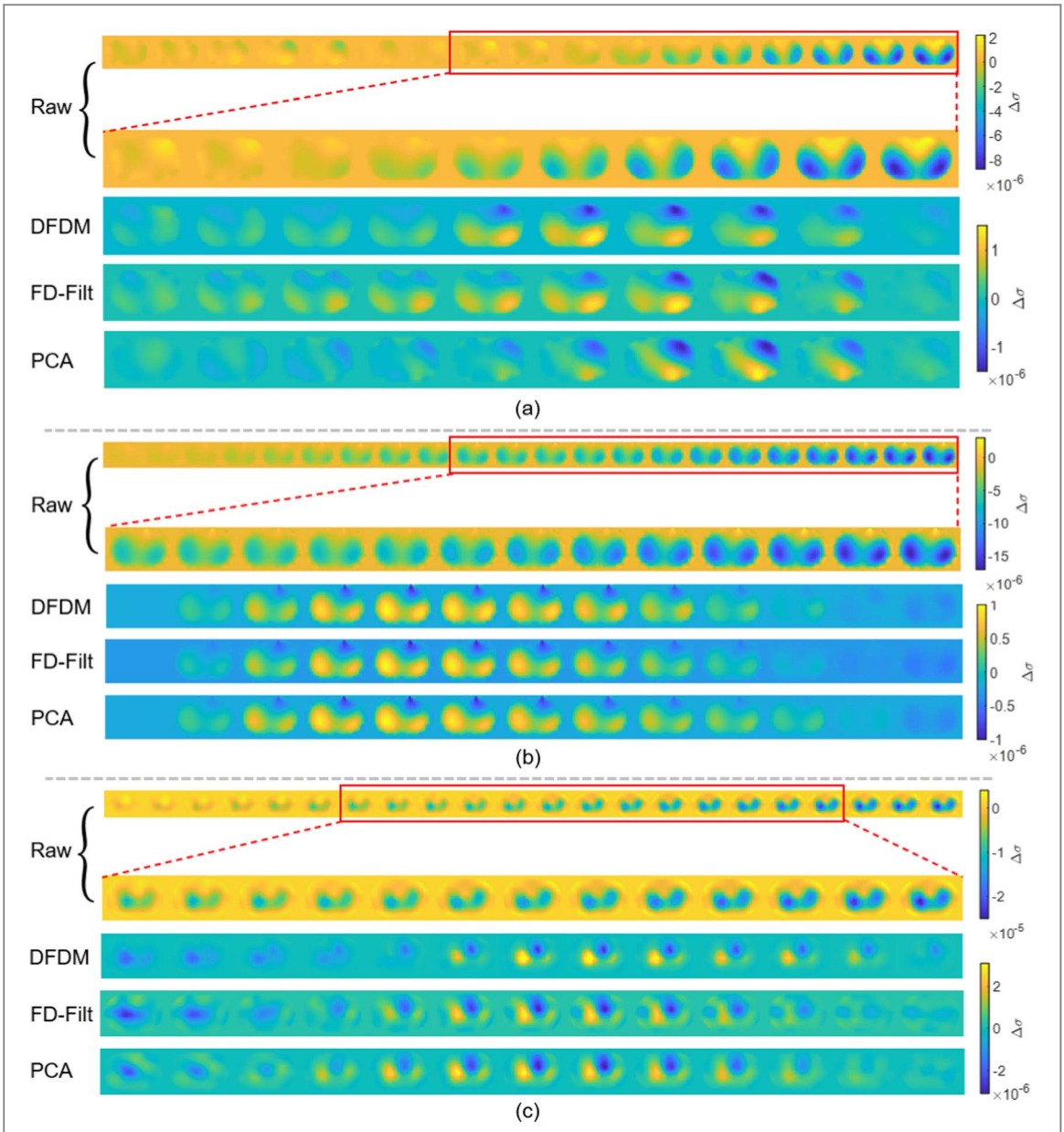

Figure 12. Separation results of (a) Patient 1, (b) Patient 2, and (c) Patient 3 using the deep feature-domain matching method (DFDM), the frequency-domain filtering method (FD-Filt), and the PCA-based method (PCA). Raw images denote the spontaneous breathing image series before separation. For the result of each patient, one cardiac cycle is marked on the raw image series (by the red box) and the separated images correspond to this cardiac cycle. Images are in the digital imaging and communications in medicine (DICOM) orientation.

the relative errors of which are considerably larger (e.g. > 0.7). We check the ground truths of these image frames and find that they almost all have small pixel values, which poses challenges to the separation algorithms.

5.3. Test on patient data

The preparation of the patient data is described in section 3.2. Three hundred frames of spontaneous breathing images were fed into the separation network each time to separate the cardiac-related images. For comparison, the frequency-domain filtering method and the PCA-based method were also applied to 300 frames of images each time. The BPF settings of the two baseline methods are detailed in table 1.

5.3.1. Results of patient 1

The raw images and the separated cardiac-related images of Patient 1 are shown in figure 12(a). Row 1 shows the raw image series that corresponds to half a respiratory cycle starting from the end of expiration. One cardiac cycle (starting from the end of diastole) is marked on the raw image series (by the red box) and the separation results corresponding to this cardiac cycle are shown in row 3–5 of figure 12(a). The conductivity changes in the separated images may be interpreted as follows. During the systole of the heart, the blood is pumped by the heart

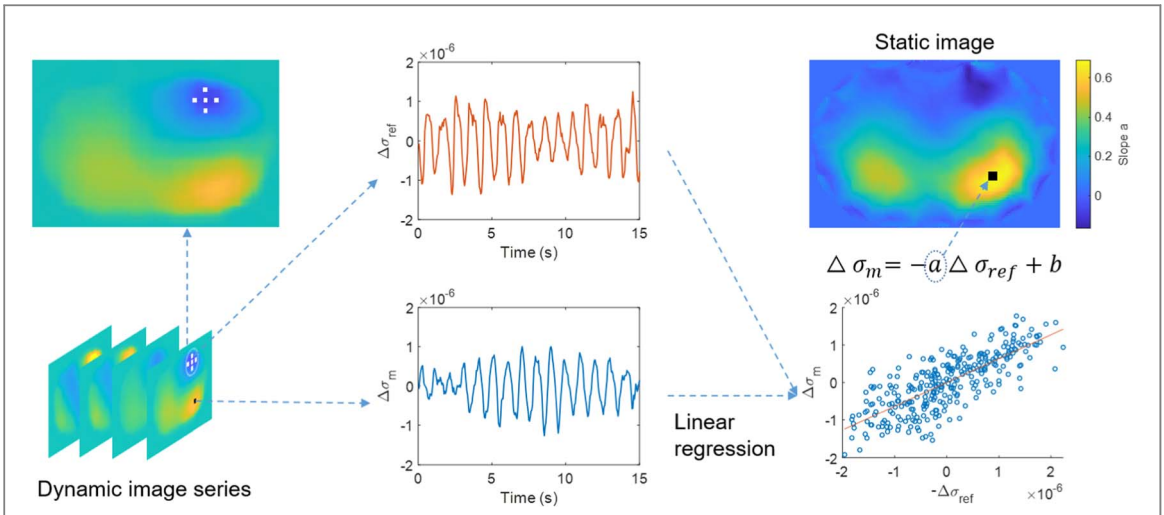

Figure 13. Illustration of the method to generate the static pulsatility image from the dynamic image series. Five pixels (cross-shaped, as shown in the top left of the figure) are selected from the heart region of the dynamic image series and then the temporal signals of these pixels are averaged to obtain the reference signal ($\Delta\sigma_{ref}$). Linear regression is performed between $\Delta\sigma_m$ (the temporal signal of Pixel m) and $-\Delta\sigma_{ref}$. The slope of linear regression is the mth pixel value in the static image.

to the lungs, hence the heart conductivity decreases whereas the lung conductivity increases; during the diastole of the heart, the blood returns from the lungs to the heart, hence the heart conductivity increases whereas the lung conductivity decreases. The separated images by the three methods show approximately synchronous conductivity changes, but the profiles of the conductivity changes exhibit some differences. Overall, the conductivity change of the right lung is weaker than that of the left lung.

To obtain a more stable representation of the cardiac-related conductivity change, a static image is calculated based on a linear regression analysis (Frerichs *et al* 2009) of the above separated image series, as illustrated in figure 13. Here, the term 'static' is used because this image is generated from a series of dynamic images, i.e. the separated image series. Firstly, five pixels (cross-shaped, as shown in the top left of figure 13) are selected from the heart region of the separated cardiac-related image series and then the temporal signals of these pixels are averaged to obtain the reference signal ($\Delta\sigma_{ref}$ in figure 13). Secondly, linear regression is performed between the temporal signal of each pixel and $-\Delta\sigma_{ref}$. Thirdly, the slopes of linear regression are combined as a static image. In this static image, positive pixel values represent that the temporal signals of these pixels are negatively correlated with the temporal signal of the heart, and negative pixel values represent a positive correlation. In addition, the larger the absolute pixel values in the static image, the larger the amplitudes of their corresponding temporal signals.

Furthermore, a lung ROI is selected as pixels whose linear regression slopes exceed 40% of the maximum slope. In this way, the region of the heart is removed as the slopes of heart pixels are negative. We define the static image with the removed heart region as the *static pulsatility image*. Similarly, we can also calculate the static image for ventilation using the corresponding ventilation image series[5]. Specifically, linear regression is performed between the temporal signal of each pixel and the average temporal signal of all the pixels, and then the slopes of linear regression are combined as the static image for ventilation. A 40% maximum slope-lung ROI is also applied. We call this ROI-applied static image for ventilation the *static ventilation image* in the rest of this paper. The calculated static pulsatility and ventilation images of Patient 1 are shown in the first and second rows of figure 14(a), respectively.

To compare between the pixel values of the two lungs, the above static images are further divided into the left half and the right half (by the white dashed lines in figure 14) containing the right lung and the left lung, respectively. Then the percentage of each half is calculated as the sum of pixels in each half divided by the sum of pixels in the whole image. The calculated percentages are shown in each static image of figure 14. In addition, the lung contours of static ventilation and pulsatility images are extracted by an edge-detection algorithm and superimposed to compare their relative positions, as shown in the third row of figure 14(a).

For Patient 1, the static ventilation images of the three separation methods are almost the same, and the percentages show balanced ventilation of the left and right lungs. The static pulsatility images of PCA and FD-Filt show large connections between the left and right lungs, which are not anatomically plausible. In all the static

[5] The ventilation image series of the proposed method and frequency-domain filtering method are obtained by subtracting the cardiac-related image series from the raw spontaneous breathing image series, whereas the ventilation image series of the PCA-based method is obtained following the method in Deibele *et al* (2008).

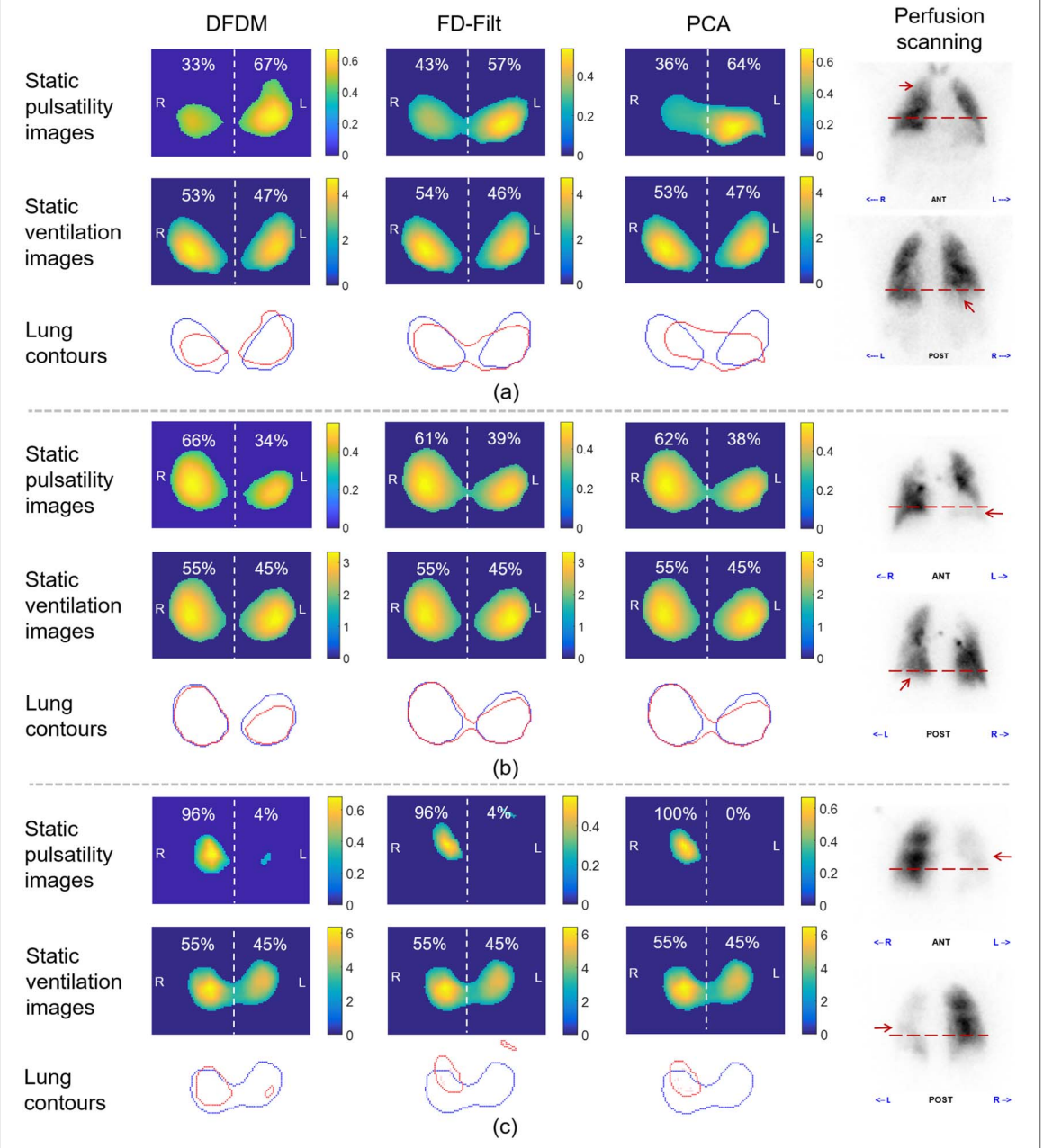

Figure 14. Static pulsatility and ventilation images of (a) Patient 1, (b) Patient 2, and (c) Patient 3 calculated based on a linear regression analysis of the separated image series. In the rows of lung contours, the lung contours of static ventilation images are plotted in blue, whereas the lung contours of static pulsatility images are plotted in red. EIT images are in DICOM orientation. The last column shows the radionuclide perfusion scanning images (in coronal view) where the red dashed lines represent the level of the electrodes used for EIT data collection. (DFDM: deep feature-domain matching method, FD-Filt: frequency-domain filtering method, PCA: PCA-based method.)

pulsatility images, the percentage of the right lung is smaller than that of the left lung, although it is more significant for DFDM and PCA. As pulsatility is related to perfusion, this is likely to indicate that the perfusion of the right inferior[6] lung is weaker than that of the left inferior lung, which agrees with the diagnosis of the radionuclide perfusion scanning (pulmonary embolism in the inferior lobe of the right lung).

5.3.2. Results of patient 2

The separation results of Patient 2 are shown in figure 12(b); the corresponding static images are shown in figure 14(b). The static ventilation images of all three separation methods are almost the same, and the percentages of the right and the left lungs have no significant differences, indicating balanced ventilation for the two lungs. The static pulsatility images match the static ventilation images well in terms of the lung position (see

[6] The EIT imaging domain is at the inferior lung because the electrodes are attached at the level of the xiphoid process during the measurement.

the third row of figure 14(b)). In the static pulsatility image of DFDM, the percentage of the right lung is significantly larger than that of the left lung, and the percentage difference between the right and left lungs is 32%. The static pulsatility images of FD-Filt and PCA are similar, and the percentage differences between the right and left lungs are 22% and 24%, respectively. In summary, the static pulsatility images are likely to indicate embolism in the left lung, which is consistent with the diagnosis of the radionuclide perfusion scanning (pulmonary embolism in the inferior lobe of the left lung).

5.3.3. Results of patient 3

The separation results of Patient 3 are shown in figure 12(c); the corresponding static images are shown in figure 14(c). Similar to the results of Patient 1 and 2, the static ventilation images of the three methods have little difference, and balanced ventilation is observed for the left and right lungs. In contrast, all the static pulsatility images show an almost missing left lung. For DFDM, the lung positions in the static ventilation and pulsatility images match well; for FD-Filt and PCA, the lung position in the pulsatility image deviates from the lung position in the ventilation image (see the third row of figure 14(c)). In summary, the static pulsatility images are likely to indicate an embolism in the left lung, which agrees with the diagnosis of the radionuclide perfusion scanning (pulmonary embolism in the left lung).

6. Discussion

This work proposes a cardiac-related component separation method based on the criterion of feature-domain matching of the separated signals and breath-holding signals. The method is comprised of training a metric network for exploiting the essential features of the breath-holding signals and training a separation network constrained by the feature-matching loss. Once trained, the separation network is ready to process the spontaneous breathing image series of new subjects.

The training of the networks is designed to be unsupervised considering that the ground-truth cardiac-related component of a spontaneous breathing image series is unknown. The networks are trained using datasets consisting entirely of real-world human data. The technical path of using synthetic data to train a supervised separation network is not adopted because it is difficult to model the cardiac-related component since its mechanism has not been fully understood.

In this paper, lightweight CNNs are used to implement the metric and the separation networks. Therefore, training of the networks is fast (5 minutes and 12 minutes for the training of the metric and the separation networks, respectively, using the hardware and software configurations described in section 4). In the inference, the separation network is able to process a batch of input signals in parallel and gives results quickly.

In our experiments, breath-holding signals are obtained from several healthy subjects. However, test results show that the feature space learned from these breath-holding signals is applicable to the separation of new data from both healthy subjects and patients. In the future, more breath-holding signals from diverse subjects (including patients) may be added to the *Breath-Holding Dataset* if available. In this way, more general features of breath-holding signals are expected to be learned.

In the preparation of the *Breath-Holding Dataset*, EIT data are required to be collected from the subjects during a breath-holding process. In the experiments of this paper, the breath-holding process is designed to be 30 s, which may be too long for some spontaneous breathing subjects, especially patients. In fact, this length of time can be reduced because our networks only need inputs of 15 s signal segments. Another solution is to collect training data from mechanically ventilated patients for whom a respiratory pause can be realized by the ventilator. In addition, an averaged ROI for all the subjects is generated to select the signals from the heart and the lungs. This process can be improved by generating individualized ROIs for different subjects.

In this work, the EIT measurement system has a sampling frequency of 20 Hz, which is not enough to capture all the details of cardiac-related signals for subjects with high heart rates. However, our method is independent of the sampling frequency of the measurement system. In the future, we will use a measurement system with a higher sampling frequency.

Static images are generated based on a linear regression analysis of the separated dynamic image series. Potential limitations of this method are as follows: (1) the linear regression slope is related to both the amplitude and the phase angle of the signals. As a result, it is difficult to completely differentiate between the effect of the amplitude and the effect of the phase angle; (2) in the lung ROI selection during the generation of static pulsatility images, it is difficult to determine the lung–heart interface because the heart tends to diffuse to the lung region (especially the left lung) in the images. This is essentially caused by the low resolution of EIT image reconstruction.

For patients whose cardiac-related signals are nonstationary, it is challenging to separate the cardiac-related signals. In this case, there are two possible solutions. First, perform the separation when the patients are

relatively stable. Second, augment the training dataset by introducing variabilities (e.g. in shape, frequency, and amplitude) to the training samples. In this way, the features of nonstationary signals may be learned. In the future, we will test the proposed method on more patients and further improve its performance on challenging clinical data.

7. Conclusion

This proof-of-concept study demonstrates the feasibility to use breath-holding signals to guide the separation of cardiac-related components from a spontaneous breathing image series of chest EIT through deep feature-domain matching. The metric network can effectively learn the common features of breath-holding signals. The effectiveness of the separation network is demonstrated by synthetic data and preliminary patient data. Future improvements of this method include enlarging the training dataset as well as validating the method using more clinical data.

Acknowledgments

This work was supported by the National Science Foundation of China under Grant 61 971 263, the National Key R&D Program of China under Grant 2018YFC0603604, and the Tsinghua Precision Medicine Foundation.

Ethical Statement

The healthy human experiment was approved by the Institution Review Board of Tsinghua University (approval no: 20 200 044). The patient experiment was approved by the Institution Review Board of TEDA International Cardiovascular Hospital (approval no: (2021)-082 6-6). All the human experiments were conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. Written informed consent was obtained from all the human participants.

Appendix A. Layer-normalization and cosine-similarity layers

The layer-normalization layer performs the Z-score normalization to the input signal:

$$\bar{\mathbf{x}} = \frac{1}{\sigma_{\mathbf{x}}}(\mathbf{x} - \mu_{\mathbf{x}}\mathbf{1}), \tag{A.1}$$

where $\mathbf{x}$ is the input signal, $\bar{\mathbf{x}}$ is the normalized signal, $\mu_{\mathbf{x}}$ and $\sigma_{\mathbf{x}}$ are the mean and the standard deviation of $\mathbf{x}$, respectively, and $\mathbf{1}$ is a vector with all its elements equal to 1 and with the same shape as $\mathbf{x}$.

The cosine-similarity layer calculates the cosine similarity between two feature vectors $\mathbf{f}_1$ and $\mathbf{f}_2$:

$$\mathcal{S}_{cos}(\mathbf{f}_1, \mathbf{f}_2) = \frac{\mathbf{f}_1^T \mathbf{f}_2}{\sqrt{\mathbf{f}_1^T \mathbf{f}_1}\sqrt{\mathbf{f}_2^T \mathbf{f}_2}}. \tag{A.2}$$

In this work, the above two layers are implemented by using the custom layers in Keras (with Tensorflow backend).

Appendix B. Design of the spectral inner product loss

We start from designing the following loss:

$$\mathcal{L}_{sip}^{0} = \frac{\langle \mathrm{PSD}(\mathbf{x}^{SB} - \mathcal{G}(\mathbf{x}^{SB})), \mathrm{PSD}(\mathcal{G}(\mathbf{x}^{SB}))\rangle}{\sum \mathrm{PSD}(\mathcal{G}(\mathbf{x}^{SB})) + \epsilon}. \tag{B.1}$$

The numerator of (B.1) is the inner product of the PSDs of the separated ventilation- and cardiac-related components. A zero inner product means that the PSDs of the two components are orthogonal to each other, indicating a good separation between the two components. The denominator of (B.1) (energy of the separated signal) is introduced to prevent the separation network from giving the trivial result, which matches the features of breath-holding signals but has a close-to-zero amplitude.

In a practical spontaneous breathing signal, the cardiac-related component is often much weaker (e.g. one or two orders of magnitude weaker in amplitude) than the ventilation-related component. As a result, it is difficult to minimize (B.1), which is relatively flat around the desired cardiac-related signal $\mathbf{x}_*^{SEP}$. To alleviate this

problem, we propose to replace the signals in (B.1) with their time derivatives, leading to the modified loss 6. The consideration behind the above modification is that, when we take the time derivative of a signal, its PSD is multiplied by the square of the frequency; since the cardiac-related components tend to appear in the higher frequencies of the spectrum, this gives larger weights to the cardiac-related components. As a result, the modified loss (6) can provide larger gradients around $\mathbf{x}_*^{SEP}$, which alleviates the ill-posedness of the optimization process.

In our implementation of 6, the time derivative is approximated by the forward difference and the PSD of a discrete temporal signal $\mathbf{x}$ is estimated using fast Fourier transform (FFT) (Youngworth *et al* 2005):

$$\mathrm{PSD}(\mathbf{x}) \propto |\mathrm{FFT}(\mathbf{x})|^2, \tag{B.2}$$

where $|\cdot|$ is the modulus of the complex FFT result.

Appendix C. PCA noise reduction

As described in section 3.1, the measured breath-holding voltages are preprocessed by two steps. Assume the voltages after the first step of processing (i.e. band-pass filtering) can be arranged into a matrix $\mathbf{D}$. The size of $\mathbf{D}$ is $M \times N$, where M is the number of voltages in a frame of data and N is the number of frames. Each row of $\mathbf{D}$ has a zero mean because it has been band-pass filtered. Therefore, the covariance matrix of rows of $\mathbf{D}$ can be calculated as $\mathbf{C} = \frac{1}{N}\mathbf{D}\mathbf{D}^T$. The eigen decomposition of $\mathbf{C}$ can be written as $\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$, where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M]$ is an orthogonal matrix, the columns of which are the eigenvectors of $\mathbf{C}$ and $\mathbf{\Lambda}$ is a diagonal matrix. Then define a matrix $\mathbf{P}$ composed of the first K columns of $\mathbf{Q}$, i.e. $\mathbf{P} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K]$. We set $K = 4$ in this paper. Then K time-domain templates can be calculated as $\mathbf{T} = \mathbf{D}^T\mathbf{P}$. Each column of $\mathbf{T}$ is a time-domain template. Using these templates, we can perform a time-domain filtering of $\mathbf{D}$ by solving the following least-squares problem:

$$\min_{\mathbf{W}} \|\mathbf{T}\mathbf{W} - \mathbf{D}^T\|^2, \tag{C.1}$$

where $\mathbf{W}$ is the coefficient matrix. The above problem has a closed form solution: $\mathbf{W} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{D}^T$. Finally, the denoised voltage matrix can be written as

$$\tilde{\mathbf{D}} = (\mathbf{T}\mathbf{W})^T = \mathbf{D}\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T. \tag{C.2}$$

ORCID iDs

Maokun Li https://orcid.org/0000-0002-7258-6413

References

Abadi M *et al* 2016 Tensorflow: A system for large-scale machine learning *XII {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* pp 265–83 (https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi)

Adler A and Boyle A 2017 Electrical impedance tomography: Tissue properties to image measures *IEEE Trans. Biomed. Eng.* **64** 2494–504

Adler A *et al* 2009 Greit: a unified approach to 2d linear eit reconstruction of lung images *Physiol. Meas.* **30** S35–55

Bayford R 2006 Bioimpedance tomography (electrical impedance tomography) *Annu. Rev. Biomed. Eng.* **8** 63–91

Borges J B *et al* 2012 Regional lung perfusion estimated by electrical impedance tomography in a piglet model of lung collapse *J. Appl. Physiol.* **112** 225–36

Carlisle H R *et al* 2010 Regional distribution of blood volume within the preterm infant thorax during synchronised mechanical ventilation *Intensive Care Medicine* **36** 2101–8

Cheney M *et al* 1990 Noser: An algorithm for solving the inverse conductivity problem *Int. J. Imaging Syst. Technol.* **2** 66–75

Cheney M *et al* 1999 Electrical impedance tomography *SIAM Rev.* **41** 85–101

Cheng K-S *et al* 2022 Separation of heart and lung-related signals in electrical impedance tomography using empirical mode decomposition *Current Medical Imaging* **18** 1396–415

Chollet F *et al* 2015 Keras (https://keras.io)

de Castro Martins T *et al* 2019 A review of electrical impedance tomography in lung applications: Theory and algorithms for absolute images *Annual Reviews in Control* **48** 442–71

Deibele J M *et al* 2008 Dynamic separation of pulmonary and cardiac changes in electrical impedance tomography *Physiol. Meas.* **29** S1–14

Eyuboglu B M and Brown B H 1988 Methods of cardiac gating applied potential tomography *Clin. Phys. Physiol. Meas.* **9** 43–8

Fagerberg A *et al* 2009a Electrical impedance tomography applied to assess matching of pulmonary ventilation and perfusion in a porcine experimental model *Critical Care* **13** 1–12

Fagerberg A *et al* 2009b Monitoring pulmonary perfusion by electrical impedance tomography: an evaluation in a pig model *Acta Anaesthesiologica Scandinavica* **53** 152–8

Frerichs I *et al* 2002 Regional lung perfusion as determined by electrical impedance tomography in comparison with electron beam ct imaging *IEEE Trans. Med. Imaging* **21** 646–52

Frerichs I *et al* 2009 Assessment of changes in distribution of lung perfusion by electrical impedance tomography *Respiration* **77** 282–91

Grant C A *et al* 2011 Measurement of ventilation and cardiac related impedance changes with electrical impedance tomography *Critical Care* **15** 1–9

Hastie T *et al* 2005 The elements of statistical learning: data mining, inference, and prediction *The Mathematical Intelligencer* **27** 83–5
Hentze B *et al* 2021 A model-based source separation algorithm for lung perfusion imaging using electrical impedance tomography *Physiol. Meas.* **42** 084001
Hoffer E and Ailon N 2015 Deep metric learning using triplet network *Lecture Notes in Computer Science* **9370** 84–92
Holder D S 2004 *Electrical Impedance Tomography: Methods, History and Applications* (Boca Raton, FL: CRC Press)
Isaacson D *et al* 2006 Imaging cardiac activity by the d-bar method for electrical impedance tomography *Physiol. Meas.* **27** S43–50
Kircher M *et al* 2020 Regional lung perfusion analysis in experimental ards by electrical impedance and computed tomography *IEEE Trans. Med. Imaging* **40** 251–61
Krivoshei A *et al* 2008 Decomposition method of an electrical bio-impedance signal into cardiac and respiratory components *Physiol. Meas.* **29** S15–25
Kunst P W A *et al* 1998 Ventilation and perfusion imaging by electrical impedance tomography: A comparison with radionuclide scanning *Physiol. Meas.* **19** 481–90
Leonhardt S and Lachmann B 2012 Electrical impedance tomography: the holy grail of ventilation and perfusion monitoring? *Intensive Care Medicine* **38** 1917–29
Liu D *et al* 2019 A parametric level set-based approach to difference imaging in electrical impedance tomography *IEEE Trans. Med. Imaging* **38** 145–55
Nguyen D T *et al* 2012 A review on electrical impedance tomography for pulmonary perfusion imaging *Physiol. Meas.* **33** 695–706
Noordegraaf A V *et al* 1998 Pulmonary perfusion measured by means of electrical impedance tomography *Physiol. Meas.* **19** 263–73
Pikkemaat R and Leonhardt S 2010 Separation of ventilation and perfusion related signals within eit-data streams *J. Phys.: Conference Series* **224** 012028
Rahman T *et al* 2013 Extraction of cardiac and respiration signals in electrical impedance tomography based on independent component analysis *J. Electrical Bioimpedance* **4** 38–44
Silva O L *et al* 2017 Influence of current injection pattern and electric potential measurement strategies in electrical impedance tomography *Control Eng. Pract.* **58** 276–86
Sohn K 2016 Improved deep metric learning with multi-class n-pair loss objective *NIPS'16 Proceedings of the XXX International Conference on Neural Information Processing Systems* **29** 1857–65
Song H O *et al* 2016 Deep metric learning via lifted structured feature embedding *Proceedings of the IEEE conference on computer vision and pattern recognition* pp 4004–12 (https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Song_Deep_Metric_Learning_CVPR_2016_paper.html)
Sun X *et al* 2017 Separation of cardiac-and ventilation-related signals within electrical impedance tomography data based on multi-dimensional ensemble empirical mode decomposition *IFAC-PapersOnLine* **50** 4436–41
Tikhonov A 1963 Solution of incorrectly formulated problems and the regularization method *Soviet Math.* **4** 1035–8
Wang X *et al* 2019 Multi-similarity loss with general pair weighting for deep metric learning *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 5022–30 (https://openaccess.thecvf.com/content_CVPR_2019/html/Wang_Multi-Similarity_Loss_With_General_Pair_Weighting_for_Deep_Metric_Learning_CVPR_2019_paper.html)
Yi D *et al* 2014 Deep metric learning for person re-identification *Proceedings of the 2014 XXII International Conference on Pattern Recognition* pp 34–9 (https://ieeexplore.ieee.org/abstract/document/6976727/)
Youngworth R N *et al* 2005 An overview of power spectral density (psd) calculations *Optical Manufacturing and Testing VI* **5869** 206–16
Zadehkoochak M *et al* 1992 Pulmonary perfusion and ventricular ejection imaging by frequency domain filtering of eit images *Clin. Phys. Physiol. Meas.* **13** 191–6