Edinburgh Research Explorer

Enhanced Multi-Scale Feature Cross-Fusion Network for Impedance-optical Dual-modal Imaging
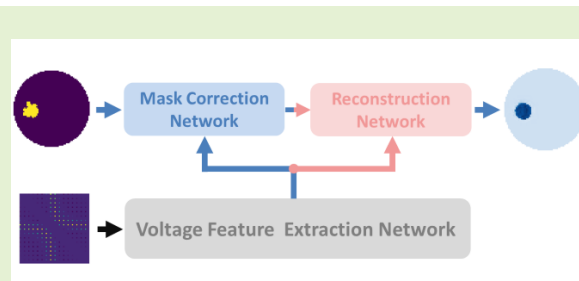
Enhanced Multi-Scale Feature Cross-Fusion Network for Impedance-optical Dual-modal Imaging

Zhe Liu, *Graduate Student Member, IEEE*, Renjie Zhao, Graham Anderson, Pierre Bagnaninchi, Yunjie Yang, *Member, IEEE*

Abstract—The intrinsic issue of low spatial resolution of Electrical Impedance Tomography (EIT) is a long-standing challenge that hinders the capability of performing quantitative analysis based on EIT image. Our recent work demonstrates an impedance-optical dual-modal imaging framework and a deep learning model named Multi-Scale Feature Cross Fusion Network (MSFCF-Net) to realize information fusion and high-quality EIT image reconstruction. However, this framework's performance is limited by the accuracy of the mask image obtained from an auxiliary imaging modality. This paper further proposes a two-stage deep neural network, which is the enhanced version of the MSFCF-Net (named En-MSFCF-Net), to automatically improve mask image and conduct information fusion and image reconstruction. Compared to MSFCF-Net, En-MSFCF-Net demonstrates the superior ability to correct the inaccurate mask image, leading to a more accurate conductivity estimation. Furthermore, the En-MSFCF-Net also maintains the best shape preservation and conductivity prediction accuracy among given learning-based and model-based algorithms. Both qualitative and quantitative results indicate that En-MSFCF-Net could make dual-modal imaging more robust in real-world situations.

Index Terms—Dual-modal imaging, electrical impedance tomography, mask image correction, image reconstruction, deep learning, information fusion

ABBREVIATIONS

En-MSFCF-Net	Enhanced Multi-Scale Feature Cross Fusion Network
MSFCF-Net	Multi-Scale Feature Cross Fusion Network
VFEN	Voltage Feature Extraction Network
MCN	Mask Correction Network
RN	Reconstruction Network
EIM	Electrical Impedance Map
SEIM	Sparse Electrical Impedance Map
DMFF	Dual-Modal Feature Fusion Modules. There are two types of DMFF and they are labelled as DMFF-V1 and DMFF-V2
MSFF	Multi-Scale Feature Fusion Modules

Z. Liu, R. Zhao and Y. Yang are with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, UK, EH9 3JL (Corresponding author: Yunjie Yang, e-mail: y.yang@ed.ac.uk).

G. Anderson and P. Bagnaninchi are with the Centre for Regenerative Medicine, Institute for Regeneration and Repair, The University of Edinburgh, Edinburgh EH16 4UU. (e-mail: Pierre.Bagnaninchi@ed.ac.uk).

Manuscript received xxx, 2022.

I. INTRODUCTION

Electrical Impedance Tomography (EIT) is a tomographic imaging modality that can reconstruct the conductivity distribution of the sensing region through current injection and boundary voltage measurements [1]–[3]. In recent years, its advantages of non-destructiveness, non-radiation and high temporal resolution have made EIT a promising method for continuous monitoring of industrial processes or tissue status [4] [5]. For instance, cell viability can be inferred by measuring the conductivity distribution with miniaturized EIT sensors [6]. However, a wider adoption of EIT is limited by its image quality, which is one of the crucial problems to be solved.

In the past decades, many model-based EIT image reconstruction algorithms have been reported to improve the EIT image quality. A prevailing type is based on regularization, such as sparse regularization [7] [8], Total Variation (TV) regularization [9]–[11], and Group Sparsity regularization [12] [13]. These methods introduce carefully-designed global or local prior information into the EIT inversion, attempting to find a solution close to the ground truth. They are often based on the linearized EIT model, which intrinsically introduces
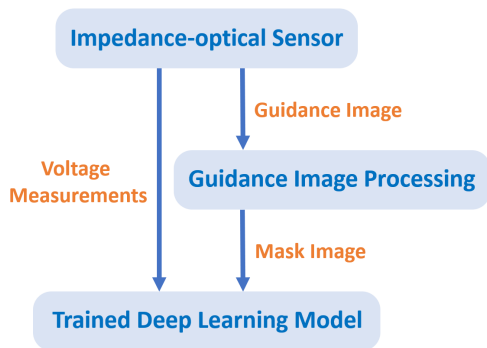
Fig. 1. Impedance-optical dual-modal imaging framework.

modelling errors. In addition, the system (or sensitivity) matrix of the EIT linearized model is ill-conditioned, making the corresponding inverse problem ill-posed. Recently, due to the fast development of computational techniques and artificial intelligence, deep-learning-based methods have been introduced to solve the EIT image reconstruction problem, which can be generally divided into three categories. The first type is the pure learning-based methods, which directly learn the non-linear inversion of EIT in an end-to-end manner [14]–[17]. In addition to the network design, its performance is mainly determined by the quality of the dataset. The second type combines the learning algorithm with model-based algorithm [18]–[20]. These methods utilize the non-linear fitting of the learning approach while retaining a certain degree of interpretability. The last type treats deep learning as a post-processing algorithm [21]. When the prior information of the imaging targets is sufficient, learning-based methods are usually superior to model-based methods in background artefact suppression, shape preservation, and conductivity prediction accuracy. However, the EIT image quality improvement is still circumscribed as the aforementioned methods are mainly single-modal.

To further promote EIT development, multi-modal methods have also been proposed, which incorporate complementary information from other imaging modalities into EIT inversion. However, there is limited literature investigating multi-modal methods. For instance, Gong et al. proposed to incorporate structural information from CT or preliminary EIT reconstructions into the EIT inversion process using group lasso [22]. Liu et al. reported a group lasso based dual-modal reconstruction method and their grouping method is based on the semantic segmentation of the prior image [23]. The segmentation algorithm needs to be carefully chosen and tuned for a specific application, and its complexity is even higher than the reconstruction algorithm in some cases. Liu et al. further reported a kernel-based, segmentation-free dual-modal image reconstruction algorithm, which can alleviate the burden of selecting and tuning the segmentation algorithm meanwhile preserving inclusion's structure [24]. In addition, Li et al. combined CT with EIT through Cross Gradient regularization [25] and Liang et al. integrated ultrasound image into EIT [26] [27]. Apart from these model-based image reconstruction algorithms, Liu et al. presented a pioneering

study on learning-based dual-modal EIT imaging [28]. These multi-modal methods show the potential of improving EIT image quality by introducing auxiliary imaging information.

In our previous work [28], an impedance-optical dual-modal imaging framework (see Fig. 1) was proposed and demonstrated noticeable improvements in the quality of EIT images. Specifically, it is proved that the accuracy of the mask image is crucial to the image reconstruction quality, and the accurate mask images cannot always be acquired in practice due to the measurement error of the optical sensor, the imaging circumstance, and the imperfection of the guidance image processing algorithm. Thus, a method to ensure the mask images as accurate as possible is expected. In later analysis, we will prove that the adverse impact on the target structure originating from the inaccurate mask image can not be satisfactorily addressed based on the MSFCF-Net trained with inaccurate mask images. Therefore, a two-stage deep neural network which is the enhanced version of the MSFCF-Net (En-MSFCF-Net) is proposed to compensate for the inaccuracy of the mask image and perform information fusion and image reconstruction. We call En-MSFCF-Net a two-stage network because its functionality can be divided into two sequential stages. The advantages of the proposed method is summarised as:

- En-MSFCF-Net introduces a new network structure for dual-modal EIT image reconstruction, separating the inaccurate mask correction and conductivity estimation into two sequential operations. Compared with MSFCF-Net, which corrects the inaccurate mask and estimates conductivity simultaneously, such a function-separated network can better recover the structures when trained with inaccurate mask images.
- Compared with the given model-based and learning-based algorithm, the proposed En-MSFCF-Net can generate the results with the best structure preservation and most accurate conductivity estimation.

The remainder of this paper is organized as follows. Section II introduces the principle of EIT and the new En-MSFCF-Net. Section III describes the dataset and network training. Section IV presents simulation and experimental results. Section V concludes this paper and discusses the future work.

II. METHODOLOGY

A. Principle of EIT

EIT comprises two sub-problems, i.e. the forward and inverse problems. The relationship between conductivity $\sigma \in \mathbb{R}^n$ and boundary voltage measurement $V \in \mathbb{R}^m$ is described as follows:

$$V = \mathrm{F}(\sigma), \qquad (1)$$

where F is the nonlinear forward mapping. The inverse problem of EIT can be expressed as:

$$\sigma = \mathrm{F}^{-1}(V), \qquad (2)$$

where F^{-1} is the inverse mapping of F. In this study, $m = 104$ and $n = 3228$. As we conduct the time-difference imaging, we focus on the following modified inverse mapping $\mathcal{F}$:

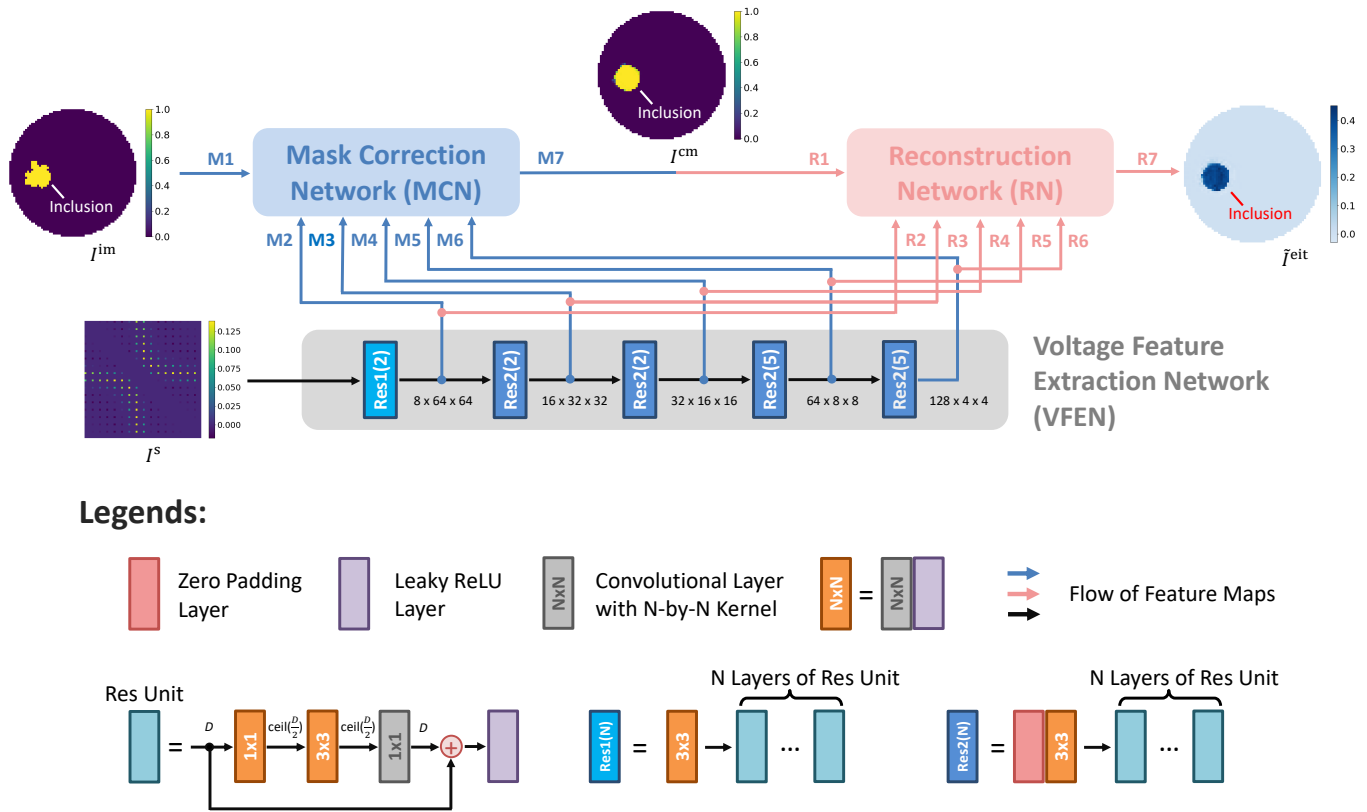$$\Delta\sigma = \mathcal{F}(\Delta V), \qquad (3)$$

Fig. 2.   Architecture of the En-MSFCF-Net. 'Res Unit' does not change the input feature map's height, width and the number of channels. D denotes the number of channels of the feature map and ceil($\cdot$) represents round up function. Res1(N) only change the number of channels of the input feature map while keeping its height and width. Zero padding layer in Res2(N) applies left and top padding for the feature map, and the stride of the convolutional layer after it is set as 2. Thus, Res2(N) can change the number of channels of the input feature map while its height and width will be reduced to half. The size of the feature map after each residual block is illustrated in VFEN.
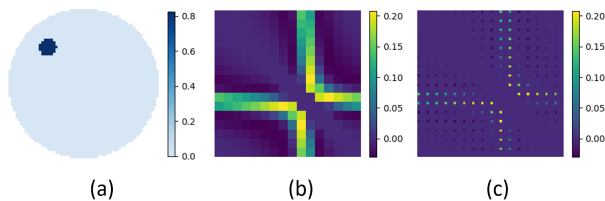

Fig. 3.   Examples of (b) EIM and (c) SEIM. (a) is the true conductivity image corresponding to (b) and (c).

where $\Delta\sigma \in \mathbb{R}^n$ and $\Delta V \in \mathbb{R}^m$ are defined as:

$$\Delta\sigma = -(\sigma_o - \sigma_r)./\sigma_r, \tag{4}$$

$$\Delta V = (V_o - V_r)./V_r, \tag{5}$$

where ./ represents element-wise division. Subscripts 'o' and 'r' represent the corresponding quantity at the observation time point and the reference time point.

B. Enhanced Muti-Scale Feature Cross Fusion Network (En-MSFCF-Net)

The architecture of En-MSFCF-Net is illustrated in Fig. 2. It consists of three subnetworks, i.e., Voltage Feature Extraction Network (VFEN), Mask Correction Network (MCN) and Reconstruction Network (RN). The VFEN extracts latent information of the ΔV and the extracted high-dimensional features are shared with MCN and RN. VFEN adopts the same residual blocks (Res1(N) and Res2(N)) as those in the backbone networks of the MSFCF-Net. To reduce the network complexity, the number of residual blocks and residual units (Res Unit) in an individual residual block is reduced. Compared with MSFCF-Net, the voltage data ΔV is rearranged as a 64×64 matrix named Sparse Electrical Impedance Map (SEIM) rather than a 104-element vector. VFEN eliminates the fully connected layers in MSFCF-Net and further reduces the network's complexity. For the circular 16-electrode EIT sensor, Hu et al. proposed Electrical Impedance Map (EIM) of the size 16×16, which is more suitable for learning-based EIT reconstruction as it uses the symmetric property of the EIT sensor and naturally integrates more spatial information [15]. To better adapt the shape of the voltage data to the network input, we enlarge the original EIM and define the aforementioned SEIM as $I^s \in \mathbb{R}^{64 \times 64}$. $I^s_{4(i-1)+1,4(j-1)+1}$ denotes the voltage data value on the electrode pair (e_j, e_{j+1}) when injecting currents into electrode pair (e_i, e_{i+1}), where i or j = 1, 2, ..., 16, $e_{16+1} \triangleq e_1$. The voltage data is not collected on the activation electrodes. According to the reciprocal theory [29], we define $I^s_{4(j-1)+1,4(i-1)+1} = I^s_{4(i-1)+1,4(j-1)+1}$. The rest elements of I^s are set as zeros. An example of EIM and SEIM is illustrated in Fig. 3. It is
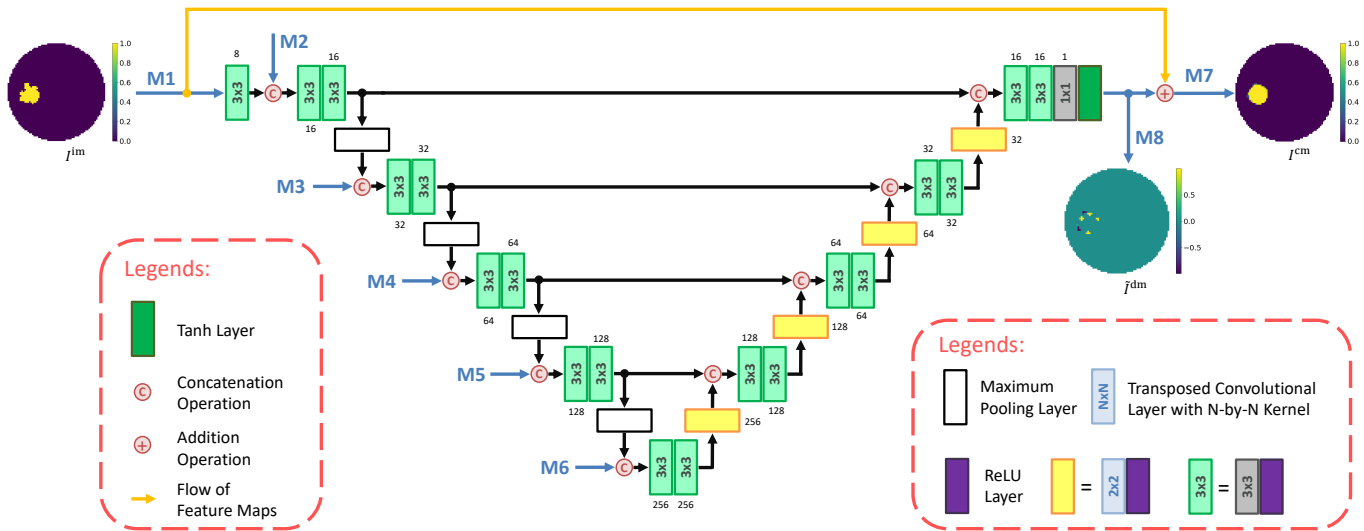
Fig. 4. Architecture of the MCN. The maximum pooling layer reduces the height and width of the feature map to half. The black number close to each neural network component denotes the number of channels of the feature map output by this component. The meaning of un-annotated components is the same as those in Fig. 2.
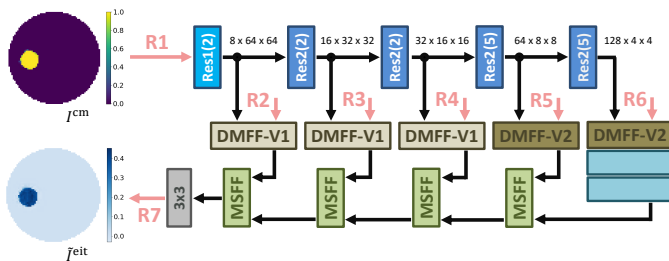

Fig. 5. Architecture of the RN. The black numbers represent the size of the feature map after each residual block. Except to DMFF-V1, DMFF-V2 and MSFF, other network components are annotated in Fig. 2.
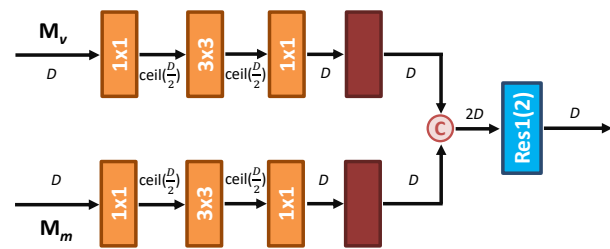

Fig. 6. Architecture of the DMFF. The dark red block means the attention modules described in [31]. It represents a channel attention module followed by a spatial attention module for DMFF-V1 and denotes a channel attention module for DMFF-V2. M_v stands for feature maps extracted from the voltage data and M_m means feature maps extracted from the mask image. D denotes the number of channels of the feature map and ceil($\cdot$) represents round up function. Other network components are described in Fig. 2.

obvious that SEIM keeps the properties of the EIM, although it is sparser.

The MCN corrects the input inaccurate mask image by combining voltage information from VFEN. Inspired by the FC-UNet which is proposed to generate the mask image for EIT reconstruction [16], MCN employs an UNet structure as shown in Fig. 4. Suppose the input inaccurate mask image and the true accurate mask image are denoted by $I^{\mathrm{im}} \in \mathbb{R}^{64 \times 64}$ and $I^{\mathrm{am}} \in \mathbb{R}^{64 \times 64}$, the MCN directly predicts the difference mask image $I^{\mathrm{dm}} \triangleq I^{\mathrm{am}} - I^{\mathrm{im}}$ since the residual is easier to learn [30]. Afterwards, the corrected mask image represented by $I^{\mathrm{cm}} \triangleq \tilde{I}^{\mathrm{dm}} + I^{\mathrm{im}}$ and voltage features extracted by VFEN will be fed into RN to perform final EIT image reconstruction, where $\tilde{I}^{\mathrm{dm}} = \tilde{I}^{\mathrm{dm}}\left(I^{\mathrm{s}}, I^{\mathrm{im}}\right) \in \mathbb{R}^{64 \times 64}$ accounts for the predicted difference mask image by VFEN and MCN.

For RN, the number of residual blocks and the number of residual units in a single residual block also diminish like VFEN to reduce the network complexity (see Fig. 5). The Dual-Modal Feature Fusion Modules (DMFF, which includes DMFF-V1 and DMFF-V2; see Fig. 6) and Multi-Scale Feature Fusion Modules (MSFF; see Fig. 7) have similar architectures as those in MSFCF-Net [28]. However, their number of resid-

ual units is also reduced. The output of RN is the predicted EIT image (denoted by $\tilde{I}^{\mathrm{eit}} \in \mathbb{R}^{64 \times 64}$), which is the two-dimensional version of $\Delta\sigma$. Unless specially specified, all convolutional layers in the En-MSFCF-Net would not change the height and width of their input feature maps and all transposed convolutional layers always enlarge twice the height and width of their input feature maps (kernel size: 2×2; stride: 2). It is worth mentioning that the architecture of the En-MSFCF-Net can be further optimized by decreasing its complexity while maintaining its performance, or by introducing more advanced architecture or attention mechanism to improve its performance. However, this is beyond the scope of this research since the main objective is to demonstrate the effectiveness of the two-stage network framework, which corrects the mask image followed by the conductivity estimation.

In practical implementation, I^{im}, I^{s}, I^{dm}, $\tilde{I}^{\mathrm{dm}}$, $\tilde{I}^{\mathrm{eit}}$ and the true reconstructed EIT image I^{eit} are denoted by tensors of size $C \times 64 \times 64$, where $C = 1$ is the channel number. For mask images and EIT images, pixels outside the circular

sensing region are filled by zero and pixels inside the sensing region are visually displayed to emphasize the shape of the sensing region throughout this paper. In this study, En-MSFCF-Net is trained with two stages. The VFEN and MCN are preliminarily trained at the first stage and then the whole network is trained at the second stage. Given a training set $\Psi = \left\{ I_k^s, I_k^{im}, I_k^{dm}, I_k^{eit} \right\}_{k=1}^{U}$, U is the number of samples, the loss function for the first stage and the second stage can be formulated as (6) and (7), respectively:

$$L_{s_1} = \frac{1}{U} \sum_{(I^s, I^{im}, I^{dm}) \in \Psi} \left\| I^{dm} - \tilde{I}^{dm} \left(I^s, I^{im} \right) \right\|^2$$
$$+ \frac{\lambda_1}{2} \left\| (\theta_1, \theta_2) \right\|^2, \tag{6}$$

$$L_{s_2} = \frac{1}{U} \sum_{(I^s, I^{im}, I^{dm}, I^{eit}) \in \Psi} \left[\left\| I^{dm} - \tilde{I}^{dm} \left(I^s, I^{im} \right) \right\|^2 \right.$$
$$\left. + \left\| I^{eit} - \tilde{I}^{eit} \left(I^s, I^{im} \right) \right\|^2 \right] + \frac{\lambda_2}{2} \left\| (\theta_1, \theta_2, \theta_3) \right\|^2, \tag{7}$$

where θ_1, θ_2, and θ_3 represent the network parameters of VFEN, MCN and RN respectively. λ_1 and λ_2 denote the l_2 regularization parameters for the first training stage and second training stage. The learned θ_1 and θ_2 at the first training stage is utilized to initialize VFEN and MCN at the second training stage.

III. DATA GENERATION, AUGMENTATION AND NETWORK TRAINING

A. Dataset Building

The dataset adopted in this study is constructed by using COSMOL Multiphysics. A 2D 16-electrode sensor is first modelled (see Fig. 8), then different types of inclusions are added to it, forming a variety of conductivity distributions. Four types of inclusions are considered, including one-circle object with multi-level conductivity (from 0.002 S/m to 1.998 S/m; see examples in row one and row two, column one in Fig. 9), two-circle objects with fixed conductivity (one is 4 S/m and the other is 0.002 S/m; see an example in row three, column one in Fig. 9), embedded-circle object with fixed conductivity (the embedded circlular region is 1 S/m and the outer ring is 0.1 S/m; see an example in row four, column one in Fig. 9), and layered-square object with fixed conductivity (one layer is 1 S/m and the other layer is 0.1 S/m; see an example in row five, column one in Fig. 9). For all cases, the background conductivity is set as 2 S/m and objects are randomly distributed in the circular sensing area of the EIT. Accurate mask images are generated by setting one in the inclusion region and setting zero in the background. Following the settings, we generated 19,000 samples, and each sample includes a frame of voltage measurements (arranged as SEIM), an accurate mask image, and a true conductivity image. There are 5,000 one-circle samples, 4,000 two-circle samples, 5,000 embedded-circle samples and 5,000 embedded square samples. For each type of data, randomly selected 10% samples are used as the test set, randomly selected 10% samples from the rest
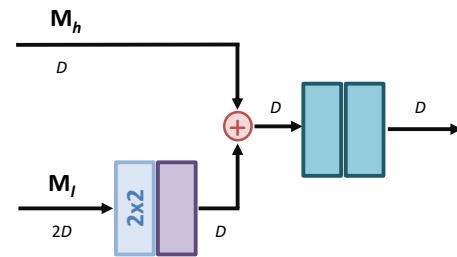

Fig. 7. Architecture of the MSFF. D denotes the number of channels of the feature map. M_l means the small-scale feature map and M_h means the large-scale feature map. Other network components are described in Fig. 2. and Fig. 4.

data are adopted as the validation set, and the remaining data serves as the training set. Eventually, there are 1,900 samples for testing, 1,710 samples for validation, and 15,390 samples for training.

B. Data Augmentation

Data augmentation is adopted during the En-MSFCF-Net training and testing. Since deep learning models are robust to voltage noise to some extent as shown in [28], there are enough training samples, and the diversity of samples is properly controlled, we take a static way to augment voltage data. In the training process, we separately add additive Gaussian noise with Signal-to-Noise Ratio (SNR) of 50 dB and 40 dB to a quarter of training and validation sets for each type of data. In [28], the MSFCF-Net is proved more sensitive to the inaccurate mask image compared with noisy voltage data. Therefore, for the mask image augmentation, we adopt a dynamic way in training to introduce more diverse inaccurate mask images. For each mask image in the training and validation set, we perturbed the boundary of the accurate mask image (this will result in the inaccurate mask image) with a probability of 50% before feeding it into the network. The concrete perturbation method is stated as follows. First, we randomly select some points on the boundary of the inclusions in the accurate mask image. According to the number of inclusions, the number of randomly selected points are set as 6 for one-circle samples, embedded-circle samples, and layered-square samples, and set as 10 for two-circle samples. Then, for each selected point, we either set its eight-neighboring pixels as one or its four-neighboring pixels as zero with a probability of 50%.

In the testing process, additive Gaussian noise with Signal-to-Noise Ratio (SNR) of 50 dB, 40 dB and 30 dB is separately added into the voltage data of the whole noise-free test set, forming four test sets. For each one of the four test sets, half of mask images are randomly selected to be perturbed using the method stated in the previous paragraph.

C. Network Training

We use Pytorch to implement the En-MSFCF-Net and adopt AdamW [33] to realize optimization. The training and testing of the En-MSFCF-Net is conducted on a server equipped with GeForce RTX 2080 GPU. For the first training stage, the hyper parameters are set as follows: the learning rate is set as 0.001
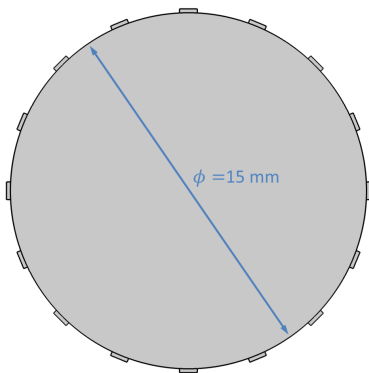
Fig. 8. Simulated 2D 16-electrode impedance sensor.

and the penalty parameter λ_1 for l_2 regularization is set as 0.0001. The maximum number of training epochs is set as 150 and the batch size is set as 128. For the second training stage, the learning rate is set as 0.0005 and the penalty parameter λ_2 for l_2 regularization is 0.00005. The maximum number of training epochs is set as 200 and the batch size is set as 96. To avoid overfitting, early-stopping is employed in each training stage. The tolerance is set as 10 for the first training stage and set as 15 for the second training stage. Ultimately, the first training stage stops at epoch 141 and the second training stage stops at epoch 163.

Comparative neural networks are also implemented by using the Pytorch and optimized with AdamW [33]. The same data augmentation method used in the training and testing of the En-MSFCF-Net is also adopted. VD-Net only involves voltage data augmentation while the MSFCF-Net involves both voltage data augmentation and mask image perturbation. For VD-Net, the learning rate is set as 0.001 and the penalty parameter for l_2 regularization is selected as 0.0001. For MSFCF-Net, the learning rate is set as 0.0005 and the penalty parameter for l_2 regularization is set as 0.00005. The tolerance of the early stopping is set as 15 for both networks. Eventually, The VD-Net experiences 134 epochs and the MSFCF-Net experiences 136 epochs.

IV. RESULTS

In this section, the proposed En-MSFCF-Net is evaluated by simulation and real experimental data. Its performance is compared with TV regularization algorithm [32], Cross Gradient based method [25], VD-Net [14] and MSFCF-Net [28].

A. Simulation Results

1) Quantitative Metrics: As the true conductivity distribution is known in simulation study, various metrics is chosen to quantitatively evaluate the algorithm performance. Relative Image Error (RIE, a positive number) and Mean Structural Similarity Index (MSSIM, ranging from $0 \sim 1$) [34] are selected as the indicators to evaluate the quality of individual reconstructed EIT image. RIE measures how accurate the conductivity values of a predicted EIT image are compared with those of the ground truth EIT image. The lower the

RIE, the more accurate the estimated conductivity values will be. MSSIM measures the degree of the structural similarity between the predicted EIT image and the ground truth EIT image. The higher the MSSIM, the more similar the predicted structure will be to the true structure. The definitions of RIE and MSSIM are expressed as:

$$\text{RIE} = \frac{\|\sigma_P - \sigma_G\|}{\|\sigma_G\|} \tag{8}$$

$$\text{MSSIM} = \frac{1}{HW} \sum_h \sum_w \frac{(2\boldsymbol{\mu}_P \boldsymbol{\mu}_G + C_1)(2\boldsymbol{\delta}_{PG} + C_2)}{(\boldsymbol{\mu}_P^2 + \boldsymbol{\mu}_G^2 + C_1)(\boldsymbol{\delta}_P^2 + \boldsymbol{\delta}_G^2 + C_2)} \tag{9}$$

where subscripts P and G represents the marked quantities are related to the predicted EIT image and the ground truth EIT image, respectively. σ_P and σ_G denotes the vectored version of the predicted EIT image and the ground truth EIT image. h and w are position indexes of an image, and H and W are height and weight of an image. $\mu_P = \mu_P(h,w), \mu_G = \mu_G(h,w), \delta_P = \delta_P(h,w), \delta_G = \delta_G(h,w)$, and $\delta_{PG} = \delta_{PG}(h,w)$ are the local means, standard deviations and the cross-covariance for image P and G. $C_1 = (K_1 L)^2$, $K_1 = 0.01$; $C_2 = (K_2 L)^2$, $K_2 = 0.03$. L is set as 1. In addition, to evaluate algorithm's performance on a dataset, Mean RIE (M-RIE) and Mean MSSIM (M-SSIM) are adopted, which are defined as the average RIE and MSSIM over all images for a specified dataset.

2) Results and Discussion: First, we compare algorithms on individual samples by visual comparison. The reconstructed images and corresponding quantitative metrics are shown in Fig. 9. The samples are selected from the test set with voltage data of SNR = 50 dB. Results of each algorithm occupy two columns. Images in the left column are reconstructed EIT images, and images in the right column are error images which are the absolute difference between the predicted EIT images and the ground truth EIT images. Fig. 10 shows the input mask images for MSFCF-Net and En-MSFCF-Net (first row, i.e. I^{im}), the difference mask images predicted by the En-MSFCF-Net (second row, i.e. $\tilde{I}^{\text{dm}}$), and the corrected mask images generated by En-MSFCF-Net (third row, i.e. I^{cm}). The results in Fig. 9 and Fig. 10 obviously indicate the En-MSFCF-Net can effectively correct the inaccurate mask images, which leads to a more precise conductivity predication and a more robust structural preservation. Especially, when the mask image is accurate (see Phantom 1), the En-MSFCF-Net applies little changes to the input mask image. The best performance of the En-MSFCF-Net can also be reflected by RIE and MSSIM. Contrarily, the results based on TV are the worst because they cannot accurately predict the conductivity values, meanwhile losing most structural information. The Cross Gradient-based algorithm introduces some structural information. However, it only incorporates boundary information while cannot make any improvements on the accuracy of the conductivity prediction and background artefact suppression. Although VD-Net makes an improvement compared with TV, it reconstructs a wrong image for Phantom 5. However, both MSFCF-Net and EN-MSFCF-Net do not make such mistake because the prior information provided by the mask image can
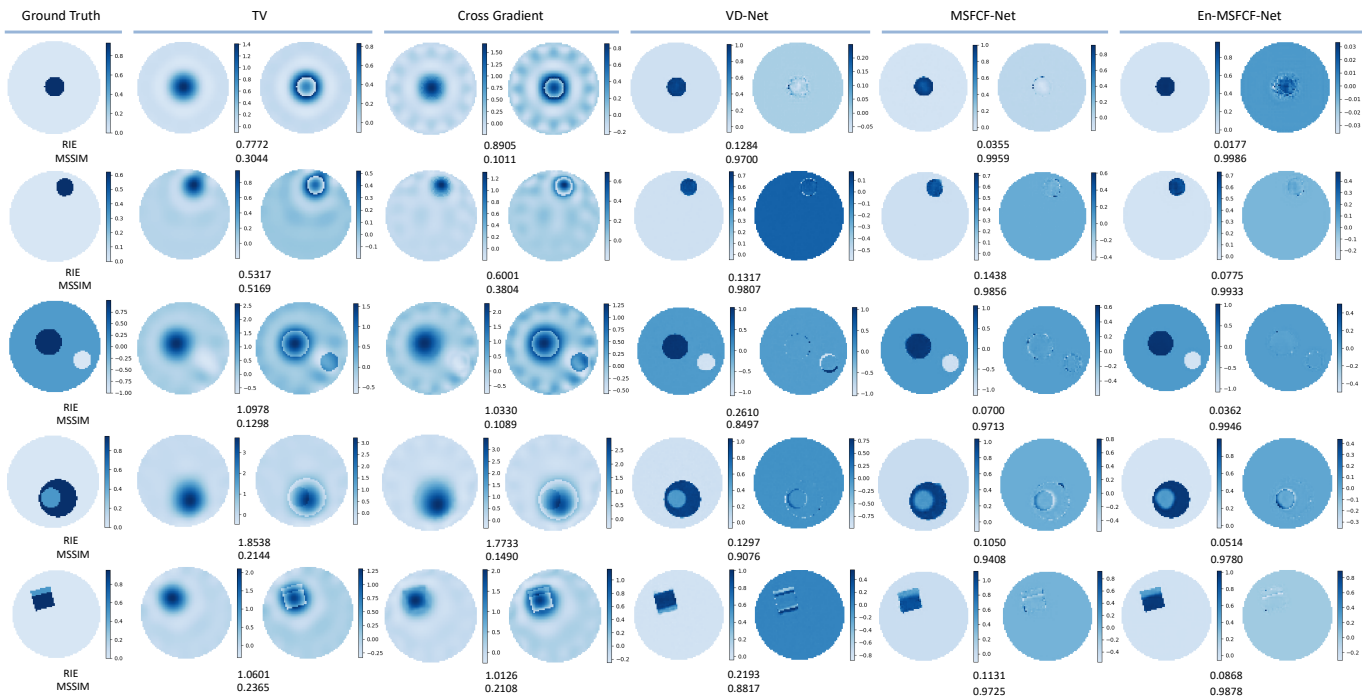
Fig. 9. Visual comparison of different algorithms on five representative phantoms (for each algorithm, left column: reconstruction; right column: error image). Phantoms are labeled by Phantom 1 - Phantom 5 from top to bottom.
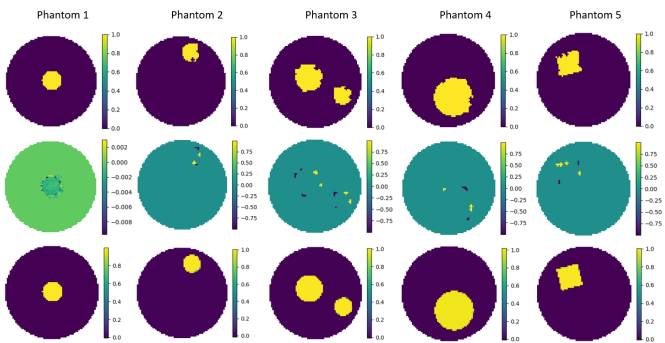

Fig. 10. Input mask images (first row), predicted difference mask images (second row) and corrected mask images (third row) corresponding to phantoms in Fig. 9.

help constrain the EIT solution space, thus generating a more reliable solution. Since the En-MSFCF-Net can acquire a more accurate mask image for reconstruction through mask image correction, its results are better than those of MSFCF-Net.

Fig. 11 and Fig. 12 further demonstrate the superiority of the proposed En-MSFCF-Net. Fig. 11 displays the enlarged inclusion at the bottom right of the Phantom 3. It is shown that TV is hard to locate this inclusion and the shape of the inclusion predicted by VD-Net is inaccurate. The boundary of the inclusion can be roughly identified by the Cross Gradient-based method. It is more evident this algorithm cannot improve the accuracy of the conductivity prediction and background artefact suppression, which is consistent with our previous conclusion. In comparison with the result of MSFCF-Net, En-MSFCF-Net reconstructs such inclusion with more homogeneous conductivity and a clearer boundary. We

select two line segments (indicated by red line segments) in Phantom 1 and show the 1D profiles of them. The results are illustrated in Fig. 12. For the bottom-line segment profiles, the results of TV and Cross Gradient are not shown because the background artefacts of the images generated by TV and Cross Gradient are more severe than those generated by learning-based algorithms. The results of the En-MSFCF-Net are closest to the ground truths. In summary, by introducing the mask image, the MSFCF-Net and En-MSFCF-Net can better suppress the background artefacts while increasing the accuracy of the conductivity prediction. Moreover, as the En-MSFCF-Net can correct the inaccurate mask image, it can gain the most precise conductivity predictions and the most accurate inclusion structure among the given algorithms.

Table I quantitatively compares the learning models' performance on the four test sets and Table II shows the quantitative metrics of the three learning-based models on different types of data with the SNR of 50 dB. From Table I, the MSFCF-Net and En-MSFCF-Net illustrate superior robustness on voltage noise because of the mask image. In addition, the M-RIE of En-MSFCF-Net is lower and the M-MSSIM is higher than those of the MSFCF-Net, which is due to the mask correction capability of En-MSFCF-Net. Table I displays the results from a holistic perspective while Table II offers a more detailed result. It indicates that En-MSFCF-Net outperforms all models on each type of data. These results verify the conclusion drawn from the previous analysis.

To demonstrate the effectiveness of the En-MSFCF-Net processing samples with inaccurate mask images, we trained the En-MSFCF-Net on the dataset with accurate mask images. The method of the voltage data augmentation remains unchanged.
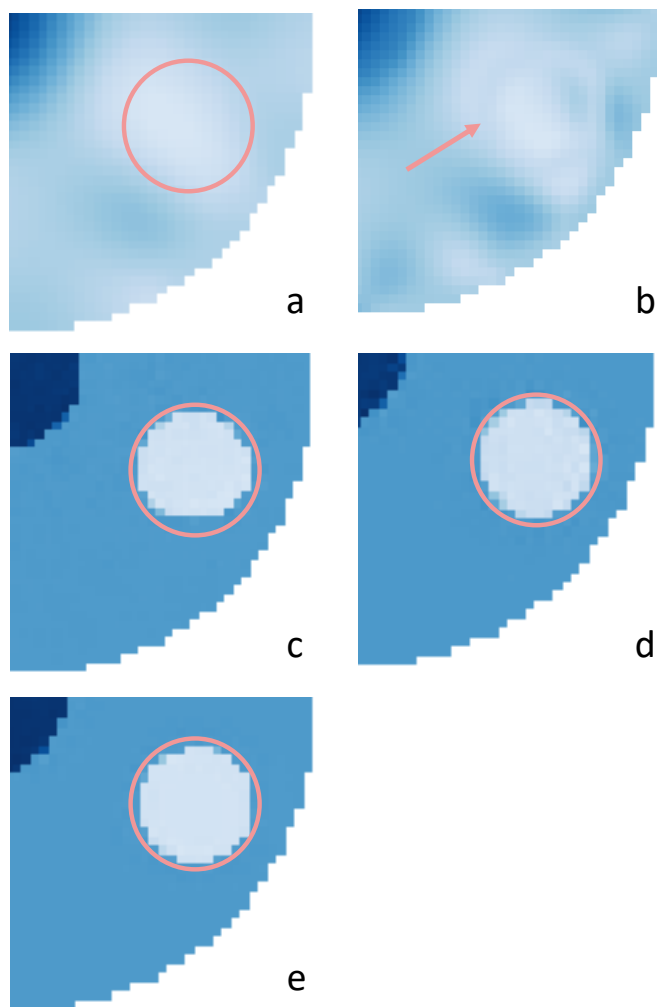
Fig. 11. Visual comparison of algorithms on an enlarged inclusion. Pink circle defines the true boundary of the inclusion and pink arrow indicate the inclusion reconstructed by Cross Gradient regularization. (a) the result of TV; (b) the result of Cross Gradient; (c) the result of VD-Net; (d) the result of MSFCF-Net; (e) the result of En-MSFCF-Net.
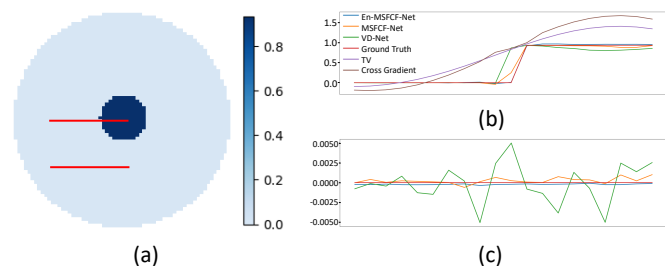

Fig. 12. Visual comparison of algorithms on red lines in (a). (a) is the true conductivity image of the Phantom 1 in Fig. 9. (b) is the results on the top line and (c) is the results on the bottom line.

Then the previously trained En-MSFCF-Net and this new model are tested on two test sets. One contains mask images that are all accurate and the other contains mask images that are all inaccurate. The SNR of all voltage data of the two test sets is set as 50 dB. Thus, there are four types of combinations concerning the trained model types and test set types. The results are shown in Table III. Obviously, if the En-MSFCF-

Net is trained with accurate mask images and tested on the dataset with accurate mask images, the results are the best. If we use dataset with inaccurate mask images to attack the model trained with the accurate mask images, the performance decreases, which is consistent with the conclusions in [28]. However, if we train the En-MSFCF-Net with inaccurate mask images, the advert influence of the inaccurate masks alleviates, which can be indicated by the descent of the M-RIE and ascent of the M-MSSIM. Combined with the results in Table I and II, we can conclude that the En-MSFCF-Net trained with inaccurate mask images can effectively address the situation suffering inaccurate mask images and its performance is better than the MSFCF-Net trained with inaccurate mask images.

Finally, it is worth mentioning another two challenging circumstances the dual-modal imaging may face. The first is that the optical sensor cannot recognize some imaging targets, which leading to such targets not appearing in both mask images and reconstructed EIT images. The second is that targets in the mask image are noticeably smaller or larger than the real ones. We do not deal with these situations by the En-MSFCF-Net as the first case already violates the concept of dual-modal imaging. The latter case will introduce considerable incorrect information, which causes the mask image not to provide enough useful information.

B. Experiment Results

1) Dual-modal Imaging System : The same dual-modal sensor in [28] is used to collect the real-world experimental data. For the impedance sensor, the adjacent-injection adjacent-measurement protocol [35] is adopted and the frequency of the injected current is set as 10 kHz. As discussed previously, the dual-modal sensor simultaneously provides a frame of voltage measurements and a guidance image for each imaging.

2) Experimental Phantoms: The experiments are carried out at the room temperature ($\sim$ 20 °C) and we made use of three distinct phantoms. The background medium was chosen as Phosphate Buffered Saline (PBS) whose conductivity is measured to be 1.47 S/m. The first phantom is a 3D printed cylinder which was fabricated via stereolithography (SLA), using black photopolymer resin (FormLabs Inc., MA) (see the first row in Fig. 13). The height and the diameter of the cylinder are 1 mm and 2 mm, respectively. As the black resin is almost non-conductive, the relative conductivity change of it is around 1. The second phantom is a fresh cylindrical mango tissue whose height and the diameter are 1 mm and 3 mm (see the second row in Fig. 13). [36] studies the conductivity change of the mango juice with the increasing of the temperature. It is shown that the conductivity of the mango juice at the room temperature is around 0.12 S/m. Therefore, we can reasonably infer that the solid mango tissue at the room temperature is lower than 0.12 S/m and the relative conductivity change is around 1. The last phantom is a cylindrical gel whose height and diameter are 1 mm and 5 mm, respectively (see the third row in Fig. 13). The gel was carefully made in lab to adjust its conductivity. We diluted the Phosphate Buffered Saline (PBS) by adding distilled deionized water. The conductivity of the resulting solution was measured

TABLE I

QUANTITATIVE COMPARISON OF DIFFERENT DEEP LEARNING MODELS ON THE WHOLE TEST SET

Models	Noise-free		50 dB		40 dB		30 dB	
	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM
VD-Net	0.1537	0.9506	0.1543	0.9503	0.1581	0.9488	0.1824	0.9399
MSFCF-Net	0.1317	0.9792	0.1317	0.9792	0.1320	0.9791	0.1346	0.9783
En-MSFCF-Net	**0.1250**	**0.9841**	**0.1251**	**0.9840**	**0.1256**	**0.9839**	**0.1307**	**0.9825**

Black bold numbers represent the best results.

TABLE II

QUANTITATIVE COMPARISON OF DIFFERENT DEEP LEARNING MODELS ON DIFFERENT TYPES OF SAMPLES

Models	One Circle		Two Circles		Embedded Circle		Layered Square	
	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM
VD-Net	0.1942	0.9822	0.0847	0.9605	0.1318	0.9205	0.1926	0.9403
MSFCF-Net	0.1733	0.9867	0.0972	0.9803	0.1291	0.9681	0.1510	0.9821
En-MSFCF-Net	**0.1583**	**0.9891**	**0.0641**	**0.9918**	**0.1252**	**0.9693**	**0.1415**	**0.9877**

Black bold numbers represent the best results.

TABLE III

COMPARISON OF EN-MSFCF-NETS TRAINED WITH AND WITHOUT PERTURBED MASK IMAGES ON TEST SETS WITH ACCURATE AND INACCURATE MASK IMAGES

Type	Whole Test Set		One Circle		Two Circles		Embedded Circle		Layered Square	
	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM
MNP w/ TNP	**0.0827**	**0.9903**	**0.1036**	**0.9952**	**0.0171**	**0.9979**	**0.1028**	**0.9754**	**0.0948**	**0.9944**
MNP w/ TP	0.2704	0.9495	0.3462	0.9666	0.2426	0.9384	0.2075	0.9293	0.2818	0.9620
MP w/ TNP	0.1015	0.9870	0.1287	0.9915	0.0277	0.9964	0.1210	0.9712	0.1201	0.9918
MP w/ TP	**0.1495**	**0.9811**	**0.1968**	**0.9862**	**0.1044**	**0.9877**	**0.1282**	**0.9677**	**0.1608**	**0.9844**

MNP means the En-MSFCF-Net is trained with training and validation sets with non-perturbed (accurate) mask images. MP represents the En-MSFCF-Net is trained with training and validation sets with perturbed (inaccurate) mask images. This training method is the same as that described in Subsection B and C, Section III. TP stands for all the mask images in the test set are perturbed. TNP accounts for all the mask images in the test set are non-perturbed. Black bold numbers represent the best results and blue bold numbers emphasize the results of En-MSFCF-Net trained with training and validation sets with perturbed mask images and tested on dataset with perturbed mask images.

to be 0.92 S/m. Then gelatin powder was added to the liquid to form a gel. Thus, The relative conductivity change of the gel is around 0.3741.

3) Results and Discussion: The image reconstruction results are shown in Fig. 13. The input mask images for MSFCF-Net and En-MSFCF-Net are illustrated in column six and they are generated by the guidance image processing algorithm. The guidance images are almost like images in the column one of the Fig. 13, but they only occupy the EIT sensing region. For the mask image generation, we first convert the guidance image into its gray-scale version, then apply the threshold method to segment the inclusion. The resulting binary image is further refined by morphological operation. Finally, the mask image is acquired by down-sampling the refined binary image to the EIT image size. The images in column seven and eight are predicted difference mask images and corrected mask images by En-MSFCF-Net. As the input mask images are approximately accurate, the pixel values of the predicted mask images are very small. For the gel result of each algorithm, the average conductivity in a

light pink circular region is also indicated in Fig. 13. From this figure, it is clear that the VD-Net cannot recover the shape of the phantoms, which is the defect of the single-modal learning-based algorithm. In addition, for gel imaging, it also predicts wrong conductivity values. For TV, it predicts relatively correct conductivity values except the black resin phantom, however the background artefacts are very sever and its ability to preserve the phantom shape is low. For Cross Gradient, although it shows some structural features of the phantoms, the image quality is similar to that generated by TV. As the mask image is approximately accurate, the quality of the results of the MSFCF-Net and En-MSFCF-Net is similar and they outperform other algorithms. Additionally, as shown in simulation study, the phantom in the image generated by En-MSFCF-Net is more homogeneous than that in the image generated by MSFCF-Net. This indicates the En-MSFCF-Net can better recover the conductivity distribution because the phantoms in the experiment are homogeneous.

To further verify the mask correction ability of the En-MSFCF-Net, the inaccurate mask images of the experimental
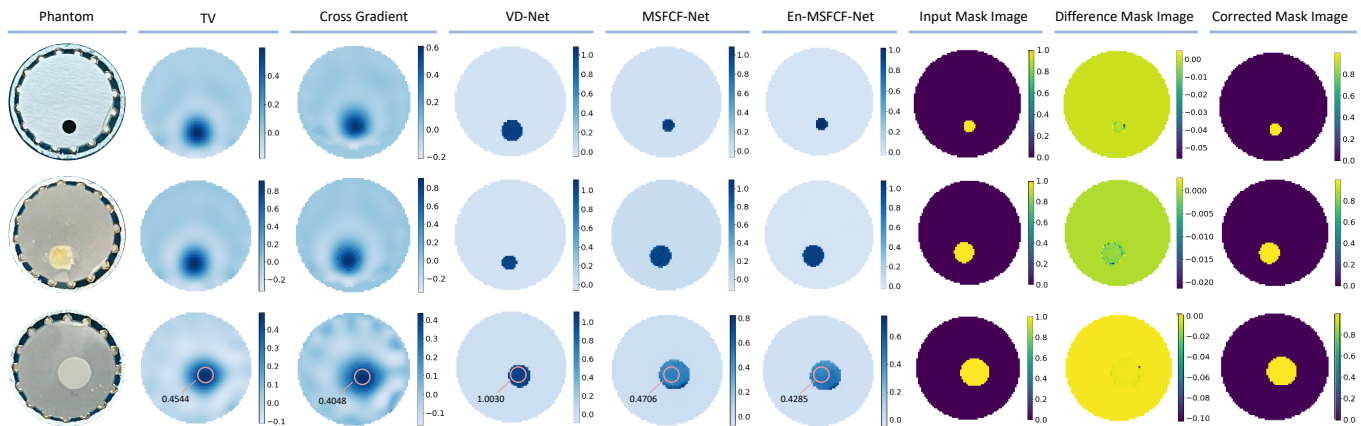
Fig. 13.   Visual comparison of different algorithms on experimental data. The first row is the results of the black resin imaging, the second row is the results of the mango imaging, and the last row is the results of the gel imaging.
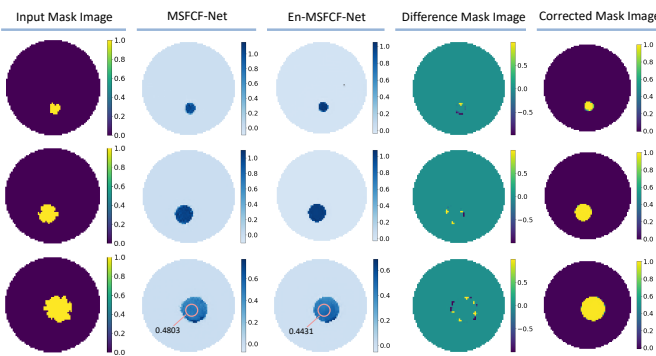

Fig. 14.   Image reconstruction results of learning-based dual-modal algorithms based on experimental data with inaccurate mask images.

phantoms are generated from the mask images in the seventh column of the Fig. 13 by simulation. The generated mask images are illustrated in the first column of Fig. 14 and they serve as the input of the learning-based dual-modal algorithms. The image reconstruction results of the En-MSFCF-Net and MSFCF-Net, and the mask correction results of the En-MSFCF-Net are also shown in Fig 14. For the gel imaging, though the inaccurate mask images affect the accuracy of the conductivity prediction, the decrease of the prediction accuracy is tolerable for both En-MSFCF-Net and MSFCF-Net. The average value in the light pink circular region varies within 0.02 for these two algorithms. Nevertheless, from this figure, we can draw the same conclusion as in the simulation study that: En-MSFCF-Net can better address the inaccurate mask image issue than MSFCF-Net.

V. CONCLUSION AND FUTURE WORK

We report a two-stage convolutional neural network named En-MSFCF-Net for impedance-optical dual-modal imaging. This network corrects the inaccurate mask image followed by the conductivity estimation. The results of simulation and experiments demonstrate that:

- Benefit from the mask correction, the En-MSFCF-Net can effectively process input data with inaccurate mask image

compared with other given algorithms. Visual results directly show that En-MSFCF-Net can preserve inclusions' boundary and internal conductivity homogeneity.

- Quantitative metrics also indicate the superiority of the En-MSFCF-Net. The RIE of En-MSFCF-Net is commonly 0.5 lower than that of model-based algorithms and the MSSIM of En-MSFCF-Net is usually 0.6 higher than that of model-based algorithms. Compared with learning-based methods, En-MSFCF-Net still achieves the lowest RIE and M-RIE, and highest MSSIM and M-MSSIM on both individual sample and dataset level evaluation.

Future work will extend the dual-modal imaging framework and the En-MSFCF-Net to 3D tissue imaging.

REFERENCES

[1] P. Metherall, D. C. Barber, R. H. Smallwood et al., "Three-dimensional electrical impedance tomography," *Nature*, vol. 380, no. 6574, pp. 509-512, 1996.

[2] R. H. Bayford, "Bioimpedance tomography (electrical impedance tomography)," *Annu. Rev. Biomed. Eng.*, vol. 8, pp. 63-91, 2006.

[3] M. Cheney, D. Isaacson, and J. C. Newell, "Electrical impedance tomography," *SIAM review*, vol. 41, no. 1, pp. 85-101, 1999.

[4] A. Adler, and A. Boyle, "Electrical impedance tomography: Tissue properties to image measures," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 2494-2504, 2017.

[5] J. Yao, and M. Takei, "Application of process tomography to multiphase flow measurement in industrial and biomedical fields: A review," *IEEE Sensors Journal*, vol. 17, no. 24, pp. 8196-8205, 2017.

[6] H. Wu, Y. Yang, P. O. Bagnaninchi et al., "Electrical impedance tomography for real-time and label-free cellular viability assays of 3D tumour spheroids," *Analyst*, vol. 143, no. 17, pp. 4189-4198, 2018.

[7] J. Li, S. Yue, M. Ding et al., "Adaptive L_p regularization for electrical impedance tomography," *IEEE Sensors Journal*, vol. 19, no. 24, pp. 12297-12305, 2019.

[8] Q. Wang, H. Wang, R. Zhang et al., "Image reconstruction based on L1 regularization and projection methods for electrical impedance tomography," *Review of scientific instruments*, vol. 83, no. 10, pp. 104707, 2012.

[9] B. Gong, B. Schullcke, S. Krueger-Ziolek et al., "Higher order total variation regularization for EIT reconstruction," *Medical & biological engineering & computing*, vol. 56, no. 8, pp. 1367-1378, 2018.

[10] Y. M. Jung, and S. Yun, "Impedance imaging with first-order TV regularization," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 193-202, 2014.

[11] A. Borsic, B. M. Graham, A. Adler et al., "In vivo impedance imaging with total variation regularization," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 44-54, 2009.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2022.3200758

Z. LIU *et al.*: PREPARATION OF PAPERS FOR IEEE TRANSACTIONS AND JOURNALS (FEBRUARY 2017)   11

[12] Y. Yang, and J. Jia, "An image reconstruction algorithm for electrical impedance tomography using adaptive group sparsity constraint," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2295-2305, 2017.

[13] Y. Yang, H. Wu, and J. Jia, "Image reconstruction for electrical impedance tomography using enhanced adaptive group sparsity with total variation," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5589-5598, 2017.

[14] F. Li, C. Tan, and F. Dong, "Electrical resistance tomography image reconstruction with densely connected convolutional neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2020.

[15] D. Hu, K. Lu and Y. Yang, "Image reconstruction for electrical impedance tomography based on spatial invariant feature maps and convolutional neural network," *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2019, pp. 1-6, doi: 10.1109/IST48021.2019.9010151.

[16] Z. Chen, Y. Yang, J. Jia et al., "Deep learning based cell imaging with electrical impedance tomography," *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2020, pp. 1-6.

[17] Z. Wei, D. Liu, and X. Chen, "Dominant-current deep learning scheme for electrical impedance tomography," IEEE Transactions on Biomedical Engineering, vol. 66, no. 9, pp. 2546-2555, 2019.

[18] Z. Wei, and X. Chen, "Induced-current learning method for nonlinear reconstructions in electrical impedance tomography," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1326-1334, 2019.

[19] D. Smyl, T. N. Tallman, D. Liu et al., "An efficient quasi-newton method for nonlinear inverse problems via learned singular values," *IEEE Signal Processing Letters*, vol. 28, pp. 748-752, 2021.

[20] M. Capps, and J. L. Mueller, "Reconstruction of organ boundaries With deep learning in the D-Bar method for electrical impedance tomography," IEEE Transactions on Biomedical Engineering, vol. 68, no. 3, pp. 826-833, 2020.

[21] S. J. Hamilton, and A. Hauptmann, "Deep D-bar: Real-time electrical impedance tomography imaging with deep neural networks," *IEEE transactions on medical imaging*, vol. 37, no. 10, pp. 2367-2377, 2018.

[22] B. Gong, B. Schullcke, S. Krueger-Ziolek et al., "Sparse regularization for EIT reconstruction incorporating structural information derived from medical imaging," *Physiological Measurement*, vol. 37, no. 6, pp. 843, 2016.

[23] Z. Liu, and Y. Yang, "Image Reconstruction of Electrical Impedance Tomography Based on Optical Image-Guided Group Sparsity," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21893-21902, 2021.

[24] Z. Liu, and Y. Yang, "Multi-modal Image Reconstruction of Electrical Impedance Tomography Using Kernel Method," *IEEE Transactions on Instrumentation and Measurement*, 2021.

[25] Z. Li, J. Zhang, D. Liu et al., "CT image-guided electrical impedance tomography for medical imaging," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1822-1832, 2019.

[26] G. Liang, S. Ren, S. Zhao, and F. Dong, "A Lagrange-Newton method for EIT/UT dual-modality image reconstruction," *Sensors*, vol. 19, no. 9, p. 1966, 2019.

[27] G. Liang, S. Ren, and F. Dong, "An adaptive local weighted image reconstruction algorithm for EIT/UTT dual-modality imaging," *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2017, pp. 1-6: IEEE.

[28] Z. Liu, P. Bagnaninchi and Y. Yang, "Impedance-optical Dual-modal Cell Culture Imaging with Learning-based Information Fusion," *IEEE Transactions on Medical Imaging*, doi: 10.1109/TMI.2021.3129739.

[29] D. B. Geselowitz, "An application of electrocardiographic lead theory to impedance plethysmography," *IEEE Transactions on biomedical Engineering*, no. 1, pp. 38-41, 1971.

[30] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[31] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "Cbam: Convolutional block attention module," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3-19.

[32] Osher et al., "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–464, 2005.

[33] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam", 2017, arXiv preprint arXiv:1711.05101.

[34] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004, doi: 10.1109/TIP.2003.819861.

[35] B. H. Brown and A. D. Seagar, "The Sheffield Data Collection System," *Clinical Physics and Physiological Measurement*, vol. 8, no. 4A, p. 91, 1987.

[36] S. Ganapathy, P. Raja, P. Vijayakumary et al., "Effect of physical properties of mango juice on the effectiveness of pulsed electric field treatment," *Journal of Pharmacognosy and Phytochemistry*, vol. 8, no. 6, pp. 691-696, 2019.

Zhe Liu (Graduate Student Member, IEEE) received his B.Eng. degree in Biomedical Engineering from Huazhong University of Science and Technology, China, in 2018, and the M.Sc. degree in Biomedical Engineering from City University of Hong Kong, HKSAR, China, in 2019. He is currently pursuing the Ph.D. degree with the Intelligent Sensing, Analysis and Control Group, School of Engineering, The University of Edinburgh, UK. His current research interests include multi-modal biomedical imaging, image processing and image reconstruction.

Renjie Zhao received his B.Eng. degree in Communication Engineering from Southwest University, China, in 2016, and the M.Sc. degree in Signal Processing and Communications from The University of Edinburgh, UK, in 2021. During his M.Sc. thesis project, he worked on learning-based multi-modal imaging for tissue engineering.

Graham Anderson (Post-Doctoral Research Fellow, Centre for Regenerative Medicine) received his B.Sc. (hons) in Human Biology from Queen Margaret University in Edinburgh, the United Kingdom, in 2015 and his Ph.D. in Clinical Brain Sciences from the University of Edinburgh in 2021. His current research interests concern 3D in vitro retinal models, photobiomodulation and imaging assay development.

Pierre-Olivier Bagnaninchi received the M.Sc. degree in theoretical physics and the Ph.D. degree in biomedical engineering both from the University of Grenoble, Grenoble, France, in partnership with Thales, in 1997 and 2001, respectively. He then accepted a two-year post-doctoral position in Prof. Maryam Tabrizian Laboratory at McGill University, Montreal, QC, Canada, to develop impedance-based biosensors for tissue engineering applications. Following a Post-Doctoral Research Associate (PDRA) in optical coherence tomography at Keele University, Keele, U.K., in 2004, he was awarded a five-year RCUK fellowship at The University of Edinburgh, Edinburgh, U.K., in 2008, where he was then appointed as a Principal Investigator in 2013. He is currently a Principal Investigator with the MRC Centre for Regenerative Medicine, College of Medicine and Veterinary Medicine, The University of Edinburgh. His main research interest is the development of label-free quantitative monitoring technologies to monitor cell proliferation and differentiation in 2-D and 3-D cell cultures (organoids, gels, and scaffolds).
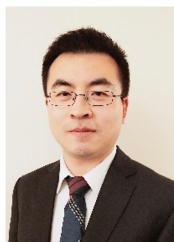
Yunjie Yang (Member, IEEE) received his B.Eng., M.Sc. and Ph.D. degrees from Anhui University, China, in 2010, Tsinghua University, China, in 2013, and The University of Edinburgh, UK, in 2018, respectively. After obtaining his Ph.D., he briefly worked as a Postdoctoral Research Associate in Chemical Species Tomography at The University of Edinburgh. Since September 2018, he has become the Lecturer and Chancellor's Fellow in Data Driven Innovation at School of Engineering, The University of Edinburgh. His research interests are in the areas of tomography, soft robotics, and machine learning for medical imaging. He is the Associate Editor of IEEE Access and the Guest Editor of IEEE Sensors Journal. He is also the recipient of the 2015 IEEE I&M Society Graduate Fellowship Award.