Edinburgh Research Explorer

Impedance-optical Dual-modal Cell Culture Imaging with Learning-based Information Fusion

Citation for published version:
Liu, Z, Bagnaninchi, P & Yang, Y 2021, 'Impedance-optical Dual-modal Cell Culture Imaging with Learning-based Information Fusion', *IEEE Transactions on Medical Imaging*, pp. 1-1.
https://doi.org/10.1109/TMI.2021.3129739

Digital Object Identifier (DOI):
10.1109/TMI.2021.3129739

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
IEEE Transactions on Medical Imaging

Impedance-optical Dual-modal Cell Culture Imaging with Learning-based Information Fusion

Zhe Liu, *Student Member, IEEE*, Pierre Bagnaninchi, Yunjie Yang, *Member, IEEE*

Abstract—**While Electrical Impedance Tomography (EIT) has found many biomedicine applications, better image quality is needed to provide quantitative analysis for tissue engineering and regenerative medicine. This paper reports an impedance-optical dual-modal imaging framework that primarily targets at high-quality 3D cell culture imaging and can be extended to other tissue engineering applications. The framework comprises three components, i.e., an impedance-optical dual-modal sensor, the guidance image processing algorithm, and a deep learning model named multi-scale feature cross fusion network (MSFCF-Net) for information fusion. The MSFCF-Net has two inputs, i.e., the EIT measurement and a binary mask image generated by the guidance image processing algorithm, whose input is an RGB microscopic image. The network then effectively fuses the information from the two different imaging modalities and generates the final conductivity image. We assess the performance of the proposed dual-modal framework by numerical simulation and MCF-7 cell imaging experiments. The results show that the proposed method could improve the image quality notably, indicating that impedance-optical joint imaging has the potential to reveal the structural and functional information of tissue-level targets simultaneously.**

Index Terms—**Cell culture, dual-modal imaging, electrical impedance tomography, deep learning, image processing**

I. INTRODUCTION

3D cell culture has far-reaching significance because it can better mimic the function of living tissues compared with cell monolayers, which has a significant impact on drug screening [1][2]. Providing better models of cell behaviors may benefit the research and treatment of human diseases and reduce animal testing. A key challenge in 3D cell culture is to determine the cellular state in depth and across time. Therefore, a suitable imaging technique is desired to monitor 3D cell culture continuously and non-destructively. Electrical Impedance Tomography (EIT) is a tomographic imaging technique that can recover the conductivity distribution within the interior of a domain through boundary current injection and induced voltage measurements [3]-[5]. It is proved that cell viability can be inferred by measuring its cellular electrical

parameters [6]. Recently, miniaturized EIT has been introduced to image the conductivity distribution of the 3D cultivated cells in both static and dynamic setups [7]-[10]. However, the low image quality of EIT has become a critical limiting factor for quantitatively analyzing the properties and behaviors of 3D cultivated cells in tissue engineering applications.

Thus far, efforts to improve EIT image quality have been primarily focused on advancing the image reconstruction algorithm. A prevailing type of EIT image reconstruction methods is based on regularization, which imposes certain prior knowledge in solving the problem. The state-of-the-art regularization methods include Total Variation (TV) regularization [11]-[13], Fidelity-Embedded Regularization [14], sparse regularization [15][16] and Adaptive Group Sparsity (AGS) regularization [17][18]. These elaborate model-based algorithms have proven effective in obtaining high image quality, but at the expense of computational efficiency or intricate parameter tuning. Deep learning [19] has recently demonstrated its potential in solving the nonlinear EIT inverse problem. A number of end-to-end deep-learning-based approaches have been reported for high-quality EIT image reconstruction [20]-[22], while the performance of such methods has a strong dependency on the quality of training datasets. An alternative is to combine model-based methods and deep learning to synthesize the strength of both approaches. Existing work includes using deep learning to assist model error compensation [23] or boost the performance of the model-based method [24]-[27]. Though these efforts lead to noteworthy advancements in EIT imaging, they mainly focus on solving the single-modal imaging problem.

Recently, dual-modal or multi-modal methods have also been explored to supplement EIT. For instance, the joint imaging of EIT and ultrasound tomography was investigated and showed improved image quality and structure preservation [28][29]. Gong *et al.* proposed a modified k-means based method to construct the regularization term to incorporate the prior information from the computed image into EIT image reconstruction [30]. Li *et al.* integrated structural information from X-ray tomography into EIT inversion by using the cross-gradient method [31]. Ren *et al.* also reported a CT-guided method in which the mask image of the chest region from CT scanning was adopted to improve the robustness of the deep-

learning-based EIT image reconstruction [32]. However, their method is not strictly a dual-modal imaging approach. Nevertheless, these predecessors have shown the evidence that combining information of other imaging modalities with EIT can improve the reconstructed EIT image quality.

Inspired by multi-modal imaging and deep learning, and to improve EIT image quality and promote EIT-based quantitative cellular assay in tissue engineering, we propose an impedance-optical dual-modal imaging framework to enable dual-modal cell imaging and learning-based information fusion. This work focuses on 2D imaging of the 3D cell culturing process, which reconstructs the cross section of 3D cell clusters. The imaging framework comprises the impedance-optical dual-modal miniature sensor for cell imaging, the guidance image processing algorithm for optical image preprocessing, and a dual-input deep learning model for information fusion and image reconstruction. The advantages of the proposed approach are:

1) Compared with single-modal methods, the proposed framework can generate EIT images with more accurate shapes by introducing optical imaging, thereby leading to more precise conductivity distribution estimation.

2) The framework develops a new indirect information fusion approach that addresses the challenge of directly using the optical image to train the deep learning model. This approach can be extended to other learning-based multi-modal image reconstruction scenarios with similar issues, e.g., where it is impossible to collect auxiliary images for training, or there are insufficient auxiliary images.

The remainder of the paper is organized as follows. Section II states the principle of EIT inverse problem. Section III describes the proposed framework. Section IV elaborates 2D simulation data generation and experiment setup. Section V illustrates simulation and experimental results. Finally, Section VI draws the conclusion and discusses future work.

II. INVERSE PROBLEM OF EIT

We describe the principle of EIT image reconstruction based on the 16-electrode configuration; as in this work, we adopt a 16-electrode miniature sensor and conduct 2D imaging. Suppose the sensing area occupies a 2D circular region $\Omega \subset \mathbb{R}^2$ (see Fig. 1). Sixteen electrodes denoted by $(e_1, e_2, \dots, e_{16})$ are attached around the boundary $\partial\Omega$ (see Fig. 1). Adjacent protocol [33] is adopted, where a specified current is applied successively to the electrode pairs $(e_\ell, e_{\ell+1}), \ell = 1, \dots, 16, e_{16+1} := e_1$; and the voltage difference between all other pairs of neighboring electrodes are collected. For each current injection, letting $\sigma = \sigma(x)$ and $u = u(x)$ denote the conductivity distribution and the electrical potential distribution in Ω respectively, the forward problem of EIT based on the Complete Electrode Model (CEM) [34] can be expressed as:

$$\nabla \cdot \big(\sigma(x)\nabla u(x)\big) = 0, \quad x \in \Omega \tag{1}$$
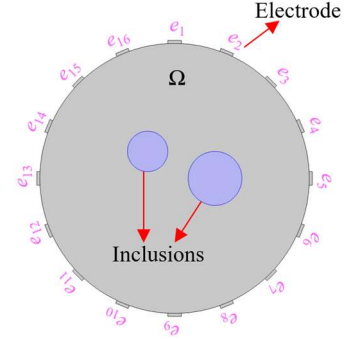

Fig. 1. Sixteen-electrode circular EIT sensor with two inclusions in it.

$$u(x) + z_\ell \sigma(x) \frac{\partial u(x)}{\partial n} = U_\ell, \quad x \in e_\ell, \ell = 1, 2, \dots, 16 \tag{2}$$

$$\int_{e_\ell} \sigma(x) \frac{\partial u(x)}{\partial n} dS = J_\ell, \quad \ell = 1, 2, \dots, 16 \tag{3}$$

$$\sigma \frac{\partial u(x)}{\partial n} = 0, \quad x \in \partial\Omega \backslash \bigcup_{l=1}^{16} e_\ell \tag{4}$$

where n is the outward unit normal of $\partial\Omega$. U_ℓ, J_ℓ and z_ℓ denotes the electrical potential, the injected current, and the contact impedance on the electrode e_ℓ, respectively.

The existence and uniqueness of the solution u should also be ensured by the charge conservation and the choice of the ground voltage defined respectively by the left and right equations below.

$$\sum_{\ell=1}^{16} J_\ell = 0, \quad \sum_{\ell=1}^{16} U_\ell = 0 \tag{5}$$

We define the measured voltage difference between electrode pairs $(e_g, e_{g+1}), g = 1, 2, \dots 16, e_{16+1} := e_1$, subject to the ℓth current injection as:

$$V^{\ell, g} := U_g^\ell - U_{g+1}^\ell \tag{6}$$

where U_g^ℓ and U_{g+1}^ℓ denote respectively the measured potential on the gth and $(g+1)^{th}$ electrode.

Time-difference EIT (td-EIT) reconstructs the conductivity variation in Ω through boundary voltage variation measurements. In this work, the electrodes directly contact the highly conductive cell culture media, and the contact impedance is negligible. Therefore, the boundary voltage variation on the gth electrode pair subject to the ℓth injection can be formulated as:

$$V_{\sigma_1}^{\ell, g} - V_{\sigma_0}^{\ell, g} = -\int_\Omega (\sigma_1 - \sigma_0) \nabla u_{\sigma_1}^\ell \nabla u_{\sigma_0}^g dx \tag{7}$$

where σ_1 denotes the conductivity distribution at the observation time point and σ_0 represents the conductivity distribution at the reference time point. $V_{\sigma_1}^{\ell, g}$ is the $V^{\ell, g}$ corresponding to σ_1, and so does $V_{\sigma_0}^{\ell, g}$. $u_{\sigma_1}^\ell$ is the electrical
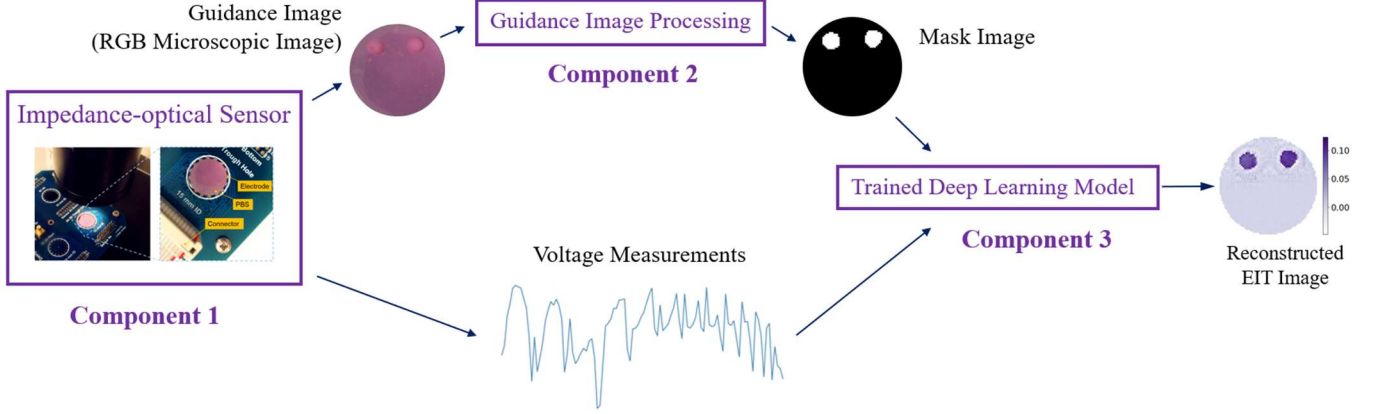
Fig. 2. Schematic of the impedance-optical dual-modal imaging framework.

potential distribution subject to the ℓth injection and σ_1. The definition of $u_{\sigma_0}^{\mathscr{g}}$ is similar with $u_{\sigma_1}^{\ell}$. Therefore, (7) can be reformulated as:

$$\Delta V^{\ell,\mathscr{g}} = H^{\ell,\mathscr{g}}(\Delta\sigma) \qquad (8)$$

where $\Delta\sigma$ is the conductivity variation in Ω and $H^{\ell,\mathscr{g}}$ is the non-linear mapping from $\Delta\sigma$ to $\Delta V^{\ell,\mathscr{g}}$. By eliminating repetitive data according to the reciprocity principle [35], we can obtain a frame of independent measurements, i.e. $\Delta V \in \mathbb{R}^{104}$. Therefore, the forward mapping can be ultimately expressed as $\Delta V = [H^{1,3}, \dots, H^{1,15}, H^{2,4}, \dots, H^{2,16}, \dots, H^{14,16}]^T \triangleq H(\Delta\sigma)$, H is the non-linear mapping from $\Delta\sigma$ to ΔV. The inverse problem can be formulated as:

$$\Delta\sigma = H^{-1}(\Delta V) \qquad (9)$$

where H^{-1} is the inverse mapping operator of H, which is to be approximated.

III. IMPEDANCE-OPTICAL DUAL-MODAL IMAGING FRAMEWORK

This section proposes an impedance-optical dual-modal imaging framework (see Fig. 2) to improve EIT image quality for 3D cell imaging. It consists of three components, i.e., the impedance-optical miniature sensor, the guidance image processing algorithm, and a deep learning model. First, the impedance-optical sensor will simultaneously output a frame of voltage measurements and an RGB microscopic image named the guidance image (I^g). Then, the guidance image processing algorithm will convert I^g into its corresponding mask image (I^m). Finally, I^m and the voltage measurements are fed into a deep learning model to generate the reconstructed EIT image.

A. Impedance-optical Dual-modal Sensor

The dual-modal sensor (see Fig. 3) combines a miniature 16-electrode EIT sensor with a digital microscope (Digital USB Microscope 1.3M, RS Components Ltd). The EIT sensor is manufactured on a Printed Circuit Board (PCB). A transparent
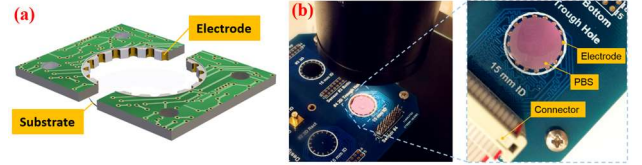

Fig. 3. Impedance-optical dual-modal sensor. (a) EIT sensor structure. (b) The manufactured dual-modal sensor.
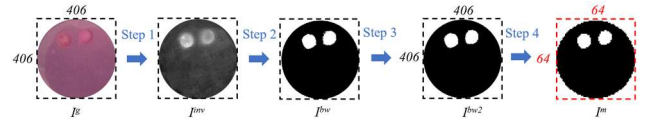

Fig. 4. An illustration of the guidance image processing procedure. The dashed square represents the circumscribed square region of the circular sensing region. The numbers mean the number of pixels for each side of the square.

glass substrate is attached at the bottom of the sensing area to support cells and enable optical imaging. The height and diameter of the sensing chamber are 1.6 mm and 14 mm, respectively. The 16 gilded microelectrodes are manufactured using the half-hole process and distributed at the periphery of the sensing area. The digital microscope is placed over the sensing chamber and is calibrated well to make its view field the same as the sensing area. This dual-modal sensor can then simultaneously record the cells' visual profiles and EIT measurements.

B. Guidance Image Processing

Guidance image processing containing four steps converts the guidance image I^g into its corresponding mask image I^m (see Fig. 4). The size of I^m is the same as that of the expected EIT image, which occupies a circular region inscribed in a 64×64 square region, while the size of I^g is much larger than it. I^g also occupies a circular region, but this circle inscribes in a 406×406 square region. It should also be noted that I^g has three color channels, i.e., R, G and B. Therefore, this algorithm starts with the processing of the high-resolution RGB image I^g.

In I^g, the illumination often causes shadow, which is invalid information and significantly affects the target segmentation. Besides, as the structure of the targets is only desired, preservation of color has seldom significance. Therefore, the first step is to obtain the 1D illuminant invariant image I^{inv} of I^g following the methods proposed by Finlayson et al. [36] in order to convert I^g into a grey-scale image while removing the influence of illumination. The equation is formulated as:

$$I^{inv}(r,c) = \exp\left(\chi_1(r,c)\cos(\Theta) + \chi_2(r,c)\sin(\Theta)\right) \quad (10)$$

where r and c are pixel indexes. Θ is the projection direction in the 2D log-chromaticity space of I^g which is a constant for a specific camera. This direction leads to the minimum Shannon's entropy for I^{inv} and can be approximately obtained by traversing every integer angle from $1°$ to $180°$. $\chi_1(r,c)$ and $\chi_2(r,c)$ is expressed as:

$$[\chi_1(r,c), \chi_2(r,c)]^T = U\rho(r.c) \quad (11)$$

Here, U is a 2×3 orthogonal matrix and take the value of $U = [v_1, v_2]^T$, $v_1 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right]^T$, $v_1 = \left[\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right]^T$. $\rho(r.c)$ is defined by:

$$\rho(r,c) = \left[\ln\left(\frac{R(r,c)}{\Xi(r,c)}\right), \ln\left(\frac{G(r,c)}{\Xi(r,c)}\right), \ln\left(\frac{B(r,c)}{\Xi(r,c)}\right)\right]^T \quad (12)$$

where $\Xi(r,c) = \sqrt[3]{R(r,c)G(r,c)B(r,c)}$. $R(r, c)$, $G(r, c)$ and $B(r, c)$ are the three components of a color image.

Then, the binary version of I^{inv} can be generated by using the following thresholding segmentation method:

$$I^{bw}(r,c) = \begin{cases} 0, & \text{if } I^{inv}(r,c) < \beta \\ 1, & \text{if } I^{inv}(r,c) \geq \beta \end{cases} \quad (13)$$

where I^{bw} denotes the binary image after thresholding. The threshold value β is selected based on empirical trials. However, such method may lead to irregular boundary and randomly distributed white pixels. To address this issue, the third step applies morphological operations to I^{bw} to acquire a clean binary image with boundary-regular targets. In this paper, open operation (14) and dilation operation (15) are successively applied to reduce background irrelevant information and recover accurate target profiles. The two operations are defined as [37]:

$$I^{bw1} = I^{bw} \circ S = \cup \{(S)_z | (S)_z \subseteq I^{bw}\} \quad (14)$$

$$I^{bw2} = I^{bw1} \oplus S = \left\{z \middle| \left(\hat{S}\right)_z \cap I^{bw1} \neq \emptyset\right\} \quad (15)$$

where, $I^{bw} \circ S$ means I^{bw} is opened by the structuring element S and $I^{bw1} \oplus S$ means I^{bw1} is dilated by S. $(S)_z$ and $\hat{S}$ are defined as [37]:

$$(S)_z = \{k|k = a + z, a \in S\} \quad (16)$$

$$\hat{S} = \{w|w = -a, a \in S\} \quad (17)$$

where $z = (z_1, z_2)$ is a fixed point in the image space where I^{bw} and I^{bw1} exist.

I^{bw2} already provides the expected structural information, but it cannot be directly used as the input of the MSFCF-Net. As stated at the beginning of this subsection, the size of I^m is required to be the same as that of the EIT image. In addition, the height and width of images generated by the first three steps (i.e., I^{inv}, I^{bw} and I^{bw2}) is the same as those of I^g. Therefore, the final step is to down-sample I^{bw2} into a smaller circular image internally tangent with a 64×64 square region. The resulting smaller image is the I^m, and it is exactly the tiny version of I^{bw2}.

C. Multi-scale Feature Cross Fusion Network

MSFCF-Net reconstructs an EIT image I^{eit} from a frame of voltage measurements $\Delta V^* \in \mathbb{R}^{104}$ and a mask image I^m. We describe I^{eit} and I^m with a tensor of size $C \times 64 \times 64$, where $C = 1$ denotes the number of channels for a multi-channel image. ΔV^* and I^{eit} are defined by:

$$\Delta V^* = \frac{V_{\sigma_1} - V_{\sigma_0}}{V_{\sigma_0}} \quad (18)$$

$$I^{eit} = -\frac{\sigma_1 - \sigma_0}{\sigma_0} \quad (19)$$

As stated in Section II, td-EIT aims to recover $\Delta\sigma = \sigma_1 - \sigma_0$ from $\Delta V = V_{\sigma_1} - V_{\sigma_0}$. In this work, we adopt the relative changes format to facilitate the training of the deep learning model [38].

Our goal is to learn an end-to-end mapping F from ΔV_i^* and I^m to I^{eit}. Given a training dataset $\{\Delta V_i^*, I_i^m, I_i^{eit}\}_{i=1}^N$, the problem can be described as:

$$\hat{\theta} = \arg\min_\theta \frac{1}{N} \sum_{i=1}^N L(F_\theta(\Delta V_i^*, I_i^m), I_i^{eit}) + \frac{\lambda}{2}\|\theta\|_2^2 \quad (20)$$

where the second term is l_2 regularization with a penalty parameter λ, which can reduce over-fitting. $\theta = \{W, b\}$ represents the weights and bias of MSFCF-Net. L is the loss function to minimize the difference between I_i^{eit} and $F_\theta(\Delta V_i^*, I_i^m)$. As EIT image reconstruction is a regression problem, the mean squared error loss function is used, and L is defined as:

$$L(F_\theta(\Delta V_i^*, I_i^m), I_i^{eit}) = \left\|F_\theta(\Delta V_i^*, I_i^m) - I_i^{eit}\right\|_2^2 \quad (21)$$

The architecture of MSFCF-Net is shown in Fig. 5. Subnetworks in MSFCF-Net can be divided into three categories, i.e. the backbone networks, dual-modal feature fusion modules, and multi-scale feature fusion modules.

1) Backbone Networks (BN)

The backbone network extracts latent features from inputs. Thus, this network should have a good ability of feature extraction. The Darknet as the backbone of YOLOV3 is proved effective and powerful in feature extraction [39]. Inspired by its architecture, we designed the Darknet-like backbone networks for our application. The backbone network for voltage measurements (BN-V) has three additional fully connected layers followed by a reshape operation because of the dimension difference between Q and I^m (see Fig. 5). The output of the reshape operation is a feature map with the size of $1 \times 64 \times 64$. The rest of BN-V is the same as the backbone network for mask image (BN-M), which consists of five

residual blocks denoted by Res(n). Res(n) starts with left-and-upper zero padding followed by a *Conv + Leaky ReLU* unit with *Kernel Size = 3 × 3, Stride Step = 2*, and the number of kernels is twice as that of input feature maps. Then *n* residual units (represented by Res Unit, see Fig. 2) follow. The idea of Res Unit is proposed in [40], in which the short connection can make the deep network easier to train. Therefore, the combination of mentioned components in Res(n) will make the height and width of the output feature maps half than those of input feature maps while the number of feature maps doubles.

2) Dual-modal Feature Fusion Module (DMFF)

Dual-modal Feature Fusion Modules (DMFF) fuse information from different sources (see Fig. 6). To maintain the main information and eliminate the trivial ones, the attention mechanism originally used in natural language processing [41] is adopted in DMFF. As the feature maps generated by each layer in CNN have both channel dimension and spatial dimension, there are two types of attention mechanism, i.e., channel-wise attention and spatial-wise attention. In BN-V and BN-M, the spatial dimension gradually decreases with the increase of the number of layers. For the feature maps with a small spatial dimension, the spatial information is lost, and information carried by this type of feature maps is usually called semantic information. The spatial relationship between each element of the feature maps is trivial. Therefore, there are two types of DMFF in MSFCF-Net, i.e., DMFF-V1 and DMFF-V2 (see Fig. 5). DMFF-V1 corresponds to the feature maps with large spatial dimension, and it will incorporate both channel-wise attention and spatial-wise attention. DMFF-V2 corresponds to the feature maps with small spatial dimension, and it will only incorporate channel-wise attention. The implementation of attention mechanisms adopts the convolutional block attention modules proposed in [42], which is proved to be an effective and efficient method. Suppose the mapping of channel attention module in CBAM is denoted by f^{CA} and that of spatial attention module is denoted by f^{SA}, the mappings of both DMFF-V1 and DMFF-V2 can be uniformly expressed as:

$$S_v = f^A \left(f_{1,1}^{CL} \left(f_{3,1}^{CL} \left(f_{1,1}^{CL} (M_v) \right) \right) \right) \tag{22}$$

$$S_m = f^A \left(f_{1,1}^{CL} \left(f_{3,1}^{CL} \left(f_{1,1}^{CL} (M_m) \right) \right) \right) \tag{23}$$

$$M_{dm} = R^3([S_v, S_m]) \tag{24}$$

where M_v is the feature map from BN-V and M_m is the feature map from BN-M. The size of the feature maps S_v, S_m, M_{dm}, M_v and M_m is the same. f^A equals to f^{CA} for DMFFM-V1 and equals to $f^{SA} \circ f^{CA}$ for DMFFM-V2, which is the only difference between the two modules. f^{CL} denotes the mapping for *Conv + Leaky ReLU* unit. The first subscript means the kernel size and the second means the convolution step used in the convolution layer in this unit. $[\cdot, \cdot]$ denotes the concatenation operation and R^3 means the mapping of Res(3).

3) Multi-scale Feature Fusion Module (MSFF)

Feature maps of different scales will provide information on different scales. It will generate a more precise result if the
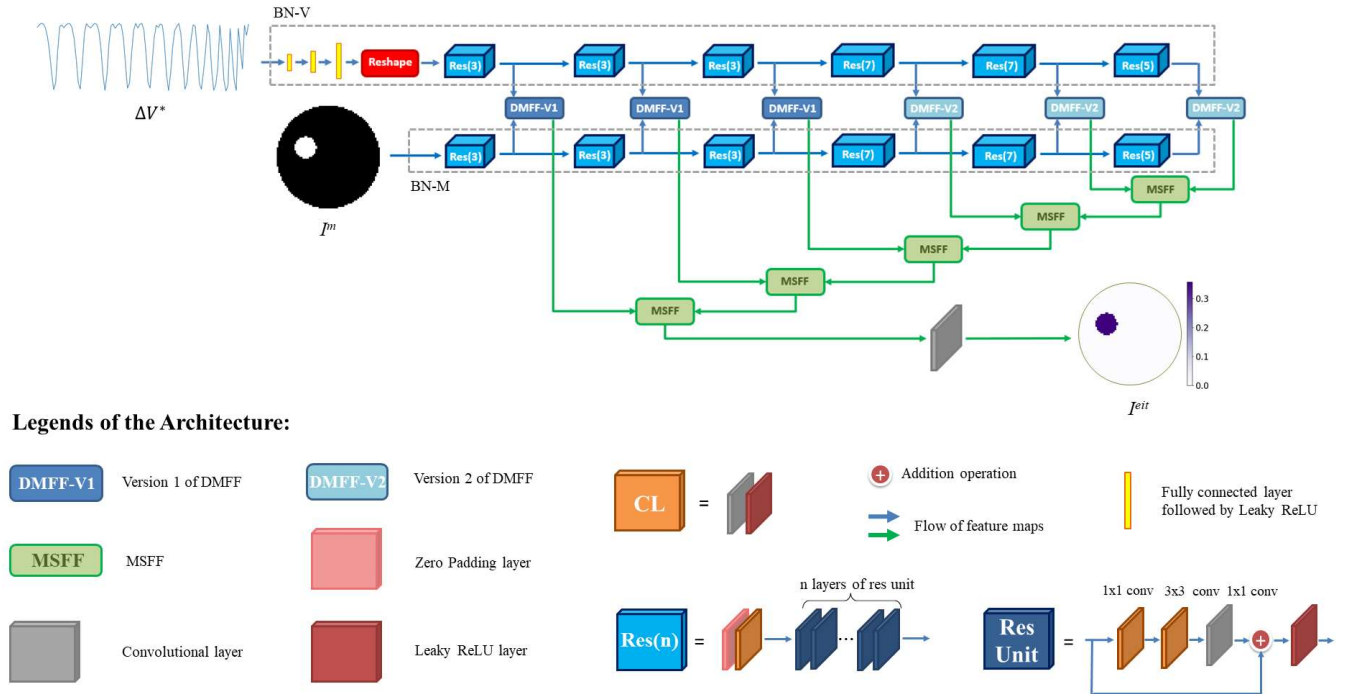

Fig. 5 Architecture of MSFCF-Net. Note, the color of arrow is only for indicating the feature maps flowing to different function block. BN-V and BN-M are the layers in dashed black squares.
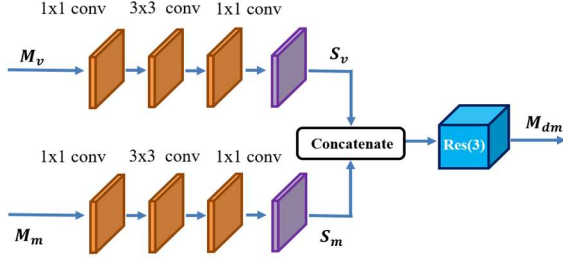
Fig. 6. Architecture of DMFF. The purple block means the mapping f^A. f^A equals to f^{CA} for DMFFM-V1 and f^A equals to $f^{SA} \circ f^{CA}$ for DMFFM-V2. The meaning of other components is the same as those of legends in Fig. 5.
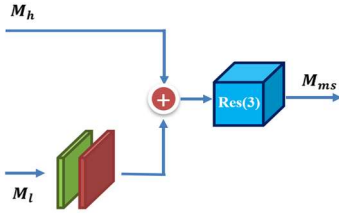

Fig. 7. Architecture of MSFF. The green block means transposed convolution layer. The meaning of other components is the same as those of legends in Fig. 5.

information of different scales can be integrated together. Many work in computer vision and image processing demonstrates that fusing feature maps of different scales is an efficient way to improve the performance of the network [39][43][44]. In addition, Chen et al. [21] and Li et al. [20] both adopted this method in their work on EIT image reconstruction and showed good results. MSFCF-Net also adopts the same idea and the multi-scale feature fusion module (MSFF, see Fig. 7) undertakes this function. MSFF module uses a simple way to perform information fusion. First, the spatial dimension of low scale feature maps will be enlarged twice by transposed convolutional layers followed by the *Leaky ReLU* layer. Then, the output of *Leaky ReLU* layer and the output of the network block before the current block will be added together in MSFF. The addition operation here is inspired by the work on human eye-fixation prediction, where the authors also face a dual-modal information fusion problem and fuse information of different scales by addition operation [45]. Like YOLOV3, the initial fused feature maps will be fed into multiple layers for thorough information fusion. Instead of using successive convolutional layers, the basic module in BN, i.e. Res(n), is used in MSFF to conduct post information fusion. Because this module has a satisfactory feature extraction ability while it can prevent the degradation of the network [40]. Finally, the output of the current MSFF will be the input of the next MSFF. The mapping of MSFF can be represented as:

$$M_{ms} = R^3 \left(M_h + f_{2,2}^{TCL}(M_l) \right) \tag{25}$$

where, M_l is the low-scale feature map and M_h is the high-scale feature map. The size of the output feature map M_{ms} is the same with that of M_h. $f_{2,2}^{TCL}$ represents the mapping for *Transposed Conv + Leaky ReLU* unit. The first subscript means the kernel size and the second represents the convolutional step in the convolution layer of this unit. R^3 means the mapping of Res(3).

IV. DATA GENERATION AND EXPERIMENTAL SETUP

A. Sensor Modelling and Dataset Generation

We establish the training, validation and test sets to train and evaluate the proposed MSFCF-Net. In COMSOL Multiphysics we modelled the same 2D 16-electrode circular EIT sensor as that in Fig. 1 and solved the forward problem of EIT to generate simulation data. The EIT forward problem is approximately solved by the Finite Element Method (FEM), which is the primary source leading to the modelling error. We adopted the time-difference imaging method, which could eliminate the common errors to a certain extent.

To make the deep learning model suitable for 3D cell culture imaging, specifically for 3D cell spheroids imaging, we consider multi-level, multi-circular-object conductivity distributions. In the sensing region, we generate four types of data and a sample belonging to a certain type of data includes a fixed number of objects (from one to four). For a certain type of data, for example, the one including three objects, we assign three non-overlapping circular objects with random diameters (from 0.07 d to 0.3 d, d is the diameter of the sensing area), positions, and conductivity values (from $0.0001\ S \cdot m^{-1}$ to $0.0475\ S \cdot m^{-1}$). The background conductivity is $0.05\ S \cdot m^{-1}$. As the spatial resolution of EIT is about 10% of the sensor diameter [46], we cannot set the diameter of the object too small. The selection of 0.07 d alre ady poses a challenge to image reconstruction. The maximum diameter of the objects can cover most 3D cell culturing situations. Thus, the setting of the object diameter range is reasonable. For conductivity settings, if the conductivity of a certain object is very close to the background conductivity, the voltage measurements can hardly be distinguished from those without such object in the sensing region. In this case, a frame of voltage measurements might correspond to two different mask images (one contains the object and the other does not), preventing the deep learning models from extracting the main features from the training set. Therefore, the criteria avoid that the conductivity of an object is too close to the background conductivity. Mask images for training and evaluation are also generated in simulation by a simple approach of assigning number one to pixels where there are objects and number zero to the rest of pixels. Four examples of the simulated conductivity images and corresponding binary mask images are illustrated in Fig. 8.

Based on the settings above, we built a dataset with 19,177 samples. Each sample comprises a frame of voltage measurements, a ground-truth conductivity image, and a mask image. There are 4,691 1-object samples, 4,736 2-object samples, 4,936 3-object samples and 4,814 4-object samples. In
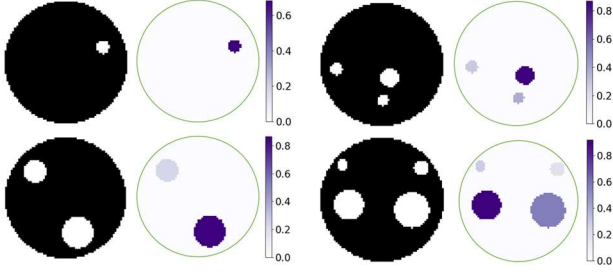
Fig. 8. Examples of simulated conductivity images and corresponding mask images. For each pair, the left is the binary mask image, and the right is the conductivity image. Green circle denotes the boundary of the sensing region.

Fig. 9. The composition of the final augmented dataset.

order to maintain the data balance in training and evaluation, we randomly select 10% samples as the test set and 10% samples from the remaining data as the validation set for each type of data. The rest will serve as the training set. As a result, we have 15,537 samples for training, 1,724 samples for validation, and 1,916 samples for testing.

Moreover, to improve the robustness of our model, additive Gaussian noise is added to the voltage measurements to augment the original dataset. For each type of data (i.e. 1-object samples, 2-object samples, 3-object samples, and 4-object samples), we separately add noise with the Signal-to-Noise Ratio (SNR) of 50 dB and 40 dB to a half of the data in the training and validation set. For the test data, the noise with the SNR of 50 dB, 40 dB and 30 dB is separately added to the entire test set. Fig. 9 displays the composition of the final dataset adopted to train and evaluate our model.

B. Dual-modal Imaging System Setup

The dual-modal sensor is connected to the in-house developed EIT system [47] to collect real-world experimental data. The frame rate of the system is set as 48 fps and its highest SNR is 82.82 dB [47]. In experiments, the frequency of the injected current is set as 10 kHz. In addition, the view field of the digital microscope and the sensing area of the impedance sensor coincide precisely.

C. Network Training

The MSFCF-Net is implemented using Pytorch, trained and tested on a workstation with a GeForce RTX 2070 Super. AdamW [48] is employed for optimization. We use the whole training set (31,074 samples) and the whole validation set (3,448 samples) to train MSFCF-Net. Early stopping is adopted to mitigate overfitting. The hyper parameters are set as follows: the learning rate is 10^{-4} and the penalty parameter λ is set as 10^{-6}; the maximum number of training epoch is 200 and the batch size of each update is 120; the tolerance is set as 10 epochs for early stopping. Finally, the training process is stopped at epoch 90 and the training time is 86.70 minutes. The model with minimum validation loss is selected as the final model.

TABLE I
QUANTITATIVE METRICS FOR COMPARING DIFFERENT ALGORITHMS ON TEST SET
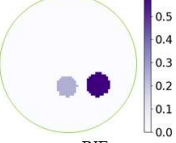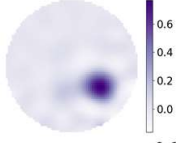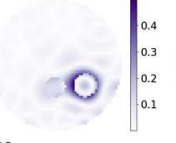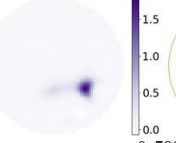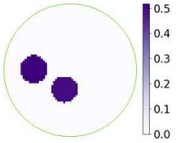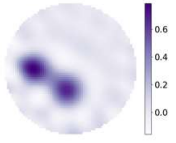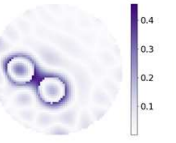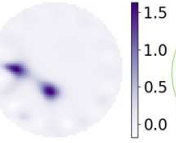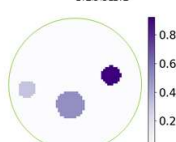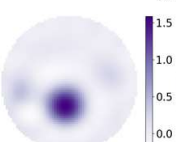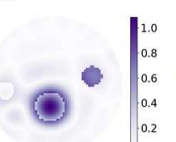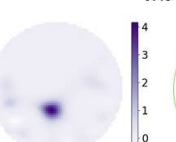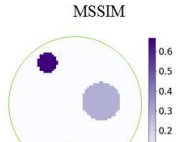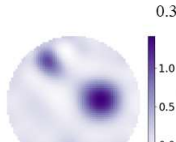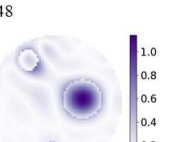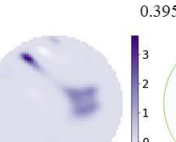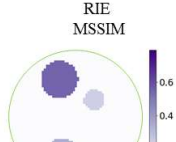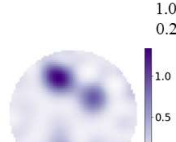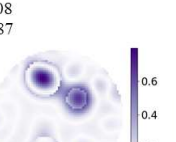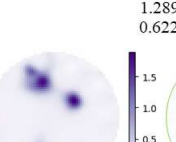
Algorithms	Noise-free		50 dB		40 dB		30 dB	
	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM
TReg-GL	0.9481	0.3879	0.9481	0.3875	0.9484	0.3782	0.9494	0.3592
SBL	1.5527	0.7553	1.5517	0.7546	1.5536	0.7389	1.5610	0.7140
CG	0.9278	0.4035	0.9279	0.4028	0.9281	0.3975	0.9297	0.3774
FC-UNet	0.4946	0.8708	0.4949	0.8707	0.4955	0.8703	0.5043	0.8665
S-MSFCF-Net	0.5150	0.8453	0.5151	0.8453	0.5158	0.8451	0.5185	0.8432
MSFCF-Net	**0.3715**	**0.9387**	**0.3715**	**0.9387**	**0.3715**	**0.9387**	**0.3715**	**0.9387**

TABLE II
COMPARISON OF DIFFERENT ALGORITHMS ON DIFFERENT TYPES OF SAMPLES

Algorithms	1-Object		2-Object		3-Object		4-Object	
	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM	M-RIE	M-MSSIM
FC-UNet	0.2334	0.9817	0.4803	0.9096	0.6018	0.8286	0.6756	0.7562
S-MSFCF-Net	0.2780	0.9644	0.4742	0.8868	0.6172	0.8008	0.7034	0.7214
MSFCF-Net	**0.0826**	**0.9920**	**0.4043**	**0.9478**	**0.4872**	**0.9182**	**0.5229**	**0.8940**

IMAGE RECONSTRUCTION RESULTS OF MODEL-BASED ALGORITHMS ON FIVE REPRESENTATIVE SAMPLES (LEFT COLUMN: RECONSTRUCTION; RIGHT COLUMN: ERROR IMAGE)

V. RESULTS AND DISCUSSION

The proposed method is evaluated by numerical simulation and MCF-7 cell spheroids experiments. The performance of MSFCF-Net is compared with other widely used single-modal based EIT image reconstruction algorithms, i.e., Gaussian-Laplace regularization (TReg-GL) [49] and Sparse Bayesian Learning (SBL) [50], and a dual-modal based image reconstruction algorithm using Cross-Gradient regularization (CG) [31]. In this work, the mask image replaces the CT image in [31] as the assisted image in both simulation and real experiments. We also compare with the recently proposed end-to-end deep learning model FC-UNet [21] and the single-modal version of MSFCF-Net (named S-MSFCF-Net). FC-UNet is originally designed for pixel-level classification for EIT image. As we treat the conductivity distribution prediction as a regression problem, to make a fair comparison, we remove the
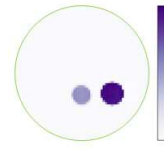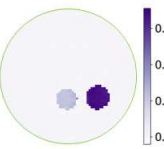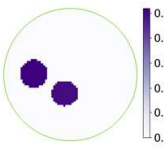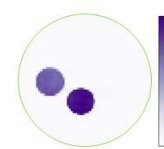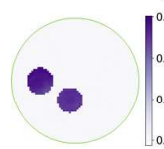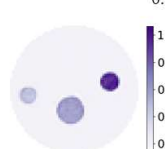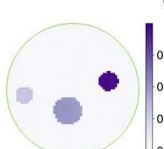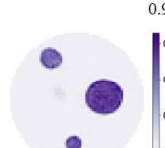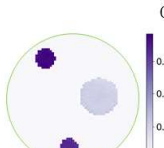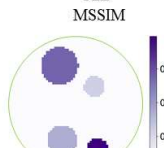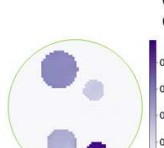
activation function in the output layer of the FC-UNet. For S-MSFCF-Net, we remove the BN-M and DMFF modules while the MSFF modules fuse different scales of feature maps from BN-V. Except for the tolerance of early stopping, FC-UNet and S-MSFCF-Net are also trained with the same loss function and settings as MSFCF- Net. The early stopping tolerance for FC-UNet and S-MSFCF-Net is set as 20 because it is beneficial to promote their convergence. It should be noted that in all reconstructed EIT images, the magnitude of each pixel denotes the quantity in (19).

A. Numerical Simulation

Relative Image Error (RIE) and mean structural similarity index (MSSIM) [51] are used to quantitatively evaluate the image quality, which are defined as:

TABLE IV
IMAGE RECONSTRUCTION RESULTS OF LEARNING-BASED ALGORITHMS ON FIVE REPRESENTATIVE SAMPLES (LEFT COLUMN: RECONSTRUCTION; RIGHT COLUMN: ERROR IMAGE)

$$\text{RIE} = \frac{\|\boldsymbol{A} - \boldsymbol{B}\|_2}{\|\boldsymbol{B}\|_2} \qquad (26)$$

$$\text{MSSIM} = \frac{1}{wh} \sum_r \sum_c \text{SSIM}(r, c) \qquad (27)$$

where $\boldsymbol{A}$ is the image to be evaluated and $\boldsymbol{B}$ is the selected reference image. r and c are the position indexes of an image. w and h are the weight and height of an image, respectively. $\text{SSIM}(r, c)$ is the structural similarity index map [51], and is defined as:

$$\text{SSIM}(r, c) = \frac{(2\mu_A\mu_B + C_1)(2\delta_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\delta_A^2 + \delta_B^2 + C_2)} \qquad (28)$$

where μ_A, μ_B, δ_A, δ_B, and δ_{AB} are the local means, standard deviations and cross-covariance for image $\boldsymbol{A}$ and $\boldsymbol{B}$, which are
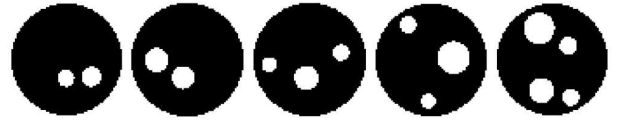
Fig. 10. Mask images corresponding to samples in Table III and Table IV.

also calculated following methods in [51]. $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$. K_1 and K_2 are constants whose values are 0.01 and 0.03, respectively. As the range for reconstructed EIT images in this work is [0, 1], L is set as 1.

RIE and MSSIM are the metrics for the evaluation of single image quality. Another two numerical metrics that evaluate the performance on the whole dataset level are the mean RIE (M-RIE) and the mean MSSIM (M-MSSIM). During the evaluation process, we calculate RIE and MSSIM for each image in the test set and then average all values.
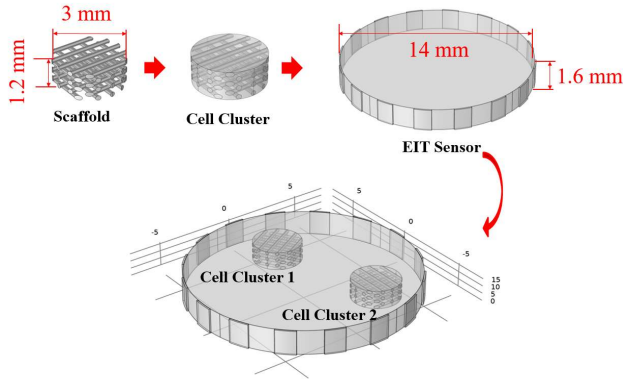
Fig. 11. Modelling of the quasi-2D EIT sensor equipped with two scaffolds.

TABLE V
IMAGE RECONSTRUCTION RESULTS BASED ON QUASI-2D EIT SENSOR

TABLE VI
IMAGE RECONSTRUCTION RESULTS BASED ON PERTURBED MASK IMAGES

Table I illustrates the quantitative evaluation results at different SNR levels on the test set. It is evident that the metrics of MSFCF-Net are superior to other given algorithms, indicating the robustness and effectiveness of the proposed dual-modal framework. Deep learning-based methods all show better voltage noise-resistance capability than conventional model-based algorithms. Especially, the M-MSSIM of deep learning models maintains a similar level with the decrease of SNR. Tables II compares the metrics of deep learning models on different types of samples with the SNR=50dB. As reconstructing multi-object and multi-level conductivity distribution is much more challenging, the metrics on 2-object samples have a big drop than those on 1-object samples for all deep learning models. For a specific type of samples, it is evident that the performance of MSFCF-Net is much better than the other two. Especially, though S-MSFCF-Net only removes the mask image related struc tures from MSFCF-Net, it still cannot reach the performance of MSFCF-Net. The reason is that: single-modal deep learning models will take the duty on both position and structure prediction and conductivity value prediction. But the proposed dual modal deep learning model in essence utilizes more structural information; thus, better conductivity prediction can be expected.

Table III and Table IV compare five representative phantoms reconstructed from test data with SNR=50dB. GT denotes the ground truth image. The left column under each algorithm is the reconstructed EIT image and the right one is the error image which is the absolute difference between the reconstructed image and the ground truth image. Mask images (from left to right) corresponding to samples in Table III and Table IV (from top to bottom) are illustrated in Fig. 10. Although both TReg-GL and SBL can predict the position of objects, the shape and conductivity values are always inaccurate (see their error images, RIE and MSSIM). For CG, the reconstructed images are very similar to images by TReg-GL and the quality of images is not improved noticeably according to their error images and numerical metrics. However, if the image generated by CG is zoomed, it is obvious that clear boundaries of objects are visible, which is exactly the result of introducing Cross-Gradient regularization. Thus, the Cross-Gradient regularization can only augment the object boundaries based on the assisted image while it cannot essentially improve the EIT image quality. For deep learning-based approaches, FC-UNet and S-MSFCF-Net can generate more accurate position, shape and conductivity values, but the errors are still more significant than those of the MSFCF-Net. Only MSFCF-Net can reconstruct the best EIT images among the given

algorithms with the most accurate position, shape, and conductivity values. Especially, the results of the second row and the fourth row indicate that the MSFCF-Net can reconstruct images correctly that the other two networks cannot do, which benefits from introducing another imaging modality.

In this study, all model-based algorithms are based on the linearized EIT forward model introducing intrinsic model error. Therefore, it is difficult for these linearized model-based methods to predict accurate results, which may lead to under or over estimation of conductivity values. Deep-learning-based methods directly fit the non-linear mapping of the EIT inverse problem, which can theoretically generate more accurate predictions. However, the performance of learning-based methods highly depends on the quality of training datasets and training strategies. We produce a large dataset and carefully train the proposed network (depicted in Section IV-C) to mitigate the adverse effect caused by such limitations.

In many tissue engineering applications, cells are cultured within the scaffold and monitoring of cell growth is vital to the process [8]. Cell growth at different stages within the scaffold will decrease the conductivity of various levels, which can be mapped by EIT [9]. To further demonstrate the effectiveness of the proposed framework, we simulated the imaging of cell growth within bio-scaffolds [8][9] by using EIT. We modelled a regular-shape scaffold, a quasi-2D EIT sensor with 16 electrodes and a cell culture model with two cell clusters (see

Fig. 11). The modelled sensor has the same dimension as the real sensor in Fig. 3. The height and diameter of the scaffold are 1.2 mm and 3 mm, respectively. The background conductivity is set as 0.05 S/m and the conductivity of the scaffold material is set as 10^{-8} S/m. The cells are modelled as evenly distributed in the space among scaffolds. We modelled two scenarios. The first contains one cell cluster and the conductivity of the cells cluster is set as 0.025 S/m. The second has two cell clusters to simulate cell growth at two different stages. The conductivities in cell cluster 1 and cell cluster 2 are set as 0.02 S/m and 0.04 S/m, respectively. In these cases, the reference conductivity distribution for the first scenario is the homogeneous medium whose conductivity is 0.05 S/m with a scaffold and the reference conductivity distribution for the second scenario is the same homogeneous medium including two scaffolds. Therefore, only the cell's conductivity contributes to the predicted conductivity variation, which is also indicated in [8].

Table V gives the image reconstruction results under settings described in last paragraph. By using the proposed approach, we could obtain reconstruction images with RIE lower than 0.22 and SSIM larger than 0.96. The results show strong evidence that the proposed framework can generate accurate conductivity distribution under a different setting. It also presents good generalization ability when dealing with the challenging scenario of scaffold-based cell culturing imaging.

It is worth further discussing the generalization ability and the limitations of our method. In practical applications of the
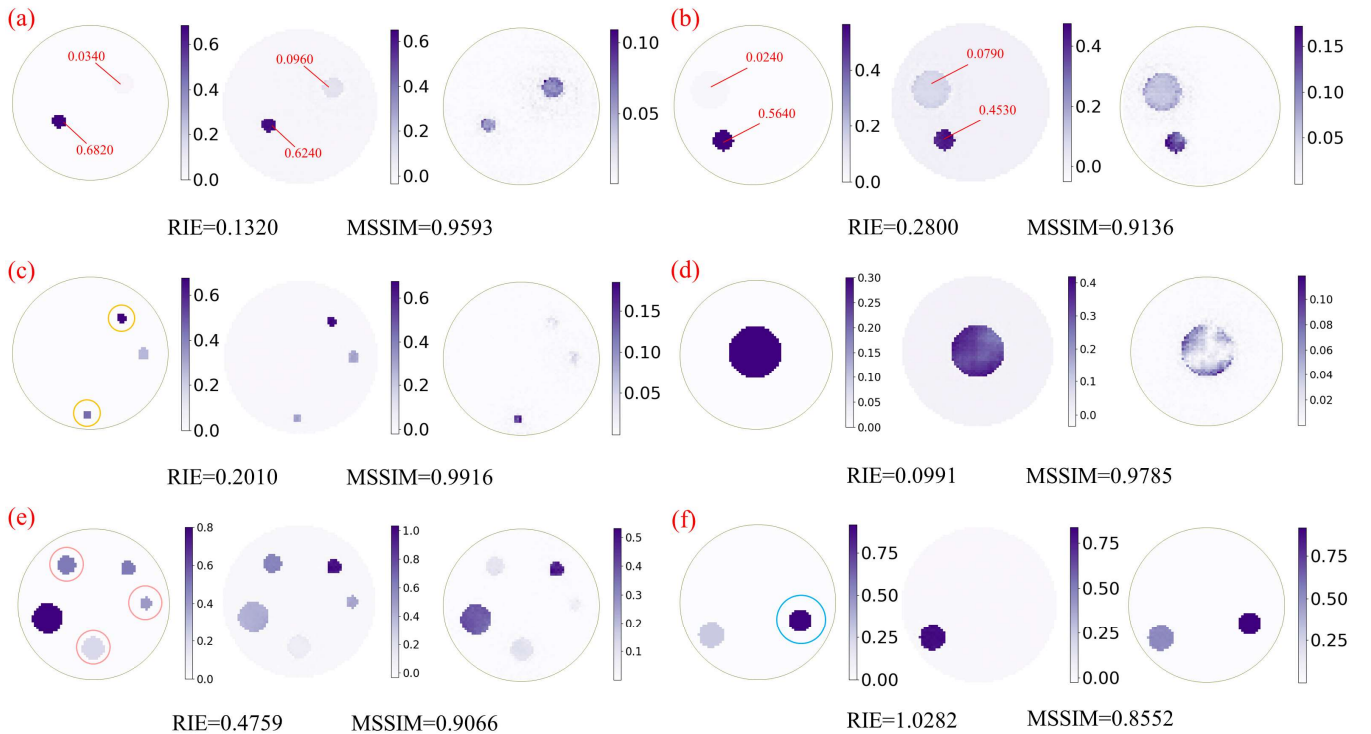
Fig. 12. MSFCF-Net image reconstruction results of six phantoms (a)-(f) based on the data which do not satisfy our dataset construction criteria. From left to right, each column represents the ground truth, reconstructed EIT image and error image, respectively. Red numbers in (a) and (b) denote the relative conductivity change, i.e. l^{eit} in (19), at the sampling points in the images. The size of the inclusions in the yellow circles in (c) is smaller than 0.07 d. The inclusions in the rose circles in (e) are the ones that are well predicted. The inclusion in the blue circle in (f) means this one is not identified by the mask image.

TABLE VII
COMPARISON BETWEEN DIFFERENT ALGORITHMS ON MCF-7 CELL EXPERIMENTS
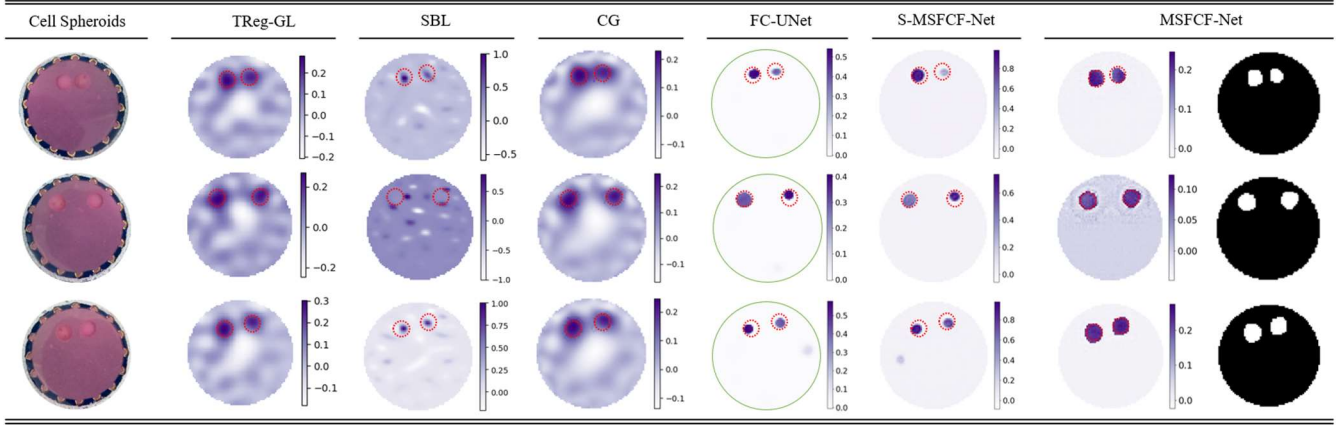
proposed method, inaccurate mask images may be generated due to many factors, such as unideal guidance image processing algorithm or noisy guidance image. To assess the robustness of the proposed method when encountering an inaccurate mask image, Table VI selects the first and fifth samples in Table III (or Table IV) for further analysis. Two different random perturbations are applied to the mask images of each sample, which are shown in the first column of the table. Each result occupying one row contains three images, i.e. the input mask image, the predicted EIT image and the error image from left to right. Observing the error images, the conductivity value can still be predicted accurately except for the pixels on the boundary while losing some structural information. Compared with the results generated by MSFCF-Net in Table IV, although the quality of the image based on perturbed mask image is lower than the quality of that based on the accurate mask (see RIE and MSSIM), the quality of these images is still much better than the quality of images generated by the model-based algorithms. This analysis implies that, in real-world experiments, we can acquire a quantitatively meaningful EIT image even if the guidance image processing algorithm cannot generate a very accurate mask image.

In Fig. 12, we also discuss another five situations which are inconsistent with our criteria of data generation but may occur in real applications. Some of them are extreme cases. There are six pairs of results, and each pair of r esults includes three images, i.e. the ground truth, reconstructed EIT image and error image. Results in Fig. 12 (a) to (e) are based on the accurate mask images. Fig. 12 (a) and (b) display the results when the conductivity of one inclusion is close to the background conductivity. It is clear that our method can still recover the conductivity contrasts though the error of conductivity prediction exists. Fig. 12 (c) illustrates the result under the situation that the size of two inclusions (around 0.05 d) is smaller than 0.07 d, and Fig. 12 (d) shows the result under the situation that the size of the inclusion (around 0.35 d) is larger than 0.3 d. The results demonstrate that our method can still reconstruct the conductivity well if the sizes of certain inclusions never appear in our training set. Fig. 12 (e) is the

result of imaging five inclusions by our approach. This type of samples does not appear in the training set neither. It is obvious that the prediction error is much larger. However, the conductivity of three inclusions can still be approximately predicted, which can be indicated from the error image. Fig. 12 (f) shows the result based on the assumption that the mask image fails to identify the inclusion in the blue circle, which means the pixel values of that inclusion in the mask image are all zeros, while the mask image provides an accurate profile for the other inclusion. This extreme scenario may occur when the optical sensor has no response on some inclusions, or the guidance image processing algorithm is imperfect. In this case, only the inclusion identified by the mask image appears in reconstructed EIT image, indicating that identifying objects in the mask image is essential in our approach. This prerequisite is acceptable as our method conducts dual-modal imaging. To summarize, the proposed dual-modal imaging approach can generate satisfactory images if the inclusions could be correctly recognized by the optical image and the guidance image processing algorithm. Our method demonstrates generalization ability and robustness when encountering voltage noise, mask image perturbation and some situations not satisfying our data generation criteria.

B. Cell Experiments

The performance of the proposed framework is further evaluated on data collected from real-world experiments (see Table VII). The imaging target is MCF-7 cell spheroids (diameter ~ 2 mm). The rightest column is the mask image generated from the guidance image processing algorithms stated in Subsection B, Section III. The threshold values β in (13) for the three guidance images (from top to bottom) are set as 0.66, 0.45 and 0.5 respectively based on a series of trials. The 3×3 kernel is adopted in the next two morphological operations for all cases.

In Table VII, the red dash line denotes the location of the cell spheroids. For conventional model-based single-modal and dual-modal algorithms, the cell spheroid structure is lost, and the reconstructed images contain too much unmeaningful

information in the background, although they can locate the position of cell spheroids. As discussed in Subsection A, Section V, the augmented object boundary is visible in images reconstructed by CG while the quality of these images is not essentially improved in the visual. Although the model-based algorithms are defeated in terms of shape preservation and noise reduction, TReg-GL and CG generate acceptable conductivity estimation (for TReg-GL, ~ 0.30 for the first and third phantom, ~ 0.25 for the second phantom; for CG, ~ 0.22 for all phantoms) according to the estimated relative conductivity change of MCF-7 cells (~ 0.39) using the simplified single-shell model and computing method in [52]. The deep learning models outperform in artifacts suppression. The MSFCF-Net could generate the most accurate shape and acceptable conductivity change (~ 0.25 for the first phantom, ~ 0.12 for the second phantom, and ~ 0.30 for the third phantom) according to the estimated conductivity change of MCF-7 cells. The reconstructed condu ctivity change of the second phantom deviates more from the reference value, possibly due to the more considerable measurement noise than the other two phantoms. It is also worth noting that such a theoretical approximation based on several assumptions might not be accurate. In the future, a possible method to better quantify the experiment results is to use hydrogel as phantoms, whose shape and conductivity value could be better controlled.

VI. Conclusion and Future Work

In this paper, we proposed an impedance-optical dual-modal imaging framework for 3D cell culture imaging. We combined optical imaging with EIT to tackle the low image quality issue of EIT. We also developed a learning-based approach to fuse the dual-modal information and reconstruct high-quality conductivity images. The results on simulation data and real-world data on MCF-7 cell spheroids demonstrate that the proposed framework could generate a more accurate estimation of conductivity distribution, which implies the possibility of quantitative imaging for EIT in tissue engineering. Future research will deal with the situation when the structure of the object in the mask image suffers more severe perturbation and develop more advanced image processing algorithms to generalize the method to other optical imaging approaches for tissue engineering, e.g., optical coherence tomography.

References

[1] F. Pampaloni *et al.*, "The third dimension bridges the gap between cell culture and live tissue," in *Nature Reviews Molecular Cell Biology*, vol. 8, no. 10, pp. 839-845, 2007.

[2] E. Cukierman *et al.*, "Taking cell-matrix adhesions to the third dimension," in *Science*, vol. 294, no. 5547, pp. 1708-1712, 2001.

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

[4] L. Miao, Y. Ma and J. Wang, "ROI-based image reconstruction of electrical impedance tomography used to detect regional conductivity variation," in *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 2903-2910, 2014.

[5] J. Yao *et al.*, "Development of three-dimensional integrated microchannel-electrode system to understand the particles' movement with electrokinetics," in *Biomicrofluidics*, vol. 10, no. 2, p. 024105, 2016.

[6] R. Pethig and D. B. Kell, "The passive electrical properties of biological systems: their significance in physiology, biophysics and biotechnology," in *Physics in Medicine & Biology*, vol. 32, no. 8, p. 933, 1987.

[7] Y. Yang, J. Jia, S. Smith, N. Jamil, W. Gamal, and P. O. Bagnaninchi, "A miniature electrical impedance tomography sensor and 3-D image reconstruction for cell imaging," in *IEEE Sensors Journal*, vol. 17, no. 2, pp. 514-523, 2016.

[8] Y. Yang, H. Wu, J. Jia, and P. O. Bagnaninchi, "Scaffold-based 3-d cell culture imaging using a miniature electrical impedance tomography sensor," in *IEEE Sensors Journal*, vol. 19, no. 20, pp. 9071-9080, 2019.

[9] H. Wu, W. Zhou, Y. Yang, J. Jia, and P. O. Bagnaninchi, "Exploring the potential of electrical impedance tomography for tissue engineering applications," in *Materials*, vol. 11, no. 6, p. 930, 2018.

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

[11] G. González, V. Kolehmainen, and A. Seppänen, "Isotropic and anisotropic total variation regularization in electrical impedance tomography," in *Computers & Mathematics with Applications,* vol. 74, no. 3, pp. 564-576, 2017.

[12] B. Gong *et al.*, "Higher order total variation regularization for EIT reconstruction," in *Medical & Biological Engineering & Computing*, vol. 56, no. 8, pp. 1367-1378, 2018.

[13] A. Borsic, B. M. Graham, A. Adler and W. R. B. Lionheart, "In Vivo Impedance Imaging with Total Variation Regularization," in *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 44-54, Jan. 2010, doi: 10.1109/TMI.2009.2022540.

[14] K. Lee, E. J. Woo and J. K. Seo, "A fidelity-embedded regularization method for robust electrical impedance tomography," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 1970-1977, 2017.

[15] J. Wang, "Non-convex ℓ_p regularization for sparse reconstruction of electrical impedance tomography," in *Inverse Problems in Science and Engineering*, pp. 1-22, 2020.

[16] J. Li *et al.*, "Adaptive ℓ_p Regularization for Electrical Impedance Tomography," in *IEEE Sensors Journal*, vol. 19, no. 24, pp. 12297-12305, 2019.

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

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

[19] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," in *Nature*, vol. 521, no. 7553, pp.436-444, 2015.

[20] F. Li, and C. Tan and F. Dong, "Electrical Resistance Tomography Image Reconstruction with Densely Connected Convolutional Neural Network," in *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-11, 2021, Art no. 4500811, doi: 10.1109/TIM.2020.3013056.

[21] Z. Chen, Y. Yang, J. Jia and P. Bagnaninchi, "Deep Learning Based Cell Imaging with Electrical Impedance Tomography," *2020 IEEE International In strumentation and Measurement Technology Conference (I2MTC)*, Dubrovnik, Croatia, 2020, pp. 1-6, doi: 10.1109/I2MTC43012.2020.9128764.

[22] 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)*, Abu Dhabi, United Arab Emirates, 2019, pp. 1-6, doi: 10.1109/IST48021.2019.9010151.

[23] D. Smyl, T. N. Tallman, J. A. Black, A. Hauptmann, and D. Liu, "Learning and correcting non-Gaussian model errors," *Journal of Computational Physics,* vol. 432, p. 110152, 2021.[23]

[24] S. J. Hamilton and A. Hauptmann, "Deep D-Bar: Real-Time Electrical Impedance Tomography Imaging with Deep Neural Networks," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2367-2377, Oct. 2018, doi: 10.1109/TMI.2018.2828303.

[25] M. Capps and J. L. Mueller, "Reconstruction of Organ Boundaries with Deep Learning in the D-bar Method for Electrical Impedance Tomography," in *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 826-833, March 2021.

[26] Z. Wei, D. Liu and X. Chen, "Dominant-current deep learning scheme for electrical impedance tomography," in *IEEE Transactions on Biomedical*

Engineering, vol. 66, no. 9, pp.2546-2555, Sept. 2019, doi: 10.1109/TBME.2019.2891676.

[27] D. Smyl, T. N. Tallman, D. Liu, and A. Ha uptmann, "An efficient Quasi-Newton method for no nlinear inverse problems via learned singular values," *IEEE Signal Processing Letters,* vol. 28, pp. 748-752, 2021.

[28] G. Liang, S. Ren, S. Zhao and F. Dong, "A Lagrange-Newton Method for EIT/UT Dual-modality Image Reconstruction," in *Sensors*, vol. 19, no. 9, p. 1966, 2019.

[29] H. Liu, S. Zhao, C. Tan and F. Dong, "A Bilateral Constrained Image Reconstruction Method Using Electrical Impedance Tomography and Ultrasonic Measurement," in *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9883-9895, 1 Nov.1, 2019, doi: 10.1109/JSEN.2019.2928022.

[30] B. Gong, B. Schullcke, S. Krueger-Ziolek, U. Mueller-Lisse, and K. Moeller, "Sparse regularization for EIT reconstruction incorporating structural information derived from medical imaging," *Physiological measurement,* vol. 37, no. 6, p. 843, 2016.

[31] Z. Li, J. Zhang, D. Liu and J. Du, "CT Image-Guided Electrical Impedance Tomography for Medical Imaging," in *IEEE Transact ions on Medical Imaging*, vol. 39, no. 6, pp. 1822-1832, June 2020, doi: 10.1109/TMI.2019.2958670.

[32] S. Ren, K. Sun, C. Tan and F. Dong, "A Two-Stage Deep Learning Method for Robust Shape Reconstruction With Electrical Impedance Tomography," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 4887-4897, July 2020, doi: 10.1109/TIM.2019.2954722.

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

[34] E. Somersalo, M. Cheney and D. Isaacson, "Existence and uniqueness for electrode models for electric current computed tomography," in *SIAM Journal on Applied Mathematics*, vol. 52, no. 4, pp. 1023-1040, 1992.

[35] D. B. Geselowitz, "An Application of Electrocardiographic Lead Theory to Impedance Plethysmography," in *IEEE Transactions on Biomedical Engineering*, vol. BME-18, no. 1, pp. 38-41, Jan. 1971, doi: 10.1109/TBME.1971.4502787.

[36] G. D. Finlayson, M. S. Drew and C. Lu, "Entropy minimization for shadow removal," in *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35-57, 2009.

[37] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 3rd ed, Upper Saddle River, N.J: Pearson, 2008.

[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning*, PMLR, 2015, pp. 448-456.

[39] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, *arXiv preprint arXiv:1804.02767*.

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

[41] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate", 2014, *arXiv preprint arXiv:1409.0473*.

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

[43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu and A.C. Berg, "Ssd: Single shot multibox detector," In *European conference on computer vision*, Oct. 2016, pp. 21-37.

[44] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In *International Conference on Medical image computing and computer-assisted intervention*, Oct. 2015, pp. 234-241.

[45] W. Liu, W. Zhou and T. Luo, "Cross-Modal Feature Integration Network for Human Eye-Fixation Prediction in RGB-D Images," in *IEEE Access*, vol. 8, pp. 202765-202773, 2020, doi: 10.1109/ACCESS.2020.3036681.

[46] P. Metherall, D. C. Barber, R. H. Smallwood, and B. H. Brown, "Three-dimensional electrical impedance tomography," Nature, vol. 380, no. 6574, pp. 509-512, 1996.

[47] Y. Yang and J. Jia, "A multi-frequency electrical impedance tomography system for real-time 2D and 3D imaging," in *Review of Scientific Instruments*, vol. 88, no. 8, p. 085110, 2017.

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

[49] Y. Yang, J. Jia, N. Polydorides and H. McCann, "Effect of structured packing on EIT image reconstruction," *2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings*, Santorini, Greece, 2014, pp. 53-58, doi: 10.1109/IST.2014.6958445.

[50] S. Liu, J. Jia, Y. D. Zhang and Y. Yang, "Image Reconstruction in Electrical Impedance Tomography Based on Structure-Aware Sparse Bayesian Learning," in *IEEE Transactions on Medical Imaging*, vo l. 37, no. 9, pp. 2090-2102, Sept. 2018, doi: 10.1109/TMI.2018.2816739.

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

[52] Cottet, Jonathan, Olivier Fabregue, Charles Berger, François Buret, Philippe Renaud, and Marie Frénéa-Robin. "MyDEP: a new computational tool for dielectric modeling of particles and cells." *Biophysical journal* 116, no. 1 (2019): 12-18.