

Image-Decomposition-Enhanced Deep Learning for Detection of Rotor Cores in Cardiac Fibrillation

Yu Shu , Tianqi Gao Smith, Shivaram P. Arunachalam , Elena G. Tolkacheva ,
and Changqing Cheng 

I. INTRODUCTION

Abstract—Objective: Rotors, regions of spiral wave reentry in cardiac tissues, are considered as the drivers of atrial fibrillation (AF), the most common arrhythmia. Whereas physics-based approaches have been widely deployed to detect the rotors, in-depth knowledge in cardiac physiology and electrogram interpretation skills are typically needed. The recent leap forward in smart sensing, data acquisition, and Artificial Intelligence (AI) has offered an unprecedented opportunity to transform diagnosis and treatment in cardiac ailment, including AF. This study aims to develop an image-decomposition-enhanced deep learning framework for automatic identification of rotor cores on both simulation and optical mapping data. **Methods:** We adopt the Ensemble Empirical Mode Decomposition algorithm (EEMD) to decompose the original image, and the most representative component is then fed into a You-Only-Look-Once (YOLO) object-detection architecture for rotor detection. Simulation data from a bi-domain simulation model and optical mapping acquired from isolated rabbit hearts are used for training and validation. **Results:** This integrated EEMD-YOLO model achieves high accuracy on both simulation and optical mapping data (precision: 97.2%, 96.8%, recall: 93.8%, 92.2%, and F1 score: 95.5%, 94.4%, respectively). **Conclusion:** The proposed EEMD-YOLO yields comparable accuracy in rotor detection with the gold standard in literature.

Index Terms—Atrial fibrillation, convolutional neural network, ensemble empirical mode decomposition (EEMD), object detection, you only look once (YOLO).

ATRIAL fibrillation (AF) is the most common arrhythmia that affects more than 2.3 million people in the U.S. [1], [2], [3]. The untreated AF could lead to stroke and heart failure, and it is associated with elevated morbidity and mortality. Studies have suggested that AF may be maintained by a rapid localized pattern, known as “rotor” [4]. Targeted rotor ablation has been deployed to terminate AF in clinical practice [5], which calls for accurate and timely identification of the rotor cores. More recently, the universal availability of biomedical data brings tremendous opportunities for data-driven approaches in diagnosis. Techniques have been developed to characterize and identify rotors based on cardiac electrograms, including dominant frequency (DF), local activation time using isochronal maps, complex fractionated atrial electrogram mean index mapping, and phase singularity analysis both in animal studies and clinical applications [6], [7], [8], [9]. However, these methods demand domain knowledge and dedicated expertise to transform the raw data into a suitable representation to specify the phenotype, or require in-depth understanding of cardiac physiology and electrogram interpretation [10]. For example, estimating AF activation rates using DF relies on the understanding of sine-wave signal decomposition and its connection to activation rates detected from electrograms, which may increase clinical workload and induce bias [10].

In comparison, deep learning approaches reveal the sophisticated underlying patterns of the raw data without strong assumptions of the data generation mechanisms, which usually remain unknown or elusive to define [11]. Recent advances in computer vision and image recognition techniques has enabled the application of deep learning in rotor detection [12], [13], [14], [15]. For instance, U-Net, an encoder-decoder deep learning architecture, was developed for the detection of rotors on simulation data [14]. Lebert et al. explored rotor detection using different architectures, including U-Net alone (as used in [14]), a combination of U-Net and LSTM, as well as an autoencoder structure. They found that all of these architectures exhibited similar performance in detecting rotors based on computing phase maps and phase singularities (PS) derived from a short sequence of excitation wave patterns. Additionally, their research demonstrated that the AI-filtering methods could be trained on simulated data and then successfully applied to experimental

Manuscript received 12 December 2022; revised 6 April 2023 and 12 June 2023; accepted 27 June 2023. Date of publication 13 July 2023; date of current version 25 December 2023. This work was supported in part by the National Science Foundation under Award 2119334 and in part by the Northeast Big Data Innovation Hub Seed Fund. (Corresponding authors: Yu Shu; Changqing Cheng.)

Yu Shu and Changqing Cheng are with the Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902 USA (e-mail: yshu1@binghamton.edu; ccheng@binghamton.edu).

Tianqi Gao Smith is with the Mayo Clinic College of Medicine and Science and Department of Radiation Oncology, USA.

Shivaram P. Arunachalam is with the Department of Medicine and Radiology, Mayo Clinic, USA.

Elena G. Tolkacheva is with the Department of Biomedical Engineering, University of Minnesota, USA.

Digital Object Identifier 10.1109/TBME.2023.3292383

data. The rotor detection method was effective across different species, from the rabbit to the pig, indicating its generalizability [15]. While U-Net is a popular deep learning architecture, training of U-Net with deep layers can be problematic due to the vanishing gradient issues, and it may take rather long time for the network to converge in terms of the Intersection over Union (IOU). The IOU here is a measure of the overlap between the predicted and actual bounding boxes, and it is commonly used to evaluate the performance of segmentation models. Notably, most U-Net architectures are not robust to variations in the input data, such as changes in scale, or orientation. LSTM includes a set of specialized units, known as memory cells, to store information over time and selectively forget or update the information based on the input data sequence, while maintaining an internal state to capture information about the proceeding elements. This makes LSTM a particularly effective RNN model in language modeling, speech recognition, and machine translation. However, the LSTM networks are generally computationally expensive to train and evaluate, especially for data with a long sequence. They also suffer from the vanishing gradient problem, which leads to poor performance [16].

To eschew those pitfalls in the encoder-decoder algorithms, we employ an object detection algorithm for rotor detection. More specifically, we use You Only Look Once (YOLO), a well-established single-stage Convolutional Neural Network (CNN) object detection algorithm. It only involves one forward pass to detect objects of interest in an image, which is simpler and faster than other two-stage detection algorithms, and YOLO is designed to detect objects at different scales and aspect ratios, which makes it more accurate than other algorithms that rely on a fixed set of anchor boxes [17]. It is further noted that high-quality data are the premise for such AI-assisted diagnosis tools. The advent of cardiac imaging techniques, including ECG imaging, catheter-based electro-anatomic mapping, electrode contact mapping, and high-density electrical mapping, has offered an unprecedented opportunity to characterize the spatiotemporal evolution of cardiac electrical activities [15], [18]. However, such medical imaging data could be contaminated with noise and blurred by artifacts, which poses a tremendous challenge for effective learning algorithms [19], [20], [21]. Therefore, data pre-processing (e.g., image denoising and sharpening) is indispensable to enhance details and contrast in imaging data and facilitate the subsequent learning and recognition tasks [19]. Empirical mode decomposition (EMD), a data-driven signal processing filtering, which decomposes a signal into a collection of intrinsic mode functions, has demonstrated competitive performance in noise reduction and pattern detection [22]. Although different filtering techniques have been integrated into CNN architecture to pre-process medical images and improve learning accuracy [23], limited success has been reported. So far, image sharpening and AI in rotor detection are largely under-explored.

In this study, we compare several algorithms in pre-processing imaging data and employ the YOLO model to detect the rotor cores using both simulated and animal cardiac imaging data. We show that accuracy of this image-decomposition-enhanced deep

learning framework for automatic rotor detection is on par with the existing gold standard of the physics-based methods.

II. METHODOLOGY

A. Model and Data Description

We employ a bi-domain simulation to mimic the intracellular electrical coupling and propagation of electrical excitation [24], shown in (1)–(3), and characterize cardiac dynamics with stationary rotor in isotropic human atrial tissues [9]. Here, the stationary rotor pivots around a circular trajectory forming the core of the spiral wave. The bi-domain simulation captures the evolution of the extracellular potential φ_e , the myocyte transmembrane voltage V^{myo} , and the fibroblast transmembrane voltage V^{fib} :

$$\begin{aligned} \nabla \cdot \left((\sigma_i^{myo} + \sigma_i^{fib} + \sigma_e) \nabla \varphi_e \right) \\ = -\nabla \cdot (\sigma_i^{myo} \nabla V^{myo}) - \nabla \cdot (\sigma_i^{fib} \nabla V^{fib}) \end{aligned} \quad (1)$$

$$\begin{aligned} C^{myo} \frac{\partial V^{myo}}{\partial t} = -I_{ion}^{myo} + \left[\frac{\nabla \cdot (\sigma_i^{myo} \nabla V^{myo}) + \nabla \cdot (\sigma_i^{myo} \nabla \varphi_e)}{\beta} \right] \\ - G_{gap} \cdot N^{fib} (V^{myo} - V^{fib}) + I_{stim} \end{aligned} \quad (2)$$

$$\begin{aligned} C^{fib} \frac{\partial V^{fib}}{\partial t} = -I_{ion}^{fib} + \left[\frac{\nabla \cdot (\sigma_i^{fib} \nabla V^{fib}) + \nabla \cdot (\sigma_i^{fib} \nabla \varphi_e)}{\beta} \right] \\ + G_{gap} (V^{myo} - V^{fib}) \end{aligned} \quad (3)$$

where I_{ion}^{myo} and I_{ion}^{fib} are the total ion channel current across the myocyte and fibroblast membranes, and C^{myo} and C^{fib} are the myocyte and fibroblast membrane capacitance. σ_i^{myo} and σ_i^{fib} are the intracellular myocyte and fibroblast conductivity, and σ_e denotes the extracellular conductivity. $N^{fib} \sim N(\mu, \sigma_N^2)$ is the local number of fibroblasts coupled to a single myocyte, and a stationary rotor is generated when $\mu = 0$. G_{gap} is the gap junction conductance of a coupled fibroblast-myocyte pair; I_{stim} is an external stimulation current, and β is the myocyte and fibroblast per volume ratio. The dynamic potential φ_e in a spatial domain of 25 mm $\times$ 25 mm with a single rotor is simulated and recorded in 4 episodes. We also generate 4 episodes for multiple rotors via reaction-diffusion equations as suggested in [25].

For training, we generate a series of 1500 snapshots or video frames with a resolution of 200 $\times$ 200 pixels for each episode (12000 in total) as illustrated in Fig. 1. Two additional episodes are generated for evaluation: one with a single rotor and the other with multiple rotors. A series of 600 snapshots are created individually from each episode and are excluded from the training process for testing. The true core of rotors are identified via the phase singularity (PS) approach [8], [9]. To evaluate the performance of image enhancement in our model, Gaussian noise with signal-to-noise ratio (SNR) of 5dB is added into the extracellular potentials φ_e of each video frame [26].

Further, we capture 4 more optical mapping recordings from different Langendorff-perfused rabbit hearts by voltage-sensitive dye and 12-bit CCD cameras, as described in [8]. For

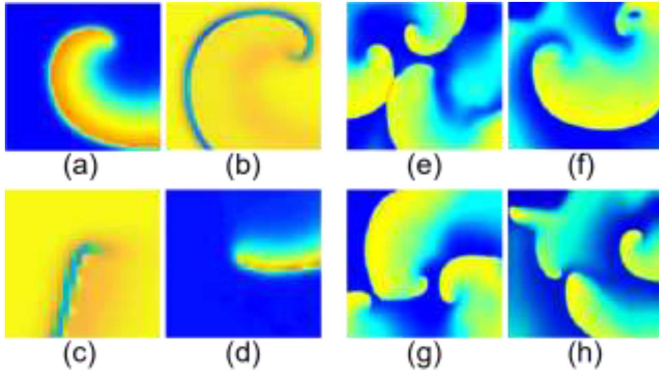


Fig. 1. (a) and (b) are stationary rotors; (c) and (d) are meandering rotors; (e)–(h) are multiple rotors.

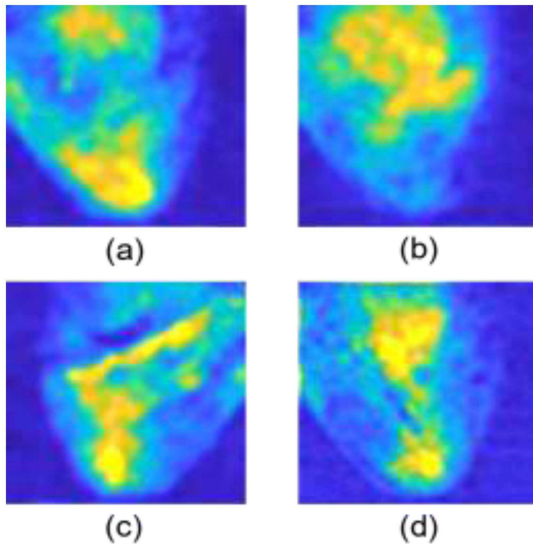


Fig. 2. Four separate optical mapping recordings obtained in different Langendorff-perfused rabbit hearts.

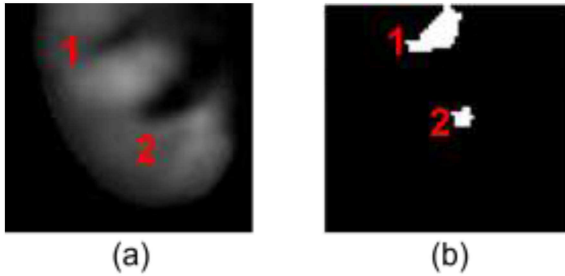


Fig. 3. (a) Animal optical mapping data with one meandering (“1”) and one stationary (“2”) rotors; (b) location of the rotor cores as time evolves for the animal optical mapping data: The meandering rotor and the stationary rotor.

each recording, we collect 1800 snapshots (7200 in total) as training datasets (Fig. 2).

The testing dataset comprises of 1500 snapshots derived from figure-of-8 optical mapping recording, which are excluded from the training dataset. A snapshot of the potential from the testing dataset is shown in Fig. 3(a). The meandering and stationary

rotor cores are indicated by “1” and “2”, respectively [8]. The clusters (the white dots) of both types of cores as time evolves are depicted in Fig. 3(b). Compared to the simulation data, the animal potential mapping is blurrier, obscuring the rotor core. To address the issue of data scarcity and validate if AI methods can be trained on simulation and then applied to experimental data, we use all the simulated images (more than 13000 video frames) to train the AI model and assess the detection performance using the randomly selected 10% of the animal optical mapping data (1500 frames). The result is compared with models trained by animal optical mapping data and the combination of simulation and animal optical mapping data.

B. Image Pre-Processing

We investigate 4 different techniques to pre-process the images: the average filter (AVF), the Gaussian filter (GF), the median filter (MF), and the EEMD. The pre-processed images are then fed into the YOLO model for rotor core detection. EEMD is inspired by the Empirical Mode Decomposition (EMD), which decomposes a time series signal into components with progressively lower frequencies, known as Intrinsic Mode Functions (IMFs). An IMF is a function with symmetric upper and lower envelopes, and the number of zero-crossings and the number of extremes is equal or different at most by one. Each IMF reflects the signal characteristics in a different frequency band, and they collectively reveal the underlying instantaneous time-frequency-energy characteristics of the signal [27]. EMD variants have been extended to spatial data, such as images [28], [29]. However, such variants are susceptible to noise contamination and mode mixing issues, as they rely on the data values at each time stamp to iteratively conduct the decomposition. Remarkably, the EEMD circumvents such issues with noise-assisted ensemble decomposition: each entity in the ensemble is the original image contaminated with white noise, and the mean of those ensemble trials constitute the IMFs [27], [28]. To alleviate computational cost, images are converted into one-dimensional vectors based on which the decomposition is then applied, as in most EMD algorithms. Specifically, an image X with size $l_1 \times l_2$ is first transformed to a $1 \times l_1 l_2$ vector $R(s)$, where $s \in [1, l_1 l_2]$ is the vector signal index, and then decomposed to derive the IMFs as below:

- 1) Add white noise $\omega_j(s)$ into $R(s)$ to generate a target vector for the j th trial:

$$q_j(s) = R(s) + \omega_j(s) \quad (4)$$

Here, the noise variance is chosen as 0.01 by trial and error so that the noise is not so large that it contaminates the texture of images yet large enough to generate a diverse ensemble.

- 2) The upper and lower envelopes, denoted as $U(s)$ and $L(s)$, of $q_j(s)$ are identified by interpolation between the local maxima and minima of $q_j(s)$, from which we get the overall trend:

$$T(s) = \frac{U(s) + L(s)}{2} \quad (5)$$

- 3) Subtract the trend from $q_j(s)$ to derive a new data $u_1(s)$ as

$$u_1(s) = q_j(s) - T(s) \quad (6)$$

- 4) Iterate the sifting operations in step 2 and 3 until $u_i(s)$ meets two conditions: over the entire range of s , the number of signal extrema and the number of zero-crossings must be equal or differ by one at most; and the mean value of the envelope defined by the local maxima and the local minima is zero at any index s . A sifting stop criterion Γ is included to avoid the issue of amplitude obliteration caused by the second condition [27],

$$\Gamma = \frac{\|u_{prev}(s) - u_{cur}(s)\|_2^2}{\|u_{prev}(s)\|_2^2} \quad (7)$$

where $u_{prev}(s)$ is the $u(s)$ value in previous iteration, $u_{cur}(s)$ is the value in current iteration, and the sifting process stops when $\Gamma < 0.2$ [27]. Then, we set $u_i(s)$ as $F_{1j}(s)$, the first IMF of the j th ensemble, which possesses the highest frequency components of the signal. The first residual $r_1(s) = q_j(s) - F_{1j}(s)$ becomes the new target vector.

- 5) Repeat step 4, the k th residual or the new target vector is given as $r_k(s) = r_{k-1}(s) - F_{kj}(s)$, where $q_j(s)$ in (4) is $r_0(s)$. This iteration continues until a certain criterion is reached, such as a maximum number of iterations or the frequency level of the residual term. Therefrom, $q_j(s)$ is decomposed into a sum of K IMFs and a residual term as

$$q_j(s) = \sum_{k=1}^K F_{kj}(s) + r_K(s) \quad (8)$$

where $F_{kj}(s)$ is the k th IMF of the j th ensemble trial, and $r_K(s)$ is the residual in the K th iteration.

- 6) With different white noise realizations to implement M ensemble trials, we obtain the ensemble mean of the k th IMF as

$$F_k = \frac{1}{M} \sum_{j=1}^M F_{kj} \quad (9)$$

The total number of ensemble trials is set as $M = 50$ in this study. The ensemble mean of the k th IMF F_k will then be transferred to 2D-image IMF G_k . The ensemble means of the corresponding IMFs portray finer details in the image, when detrended, have enhanced contrast and facilitate the image recognition [30]. Consequently, we select the best IMF G_k with the highest peak-signal-to-noise ratio (PSNR), which possesses the highest resemblance with the original image [31]. For the video frame X ,

$$PSNR(G_k, X) = 10 \log_{10} \left(\frac{255^2}{MSE(G_k, X)} \right) \quad (10)$$

where

$$MSE(G_k, X) = \frac{1}{l_1 l_2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} (G_{kij}, X_{ij})^2 \quad (11)$$

The AVF convolves a sliding window filter on the image that replaces the center value in the window with the average values of neighbor pixels. The MF replaces each pixel values by the median values of its neighbor pixels, sorted in an increasing order, in the window. The GF is achieved by convolving Gaussian sliding window with the image [32]. A 3×3 sliding window with stride of 1 is used for AVF, MF, and GF.

C. Spatial Resolution

Here, both simulation and optical potential mapping images have relatively low resolution, with 200×200 and 80×80 pixels, respectively. YOLO is exceptionally effective in detecting small objects or the rotor core in our case here. YOLO utilizes anchor boxes, which are predefined bounding boxes with specific aspect ratios and sizes based on the characteristics of the training data. During training, YOLO adjusts the coordinates of these anchor boxes to accurately localize objects in the image by fitting them to the ground truth bounding boxes. This differs from other object detection methods that employ fixed default boxes or sliding windows, which may fail to detect small objects. Additionally, YOLO demonstrates greater flexibility in handling objects with diverse shapes and sizes. It achieves this through a grid-based approach: the image is divided into numerous cells, and the presence and absence of the object in each cell will then be determined.

Further, YOLO predicts bounding boxes as offsets from the top-left corner of each cell in the grid, and the size of the bounding box can vary depending on the size and aspect ratio of the object being detected. During post-processing, the predicted bounding boxes are converted to coordinates relative to the entire image, and their sizes are adjusted based on the anchor boxes used in training. This allows YOLO to generate bounding boxes of various sizes and aspect ratios, even if they are smaller than the size of the grid cell. Moreover, rotors have been observed in various shapes (e.g., spiral, figure-eight, and multiform [1]) and sizes during rotor mapping procedures, depending on patient condition and location of the rotors in the heart. Indeed, the size of the bounding box represents the uncertainty in identification of rotor cores.

D. The YOLO Model

To detect the rotor core, we apply the YOLO-based object detection architecture with 24 convolutional layers (CLs) followed by 2 fully connected layers (FCLs), as demonstrated in Fig. 4. The first 24 CLs with max pooling layer are used to extract features from images, and the last 2 FCLs are added to predict the output probabilities and bounding boxes [33]. A linear activation function is used for the final layer, and the activation function of the leaky rectified linear unit (ReLU) are used for all other layers.

The YOLO architecture uses bounding boxes to detect objects in an partitioned image [33]. Specifically, the input image is first divided into $D \times D$ grid cells. If the center of rotor core falls into a grid cell, a bounding box (the red box in Fig. 5(a)) in that grid cell is then included and responsible for detecting that rotor core. The centroid and size of the bounding box indicate the core

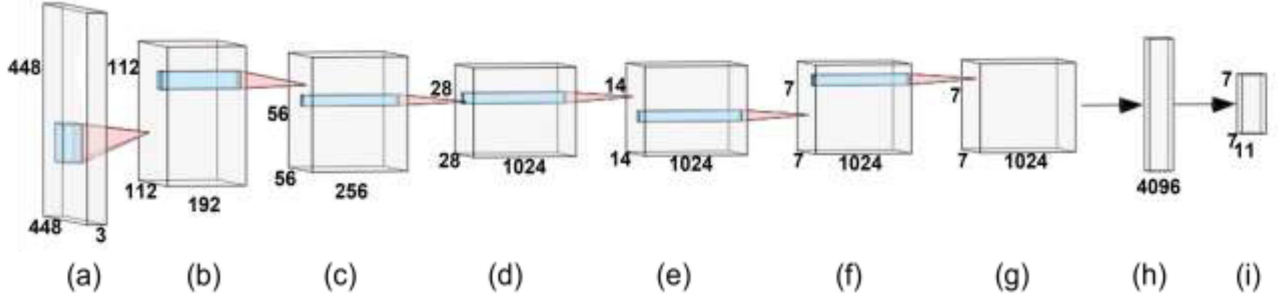


Fig. 4. CNN architecture has 24 convolutional layers followed by 2 fully connected layers. (a) The original image is resized to 448×448 . The default stride (how far the filter moves in every step along one direction) for each CL is 1; (b) the first layer includes 1 CL ($7 \times 7 \times 64$ with stride = 2) and 1 max pool (2×2 with stride = 2); (c) the second layer includes 1 CL ($1 \times 1 \times 128$ and 1 max pool (2×2 with stride = 2); (d) the third layer includes 4 CLs ($1 \times 1 \times 256$, $3 \times 3 \times 256$, $1 \times 1 \times 256$, and $1 \times 1 \times 512$) and 1 max pool (2×2 with stride = 2); (e) the fourth layer includes 10 CLs ($(1 \times 1 \times 256, 3 \times 3 \times 512) \times 4$, $1 \times 1 \times 512$, and $3 \times 3 \times 1024$) and 1 max pool (2×2 with stride = 2); (f) the fifth layer includes 6 CLs ($(1 \times 1 \times 512, 3 \times 3 \times 1024) \times 2$, $3 \times 3 \times 1024$, and $3 \times 3 \times 1024$ with stride = 2) and 1 max pool (2×2 , stride = 2); (g) the sixth layer includes 2 CLs ($(3 \times 3 \times 1024) \times 2$); (h) and (i) each include a FCL.

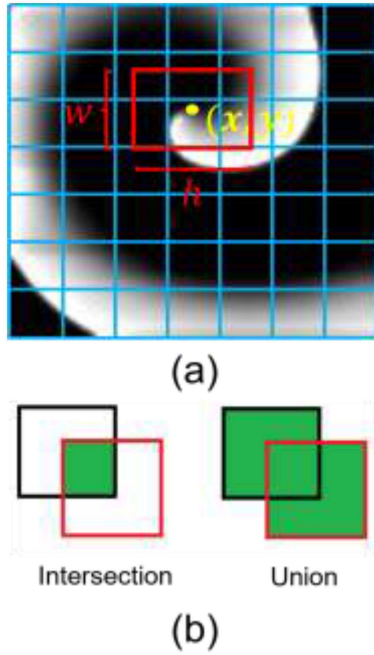


Fig. 5. (a) Image of a simulated rotor is divided into 7×7 grid cells; the centroid of the bounding box (the yellow dot) indicates the rotor core; (b) intersection and union between the predicted bounding box (red) and the ground truth (black).

location and its potential range. Each bounding box is associated with 5 variables: x , y , w , h , and C . (x , y) are the centroid coordinates of the bounding box. The valuation of (x , y) is relative to the grid cell that the core may reside in. By default, coordinates of the upper-left and the lower-right corners of each grid cell are (0, 0) and (1, 1), respectively. Thus, $x, y \in [0, 1]$. The width w and height h of the bounding box determine the prediction precision; a small bounding box implies high precision. The precision can be further quantified by the IOU ($IOU \in [0, 1]$), which hinges on intersection and union between the predicted bounding box and the ground truth (Fig. 5(b)). Intersection is the overlapping area between the predicted and ground truth

bounding boxes, and union is the total area covered by these two bounding boxes. Multiple bounding boxes (quantity denoted by B) can be adopted for each grid cell in case there are multiple rotor cores in one cell. For the multiple rotor cores in simulated datasets, $B = 4$ is used to stipulate that each grid cell can contain up to 4 cores. $B = 2$ is used for the animal optical mapping dataset, as the stationary rotor core and the meandering one may co-exist in one grid cell. $C = I_{core} \times IOU$ is the confidence for rotor core detection. $I_{core} = 0$ if no rotor exists in a cell, which also implies that $C = 0$. Otherwise, $I_{core} = 1$. Therefore, the output of the last FCL contains information on x, y, w, h, C for B bounding boxes and the corresponding probability of rotor core existence in all grid cells ($D = 7$). Thus, the dimension of the last FCL is $7 \times 7 \times (B \times 5 + 1)$.

The optimal valuation of x, y, w, h and C is derived from minimization of a loss function, as depicted in (12), which essentially corrects the centroid and size of the bounding box [33]:

$$\begin{aligned}
 Loss(x, y, w, h, C) &= \lambda_1 \sum_{i=0}^{D^2} \sum_{j=0}^B \mu_{ij}^{obj} \left[(x_i - x_{i0})^2 + (y_i - y_{i0})^2 \right] \\
 &+ \lambda_1 \sum_{i=0}^{D^2} \sum_{j=0}^B \mu_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{w_{i0}})^2 + (\sqrt{h_i} - \sqrt{h_{i0}})^2 \right] \\
 &+ \lambda_1 \sum_{i=0}^{D^2} \sum_{j=0}^B \mu_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{w_{i0}})^2 + (\sqrt{h_i} - \sqrt{h_{i0}})^2 \right] \\
 &+ \sum_{i=0}^{D^2} \sum_{j=0}^B \mu_{ij}^{obj} (C_i - C_{i0})^2 + \lambda_2 \sum_{i=0}^{S^2} \sum_{j=0}^B \mu_{ij}^{noobj} (C_i - C_{i0})^2 \\
 &+ \sum_{i=0}^{D^2} \mu_i^{obj} (p_{core}^i - p_{core}^{i0})^2
 \end{aligned} \quad (12)$$

The first term of the loss function arises from the prediction of the centroid (x, y). (x_0, y_0) are the true coordinate of the

centroid given in the training data. $\mu_{ij}^{obj} = 1$ if the j^{th} bounding box in the i^{th} cell contains the rotor core, and $\mu_{ij}^{obj} = 0$ otherwise. The true valuation of width w_0 and height h_0 are given by the PS approach [9], which points to the smallest bounding box that fully covers the rotor core. The third term is the loss related to the confidence C for each bounding box predictor. $0 < C_0 \leq 1$ if the rotor core exists in the j^{th} bounding box of the i^{th} cell, otherwise $C_0 = 0$. $\mu_{ij}^{noobj} = 1$ when there is no rotor core in the j^{th} bounding box in the i^{th} cell. The last term is the classification loss. $\mu_i^{obj} = 1$ if the rotor core resides in the i^{th} cell, and $\mu_i^{obj} = 0$ otherwise. $\lambda_1 = 5$ and $\lambda_2 = 0.5$ are the weights used for coordinate predictions and for confidence predictions when no rotor is present, respectively. p_{core} is the probability that a rotor core exists in one grid cell.

Here, we evaluate the accuracy of deep learning methods, YOLO, AVF-YOLO, GF-YOLO, MF-YOLO, and EEMD-YOLO, using both simulation and the animal optical mapping datasets to tackle the rotor core detection problem. The IOU is used as the criterion for rotor core detection, and the accuracies are compared using precision, recall, and F1 score. In addition, Tukey's honestly significant difference (Tukey's HSD) test are utilized to evaluate and compare the performance of models trained with simulation data, animal optical mapping data, and a combination of both.

III. NUMERICAL RESULTS

A. Rotors Detection on Simulated Data

Fig. 6 shows the rotor core detection using AI models on simulated dataset with single rotor. In EEMD-YOLO, the original simulated data is first decomposed by EEMD. The 2D-image IMF G_6 with the highest average PSNR = 7.91 dB is selected for rotor detection. Detection results using YOLO, AVF-YOLO, GF-YOLO, and MF-YOLO are shown in Fig. 6(b)–(e). EEMD-YOLO removes Gaussian noise while preserving fine details of the images (Fig. 6(f)). Fig. 6(g)–(l) shows detected locations of predicted rotor cores (blue) and ground truth (white) using AI models on the randomly selected snapshots over the testing dataset. It is observed that all AI models can correctly detect the location of the majority of rotor cores where ground truth are located.

Fig. 7 shows the rotor cores detection using AI models on a randomly selected video frame of simulation dataset with multiple rotors. The 2D-image IMF G_6 with the highest average PSNR = 7.87 dB is selected for rotor detection. Detection results using YOLO, AVF-YOLO, GF-YOLO, and MF-YOLO are shown in Fig. 7(b)–(e). EEMD-YOLO removes Gaussian noise while preserving fine details of the images (Fig. 7(f)). Fig. 7(g)–(l) shows detected locations of predicted rotor cores (blue) and ground truth (white) using AI models on the randomly selected snapshots over the testing dataset, indicating that rotor cores are mostly predicted in locations where ground truth are located.

As shown in Table I, we compare the performance of different models on simulated data. It appears that the EEMD-YOLO model achieves the highest IOU (0.95), precision (97.2%),

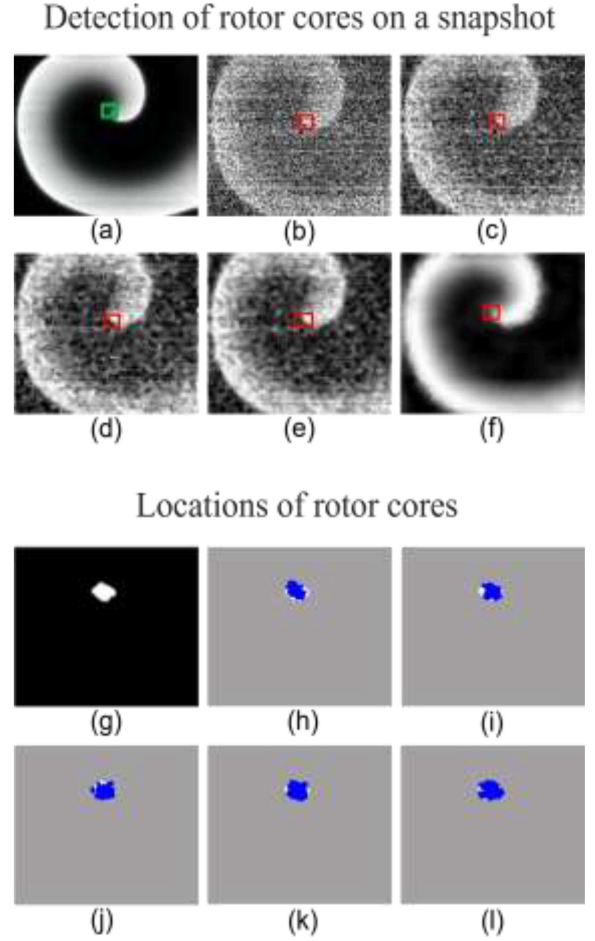


Fig. 6. Detection results on pre-processed simulated single rotor using different methods: (a) The ground truth; (b) simulated rotor core with SNR = 5dB noise, core identified via YOLO; (c)–(f) detection results using AVF-YOLO, GF-YOLO, MF-YOLO, and EEMD-YOLO; The blue clusters of dots are locations of stationary rotor cores from 300 testing snapshots of the simulated single-rotor: (g) The ground truth, and predicted locations using (h) YOLO, (i) AVF-YOLO, (j) GF-YOLO, (k) MF-YOLO, and (l) EEMD-YOLO models.

TABLE I
COMPARISON OF IOU, PRECISION, RECALL, AND F1 SCORE FOR DIFFERENT METHODS ON SIMULATED DATA

Models	IOU (SD)	Precision	Recall	F1 score
YOLO	0.82 (0.188)	89.3%	82.7%	85.9%
MF-YOLO	0.91 (0.108)	96.2%	91.3%	93.7%
AVF-YOLO	0.92 (0.129)	95.2%	85.4%	90.0%
GF-YOLO	0.93 (0.124)	96.9%	87.8%	92.1%
EEMD-YOLO	0.95 (0.072)	97.2%	93.8%	95.5%

recall (93.8%), and F1 score (95.5%) among the models listed. This suggests that EEMD-YOLO performs the best to accurately localize objects and achieve a good balance between precision and recall. The YOLO model performs relatively worse. Other models, MF-YOLO, AVF-YOLO, and GF-YOLO, demonstrate varying levels of performance, better than YOLO.

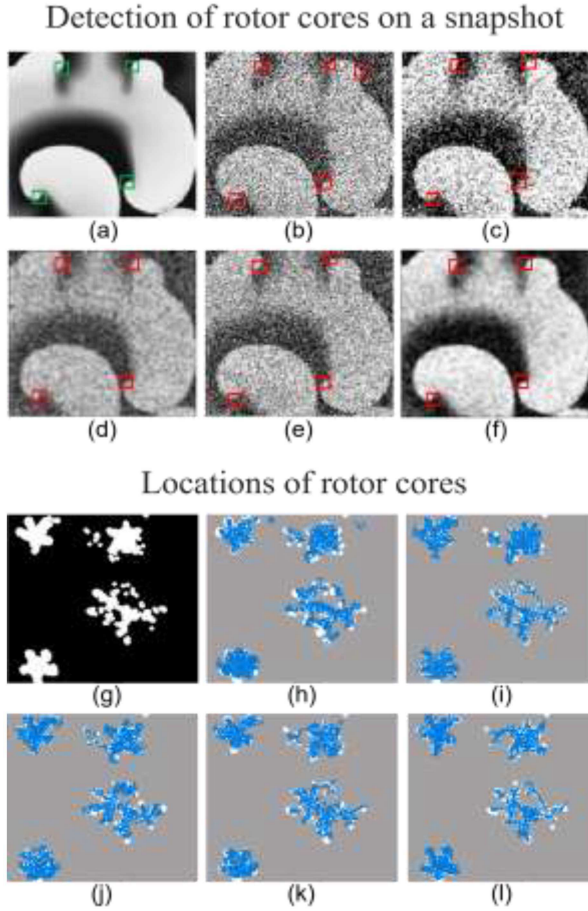


Fig. 7. Detection results on pre-processed snapshots using simulated data with multiple rotors: (a) The ground truth; (b) simulated multiple rotor cores with SNR = 5 dB noise, and cores identified via YOLO; (c)–(f) detection results using AVF-YOLO, GF-YOLO, MF-YOLO, and EEMD-YOLO; the blue clusters of dots are locations of meandering rotor cores from 300 testing snapshots of the simulated multiple-rotors data: (g) The ground truth, and predicted locations using (h) YOLO, (i) AVF-YOLO, (j) GF-YOLO, (k) MF-YOLO, and (l) EEMD-YOLO models.

B. Rotors Detection on Animal Optical Mapping

Fig. 8 depicts the rotor core detection using AI models on animal optical mapping dataset: the upper rotor is meandering, and the lower one is stationary. Fig. 8(a) is the ground truth. The AVF-YOLO, GF-YOLO, and MF-YOLO (Fig. 8(c)–(e)) slightly removes more shadow area compared with the original optical mapping image (Fig. 8(b)). In Fig. 8(f), for the EEMD-YOLO, the 2D-image IMF G_6 with an average PSNR = 6.76 dB is selected. The selected IMF from EEMD has enhanced contrast compared with other 3 AI models with filters. Fig. 8(g)–(l) shows detected locations of the predicted rotor cores (blue) and ground truth (white) using AI models on the randomly selected snapshots over the testing dataset. It can be observed that all AI models can correctly detect locations of the majority of rotor cores. However, closer inspection of Fig. 8 shows that all models miss part of the ground truth location of rotor cores.

The comparison of IOU and the SD for different models in three different groups (A, B, and C) using Tukey's HSD is

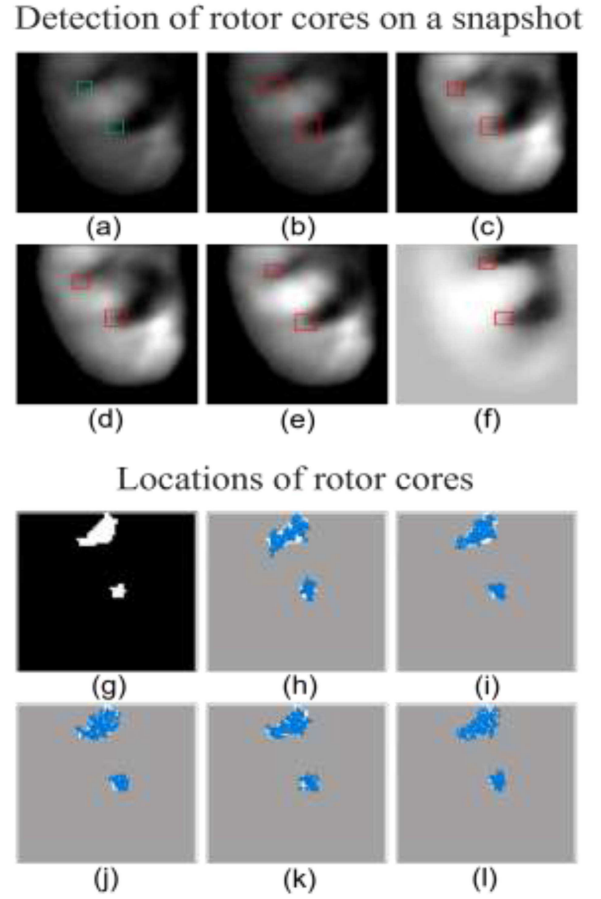


Fig. 8. Detection results on pre-processed animal optical mapping data using different methods: (a) The ground truth; (b)–(f) rotor detection results with YOLO, AVF-YOLO, GF-YOLO, MF-YOLO, and EEMD-YOLO. The blue clusters of dots are locations of rotor cores from 300 testing snapshots of the animal optical mapping: (g) The ground truth, and predicted locations using (h) YOLO, (i) AVF-YOLO, (j) GF-YOLO, (k) MF-YOLO, and (l) EEMD-YOLO models.

depicted in Table II. The models include YOLO, MF-YOLO, AVF-YOLO, GF-YOLO, and EEMD-YOLO. In group A, YOLO achieves an IOU of 0.65, while MF-YOLO and EEMD-YOLO has a higher IOU of 0.75. AVF-YOLO and GF-YOLO achieves IOU values of 0.70 and 0.68, respectively. In group B and group C, all models achieve high performance (p value > 0.33), which are higher compared to Group A (p value < 0.001). Overall, the EEMD-YOLO model registers the highest accuracy in rotor core detection in all scenarios.

As shown in Table III, all models achieve high precision, recall, and F1 scores on the animal optical mapping data. The highest precision scores are achieved by EEMD-YOLO in all three training groups (group A: 87.1%, group B: 97.5%, group C: 97.2%). In general, it can be observed that models trained with both simulation and animal optical mapping data (group B) achieve better performance than models trained with only simulation data (group A) or animal optical mapping data (group C). This is evident from the fact that all models achieve higher scores in group B than group A or C.

TABLE II

COMPARISON OF IOU USING TUKEY'S HSD FOR DIFFERENT METHODS ON ANIMAL OPTICAL MAPPING DATA WITH DIFFERENT TRAINING DATA

Models	IOU (SD)			P value (A to B)	P value (A to C)	P value (B to C)
	Group A	Group B	Group C			
YOLO	0.65 (0.112)	0.81 (0.046)	0.83(0.054)			
MF-YOLO	0.75 (0.099)	0.91 (0.052)	0.89(0.069)			
AVF-YOLO	0.70 (0.101)	0.92 (0.037)	0.90(0.031)	All<0.001	All<0.001	All>0.33
GF-YOLO	0.68 (0.098)	0.91 (0.035)	0.91(0.044)			
EEMD-YOLO	0.75 (0.076)	0.94 (0.028)	0.93(0.034)			

Group A: simulation data, group B: both simulation and animal optical mapping data, group C: animal optical mapping data.

TABLE III

COMPARISON OF PRECISION, RECALL, AND F1 SCORE USING DIFFERENT METHODS ON ANIMAL OPTICAL MAPPING DATA WITH DIFFERENT TRAINING DATA

Methods	Precision %			Recall %			F1 score %		
	A	B	C	A	B	C	A	B	C
YOLO	83.8	94.3	94.1	65.0	84.1	84.3	73.2	88.9	88.9
MF-YOLO	86.8	97.3	96.6	71.6	91.7	91.2	78.5	94.4	93.8
AVF-YOLO	85.8	96.2	96.3	70.8	85.9	85.8	77.6	90.7	90.7
GF-YOLO	84.2	97.3	97.1	70.1	89.2	89.3	76.1	93.0	93.0
EEMD-YOLO	87.1	97.5	96.8	71.4	94.2	92.2	78.4	95.8	94.4

Group A: simulation data, group B: both simulation and animal optical mapping data, group C: animal optical mapping data.

Comparing different models, EEMD-YOLO appears to be the most effective model in terms of overall performance, achieving high precision, recall, and F1 scores across all three training groups. AVF-YOLO also achieves good results in terms of precision and F1 score but had lower recall scores compared to other models. Furthermore, our images are of relatively low resolution, and YOLO is fairly fast and agile. The training time is 1166.3 seconds for the simulation data (12000 frames from 8 simulation episodes), 980.3 seconds for optical mapping data (7200 frames randomly selected from optical mapping data) and 1853.8 seconds for a combination of simulation and optical mapping data as training dataset.

IV. DISCUSSION

Addressing the scarcity of literature in AF identification using AI-assisted methods, this study presents the development and validation of an image-decomposition-enhanced CNN algorithm (EEMD-YOLO) for rotor core detection. It preserves the fine details of the original image frames, yielding the highest accuracy among different approaches for detecting both the stationary and meandering rotors. We use both simulated electric potential mapping data and optical mapping data acquired from animal experiments to train and test the models. The model learned only from the simulated patterns can identify rotors in animal optical mapping data (see Table III). In comparison, performance of models trained by the combination of both simulation and animal optical mapping data, or only animal optical mapping data register higher accuracy.

In clinical settings, the data recording is susceptible to noise contamination or artifacts. This can be attributed to the impact arising from electrode placement sites and contact areas in the high-resolution electrocardiography system. In the average filter, pixel values are replaced by average values of neighboring pixels, which, however, may distort the image texture. The median filter replaces each pixel values by the median values of neighboring pixels and protects the edge of the image well while filtering out the noise. It is less sensitive than linear techniques (e.g., AVF) to extreme changes in pixel values, yet it could lead to image discontinuity. GF works by assigning different weights to pixels at different neighboring locations with a kernel, to preserve the overall grayscale distribution of the image. The EEMD algorithm instead decomposes electric potential mapping into components with distinct frequency and energy amplitudes, the ensemble technique guarantees a robust decomposition of images. This method capitalizes the time-scale characteristics of the data itself for decomposition without predefined basis functions as required by Wavelet analysis. Thus, the EEMD method can be applied to any type of signal decomposition in theory, and has an obvious advantage in dealing with non-stationary and non-linear data [27].

One limitation of this study is that we only investigate the potential mapping on 2D surfaces, while the dynamic evolution of the potential mapping on the 3D surfaces across epicardial and endocardial regions could potentially lead to more accurate diagnosis and realistic applications [34]. As a direction for future work, cardiac imaging techniques have offered potential

opportunities to visualize and monitor the heart's electrophysiological state at high spatial and temporal resolutions for real-time and remote detection of cardiac ailments. Despite the potential benefits, cardiac imaging techniques have inherent challenges, such as the need for precise measurements using ECGI and the possible complications associated with basket catheter insertion [35], [36]. Nevertheless, exploring the effectiveness of AI-based approaches in combination with cardiac imaging techniques for real-time diagnosis and early recognition of cardiac illness in clinic settings could hold promise for the future.

V. CONCLUSION

In this study, we investigate AF rotor detection via integrating deep learning with image decomposition techniques. The novel approach is comparable with the common gold standard of physics-based methods using data from both simulation and animal optical mapping experiments. It offers a more automated, less labor- and expertise-intensive rotor detection process with promises of application in real-time dynamic cardiac monitoring and AF diagnosis.

REFERENCES

- [1] S. S. Chugh et al., "Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study," *Circulation*, vol. 129, no. 8, pp. 837–847, Feb. 2014.
- [2] A. Shamsan et al., "Intrinsic recurrence quantification analysis of nonlinear and nonstationary short-term time series," *Chaos: Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 9, Sep. 2020, Art. no. 093104.
- [3] X. Wu et al., "Pattern recognition and automatic identification of early-stage atrial fibrillation," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113560.
- [4] S. V. Pandit and J. Jalife, "Rotors and the dynamics of cardiac fibrillation," *Circulation Res.*, vol. 112, no. 5, pp. 849–862, Mar. 2013.
- [5] S. M. Narayan et al., "Treatment of atrial fibrillation by the ablation of localized sources," *J. Amer. College Cardiol.*, vol. 60, no. 7, pp. 628–636, Aug. 2012.
- [6] J. Jalife, O. Berenfeld, and M. Mansour, "Mother rotors and fibrillatory conduction: A mechanism of atrial fibrillation," *Cardiovasc. Res.*, vol. 54, no. 2, pp. 204–216, May 2002.
- [7] P. Sanders et al., "Spectral analysis identifies sites of high-frequency activity maintaining atrial fibrillation in humans," *Circulation*, vol. 112, no. 6, pp. 789–797, Aug. 2005.
- [8] E. M. Annoni et al., "Novel quantitative analytical approaches for rotor identification and associated implications for mapping," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 2, pp. 273–281, Feb. 2018.
- [9] V. Ravikumar et al., "Novel mapping techniques for rotor core detection using simulated intracardiac electrograms," *J. Cardiovasc. Electrophysiol.*, vol. 32, no. 5, pp. 1268–1280, 2021.
- [10] J. Ng and J. J. Goldberger, "Understanding and interpreting dominant frequency analysis of AF electrograms," *J. Cardiovasc. Electrophysiol.*, vol. 18, no. 6, pp. 680–685, Jun. 2007.
- [11] R. Miotto et al., "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.
- [12] M. K. Mulimani, J. K. Alageshan, and R. Pandit, "Deep-learning-assisted detection and termination of spiral and broken-spiral waves in mathematical models for cardiac tissue," *Phys. Rev. Res.*, vol. 2, no. 2, May 2020, Art. no. 023155.
- [13] M. I. Alhusseini et al., "Machine learning to classify intracardiac electrical patterns during atrial fibrillation," *Circulation: Arrhythmia Electrophysiol.*, vol. 13, no. 8, Aug. 2020, Art. no. e008160.
- [14] H. Lilienkamp and T. Lilienkamp, "Detecting spiral wave tips using deep learning," *Sci. Rep.*, vol. 11, no. 1, Oct. 2021, Art. no. 19767.
- [15] J. Lebert et al., "Rotor localization and phase mapping of cardiac excitation waves using deep neural networks," *Front. Physiol.*, vol. 12, Dec. 2021, Art. no. 782176.
- [16] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [17] Z.-Q. Zhao et al., "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
- [18] B. Yao, R. Zhu, and H. Yang, "Characterizing the location and extent of myocardial infarctions with inverse ECG modeling and spatiotemporal regularization," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1445–1455, Sep. 2018.
- [19] D. Bhonsle, V. Chandra, and P. G. Sinha, "Medical image denoising using bilateral filter," *Int. J. Image, Graph. Signal Process.*, vol. 4, pp. 36–43, Jul. 2012.
- [20] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.
- [21] S. K. Berkaya et al., "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, May 2018.
- [22] X. Wang et al., "ECG compression based on combining of EMD and wavelet transform," *Electron. Lett.*, vol. 52, no. 19, pp. 1588–1590, 2016.
- [23] M. Elhoseny and K. Shankar, "Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements," *Measurement*, vol. 143, pp. 125–135, Sep. 2019.
- [24] Y. Rudy and W. L. Quan, "A model study of the effects of the discrete cellular structure on electrical propagation in cardiac tissue," *Circulation Res.*, vol. 61, no. 6, pp. 815–823, Dec. 1987.
- [25] M. Hwang et al., "Electrophysiological rotor ablation in in-silico modeling of atrial fibrillation: Comparisons with dominant frequency, Shannon entropy, and phase singularity," *PLoS One*, vol. 11, no. 2, Feb. 2016, Art. no. e0149695.
- [26] M. Kania, M. Fereniec, and R. Maniewski, "Wavelet denoising for multi-lead high resolution ECG signals," *Meas. Sci. Rev.*, vol. 7, no. 4, pp. 30–33, 2007.
- [27] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [28] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 135–156, Apr. 2010.
- [29] A. Shamsan, W. Dan, and C. Cheng, "Multimodal data fusion using multivariate empirical mode decomposition for automatic process monitoring," *IEEE Sensors Lett.*, vol. 3, no. 1, Jan. 2019, Art. no. 7101704.
- [30] O. Aworunse et al., "Image decomposition accelerates dynamic network modeling for in situ monitoring of bio-mimic wing printing processes," *Procedia Manuf.*, vol. 39, pp. 178–184, 2019.
- [31] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.
- [32] B. K. S. Kumar, "Image denoising based on Gaussian/bilateral filter and its method noise thresholding," *Signal Image Video Process.*, vol. 7, no. 6, pp. 1159–1172, Nov. 2013.
- [33] J. Redmon et al., "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [34] L. Xu et al., "3D multifunctional integumentary membranes for spatiotemporal cardiac measurements and stimulation across the entire epicardium," *Nature Commun.*, vol. 5, no. 1, Feb. 2014, Art. no. 3329.
- [35] Y. Shu, W. Dan, and C. Cheng, "Spatiotemporal regularization in effective reconstruction of epicardial potential," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2021, pp. 1–4.
- [36] M. Cluitmans et al., "Validation and opportunities of electrocardiographic imaging: From technical achievements to clinical applications," *Front. Physiol.*, vol. 9, 2018, Art. no. 1305.