ECGVEDNET: A Variational Encoder-Decoder Network for ECG Delineation in Morphology Variant ECGs

Long Chen, Zheheng Jiang, Joseph Barker, Huiyu Zhou, Fernando Schlindwein, Will Nicolson, G. Andre Ng, and Xin Li

***Abstract*—Electrocardiogram (ECG) delineation to identify the fiducial points of ECG segments, plays an important role in cardiovascular diagnosis and care. Whilst deep delineation frameworks have been deployed within the literature, several factors still hinder their development: (a) data availability: the capacity of deep learning models to generalise is limited by the amount of available data; (b) morphology variations: ECG complexes vary, even within the same person, which degrades the performance of conventional deep learning models. To address these concerns, we present a large-scale 12-leads ECG dataset, ICDIRS, to train and evaluate a novel deep delineation model-ECGVEDNET. ICDIRS is a large-scale ECG dataset with 156,145 QRS onset annotations and 156,145 T peak annotations. ECGVEDNET is a novel variational encoder-decoder network designed to address morphology variations. In ECGVEDNET, we construct a well-regularized latent space, in which the latent features of ECG follow a regular distribution and present smaller morphology variations than in the raw data space. Finally, a transfer learning framework is proposed to transfer the knowledge learned on ICDIRS to smaller datasets. On ICDIRS, ECGVEDNET achieves accuracy of 86.28%/88.31% within 5/10 ms tolerance for QRS onset and accuracy of 89.94%/91.16% within 5/10 ms tolerance for T peak. On QTDB, the average time errors computed for QRS onset and T peak are -1.86 ± 8.02 ms and -0.50 ± 12.96 ms, respectively, achieving state-of-the-art performances on both large and small-scale datasets. We will release the source code and the pre-trained model on ICDIRS once accepted.**

***Index Terms*—ECG delineation, fiducial points detection, variational encoder-decoder, transfer learning.**

Manuscript received 16 October 2023; revised 12 January 2024; accepted 31 January 2024. Date of publication 12 March 2024; date of current version 20 June 2024. The work of Xin Li and G. Andre Ng was supported by the British Heart Foundation under Grant AA/18/3/34220. The work of Xin Li, Fernando Schlindwein, and G. Andre Ng was supported by Medical Research Council U.K. under MRC DPFS Ref. MR/S037306/1 and in part by the British Heart Foundation under BHF Project PG/18/33/33780. The work of G. Andre Ng was supported by the British Heart Foundation under BHF Programme RG/17/3/32774, and in part by SJM/Abbott, Medtronic, and Biosense Webster. This work was carried out at the National Institute for Health and Care Research (NIHR) Leicester Biomedical Research Centre (BRC). For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) license to any Accepted Manuscript version arising. *(Corresponding author: Xin Li.)*

Long Chen is with the Department of Medical Physics and Biomedical Engineering, University College London, U.K.

Zheheng Jiang and Huiyu Zhou are with the School of Computing and Mathematical Sciences, University of Leicester, U.K.

Joseph Barker is with the Faculty of Medicine, Imperial College London, U.K.

Fernando Schlindwein is with the School of Engineering, University of Leicester, U.K.

Will Nicolson and G. Andre Ng are with the Department of Cardiovascular Sciences, University of Leicester, U.K.

Xin Li is with the School of Engineering, University of Leicester, Leicester LE1 7RH, U.K. (e-mail: xin.li@leicester.ac.uk).

Digital Object Identifier 10.1109/TBME.2024.3363077

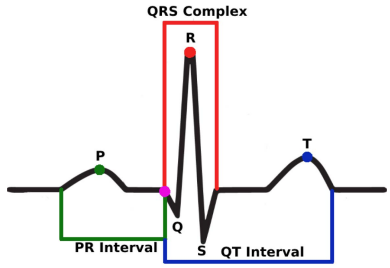

Fig. 1. ECG signal with fiducial points. The pink point is the QRS onset, and the blue point is the T peak.

I. Introduction

THE electrocardiogram (ECG) is a non-invasive, cost-efficient tool for heart activity monitoring [1] and has become a standard tool for the diagnosis of cardiac diseases, which are the leading causes of death worldwide [2]. ECG delineation [3], which aims to mark the onset, the offset and the peak of each ECG waveform (P wave, QRS complex, and T wave) as shown in Fig. 1, helps to improve the diagnosis of cardiac diseases.

ECG delineation plays an important role in clinical diagnosis as most of the clinically useful information in the ECG is found within the intervals and amplitudes determined by its fiducial points [4]. Manually marking these fiducial points is time-consuming and requires expert knowledge, therefore, automatic, digital signal processing-based ECG delineation methods [5], [6] have been adopted to assist the cardiologist with health care decision making. Digital signal processing-based ECG delineation methods [7], [8] usually start with determining wave types. Then, the peaks and boundaries, i.e., the wave onset and offset, are detected. Most of the ECG delineation algorithms start from the R-peak detection since they are visually notable in the ECG wave and easier to be detected [9]. Once the R-peaks have been detected, they serve as the reference to delineate the QRS

complex, P waves and T waves [10]. Based on ECG delineation, further important temporal parameters (e.g. RR intervals and QT intervals) and features (e.g. amplitudes and slopes) can be derived. Digital signal processing-based algorithms [11], [12] have been reported to achieve good performances on some simple and specific datasets. However, their performances highly depend on the setting of predefined thresholds, which requires considerable experience.

Several works [13] have demonstrated that deep learning methods achieve excellent delineation performance. The deep learning model [14] is composed of multiple processing layers to learn discriminative feature representations from large data. Its powerful feature learning capability enables it achieve great successes in various tasks, such as ECG classification [15] and ECG delineation [16]. Developing accurate and robust deep delineation models is, however, still challenging due to several factors: (a) large-scale ECG delineation datasets with high-quality annotations are insufficient and the performance of deep model is highly influenced by the scale of the datasets; (b) the ECG morphology across persons or even in the same person presents large variations, degrading the performances of deep delineation models.

To address these challenges, we first collect a large-scale ECG dataset, ICDIRS, and manually label the QRS onsets and T wave peaks to facilitate the development of deep learning techniques in ECG delineation. Second, we present a novel variational encoder-decoder named ECGVEDNET for ECG delineation, the biggest advantage of ECGVEDNET is its ability to handle the large morphology variations in ECG. In ECGVEDNET, we construct a novel well-regularized latent space (RLS), which can transfer large variations in raw ECG data space into small variations in the latent space. RLS greatly improves the model's robustness to morphology variations. Although the proposed model achieves prominent performance on ICDIRS, it fails to achieve satisfactory performance on smaller datasets due to the insufficiency of training data. Hence, we propose a transfer learning framework to break its limitation on smaller datasets. The transfer learning framework exploits the prior knowledge learned on ICDIRS to help the model training on small-scale datasets. This strategy builds the new state-of-the-art work on QTDB.

The main contributions of this work can be summarized as follow:

- We collect a large-scale 12-leads ECG dataset, ICDIRS, with 156,145 manually labelled QRS onset and T peak annotations. This dataset provides a useful training and evaluation environment to develop and select excellent ECG delineation frameworks.
- We propose a novel variational encoder-decoder network, ECGVEDNET, to address the large morphology variations in ECG. ECGVEDNET constructs a novel well-regularized latent space to handle the morphology variations, greatly improving the model's robustness to morphology variations.
- We present a transfer learning framework to overcome the insufficient data issue on small-scale datasets. The transfer learning framework improves deep models' generalisation capacity on small-scale datasets.
- We validate the proposed ECGVEDNET on both large-scale dataset ICDIRS and small-scale dataset QTDB. The experimental results show our ECGVEDNET achieves state-of-the-art delineation performance on both datasets.

The paper is structured as follows. Section II summarises the related works. Section III describes the proposed ECGVEDNET and the transfer learning framework. Section IV describes the experimental set-up and Section V reports and discusses the experimental results. Finally, Section VI presents the concluding remarks.

II. Related Work

Automatic ECG delineation can be broadly divided into digital signal processing based methods, traditional machine learning based methods, and deep learning based methods.

A. Digital Signal Processing-Based Delineation

Many digital signal processing-based delineation methods have been proposed within the literature, such as Wavelet Transform [7], [17], Bayesian algorithm [18], adaptive filtering [19], Kalman filter [20], and Phasor Transform (PT) [21]. Akhbari et al. [22] proposed a switching Kalman filter (SKF) model for ECG delineation, where ECG wave-forms are modeled with Gaussian functions and ECG baselines are modeled with first order auto regressive models. SKF model achieves small mean error and root mean square error, but it is sensitive to the initial parameters which need to be manually set. Wavelet Transform (WT) has been used by a number of groups to achieve high delineation accuracy. Martinez et al. [7] employed WT to delineate the P wave, QRS complex, and T wave for single-lead ECG while Rincon et al. [23] employed a multi-lead WT algorithm to delineate ECGs on a wireless body sensor network. Unfortunately, WT-based methods are notable computational expensive since they require intensive mathematical operations, and they also require considerable experiences to manually set some thresholds. In the original ECG recordings, precise detection of the onsets and offsets is a challenging task because the signal amplitudes around these fiducial points are notably low and the noise level can be even higher than the signal. To address this challenge, some enhancement techniques have been employed to enhance the ECG waves. For example, Martinez et al. [9] applied Phasor Transform (PT) to enhance the ECG waves by enlarging the amplitude of P and T waves, thus notably easing the delineation task.

Digital signal processing based methods [24], [25] are capable of removing undesired components or enhancing the desired components. However, they usually require predefined thresholds or parameters, which have great influences on the delineation performance.

B. Traditional Machine Learning-Based Delineation

The traditional machine learning based delineation methods include artificial neural network (ANN), Genetic algorithm, Bayesian model, and K-Nearest Neighbour (KNN) algorithm. Xue et al. [26] were the first to adopt ANN for QRS detection. They developed an ANN adaptive whitening filter to model

the lower frequencies of ECG, this method is very effective in removing the nonlinear noises in ECG signals. The Genetic algorithms have also been adopted for QRS detection at a early stage. Poli et al. [27] proposed a Genetic algorithm for QRS detection that can produce the optimal solution through continuous selection and evolution, however, the algorithm suffers from the local optimum problem. Lin et al. [18] proposed a Bayesian model to exploit the strong local dependency of ECG signals for accurate ECG delineation. The model achieves high accuracy on P-wave and T-wave detection but at a high computational cost. Saini et al. [28] proposed a KNN based ECG delineation method to detect the fiducial points along with the waveform boundaries. The KNN algorithm is efficient and simple but very sensitive to the local structure of the ECG signals.

The traditional machine learning methods become progressively better at performing specific tasks. However, they still require many human interventions such as hand-crafted feature design and hyper-parameter selection.

C. Deep Learning-Based Delineation

Due to the powerful feature representation capacity of deep learning, many convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have been employed for ECG delineation [29]. Camps et al. [13] proposed a two-step CNN-based framework for QRS delineation. This framework first segments the QRS waves from the ECG signals and further detects the QRS onset and offset in the QRS waves. Abrishami et al. [16] employed a bidirectional long short-term memory (BiLSTM) network to segment ECGs, this model achieves better accuracy than the traditional models.

To extract better feature representations, a few works try to combine the CNN module with the LSTM module. Londhe et al. [30] proposed a hybrid deep network that consists of a CNN module and two bidirectional LSTM modules for ECG segmentation. The hybrid network extracts better temporal features, outperforming the LSTM network and bidirectional LSTM network. Similarly, Nurmaini et al. [31] incorporated three CNN modules and one BiLSTM module into a deep delineation network, while Peimankar et al. [32] combined three CNN modules and two BiLSTM modules into a delineation network. Extensive experimental results show that the combined models achieve better performance than the models with only CNN or LSTM.

In ECG delineation, encoder-decoder is the most used deep architecture. Jimenez-Perez et al. [33] proposed an encoder-decoder with U-Net backbone for automatic ECG delineation, while Liang et al. [34] exploited an encoder-decoder architecture with BiLSTM backbone for ECG delineation. Wang et al. [35] further incorporated the domain knowledge and individual feature knowledge into the encoder-decoder architecture that further improves the delineation accuracy. Deep models show significant improvements in various supervision tasks where sufficient training data are available. Despite the encouraging performances, there is still much space for improvement in the deep architecture design for ECG delineation.

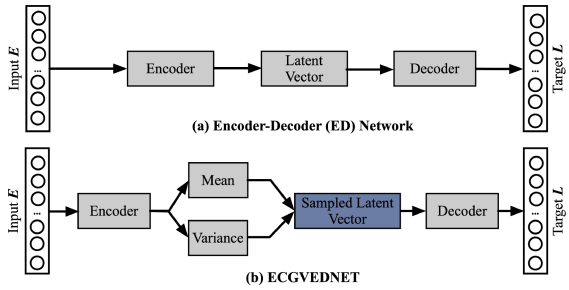

Fig. 2. Comparisons of ECGVEDNET and Encoder-Decoder (ED) Network. ECGVEDNET encodes the input into a regularized distribution (mean + variance) rather than a single point in the latent space.

III. Proposed ECGVEDNET

A. The Motivation of the Proposed ECGVEDNET

Previous works [33], [35] commonly employed the encoder-decoder (ED) networks for ECG delineation. However, the latent space of ED is extremely irregular and causes the over-fitting problem. For example, close points in the latent space can be decoded into largely different data, some points in the latent space can be decoded into meaningless data. To address the irregularity problem in latent spaces, variational autoencoder (VAE) [36], [37], one of the most famous unsupervised learning models, encodes the input into a distribution (mean + variance) rather than a single point, VAE also adds a regularization term in the loss function to ensure a better organization of the latent space.

Motivated by the unsupervised learning VAE model, we propose the supervised learning model ECGVEDNET for ECG delineation. Similarly to VAE, ECGVEDNET maps the inputs to a well-regularized distribution in the latent space. We employ the multivariate Gaussian distribution with a diagonal covariance matrix to regularize the latent feature space. Then, a regularized latent feature vector is sampled from the latent distribution and forwarded to the decoder to generate the target output. The biggest advantage of the proposed ECGVEDNET over the ED is that the large variations in raw ECG data space can be minimized and regularized in the latent space. Hence, the proposed ECGVEDNET is less sensitive to large morphology variations and well-handles the ECG data with large morphology variations. The comparisons of the ED network and the proposed ECGVEDNET can be observed in Fig. 2.

B. The Regularized Latent Space (RLS)

As shown in Fig. 3, the proposed ECGVEDNET consists of three components: the probabilistic encoder, the regularized latent space and the probabilistic decoder. In ECGVEDNET, the probabilistic encoder first maps the input $\mathbf{e}$ into a regularized distribution $q_\phi(\mathbf{z}|\mathbf{e})$ in the latent space, also named $variational\ posterior$. Then, a latent vector $\mathbf{z}$ is sampled from the regularized distribution, i.e, $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{e})$, as the latent feature representation of the input $\mathbf{e}$. Finally, the probabilistic decoder maps the sampled vector $\mathbf{z}$ into the target distribution $p_\theta(\mathbf{l}|\mathbf{z})$, where $\mathbf{l}$ is the output, i.e., the predicted label of the input $\mathbf{e}$. ϕ and θ are the parameters of the encoder and decoder, respectively.

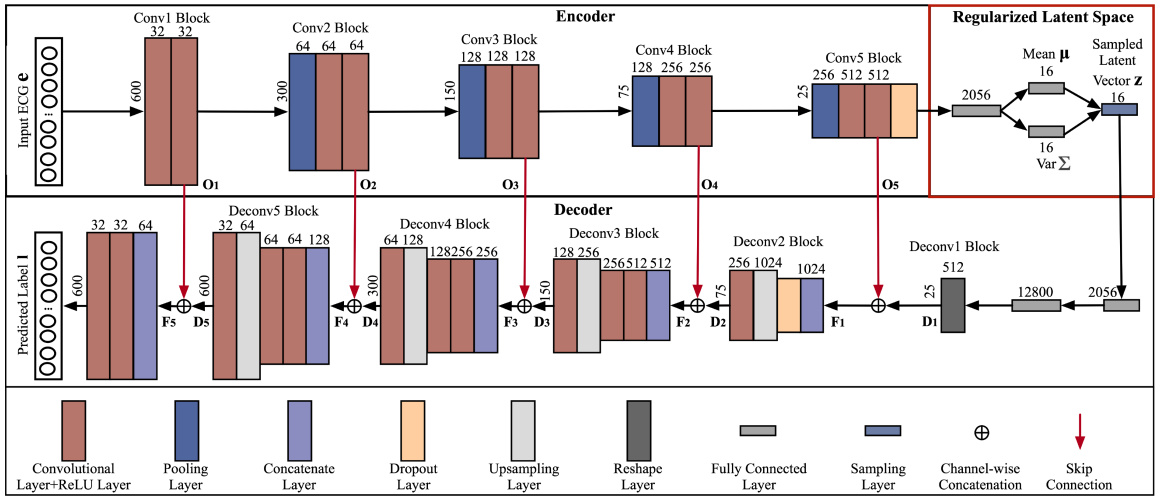

Fig. 3. Network structure of our proposed ECGVEDNET. The encoder encodes the input into a regularized multivariate Gaussian distribution (mean + variance) in the regularized latent space, in which the morphology of different ECGs present smaller variations than in the raw data space, alleviating the influence of the morphology variations on the feature learning of deep models.

The input $\mathbf{e}$, i.e., the raw ECG signal, presents large morphology variations that hinder the performance of delineation. To address this issue, in the regularized latent space, we translate the unconstrained inputs into a constrained latent distribution so that the morphology variations across different inputs can be minimized and follow a regularized distribution. Specifically, we construct the regularized latent space using the multivariate Gaussian distribution with a diagonal covariance matrix in the form of $q_\phi(\mathbf{z}|\mathbf{e}) = \mathcal{N}(\mathbf{z}|\mu_\phi(\mathbf{e}), \Sigma_\phi(\mathbf{e}))$.

$$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{e}) = \mathcal{N}(\mu_\phi(\mathbf{e}), \Sigma_\phi(\mathbf{e})) \tag{1}$$

$$= \frac{\exp(-\frac{1}{2}(x-\mu_\phi(\mathbf{e}))^T \Sigma_\phi(\mathbf{e})^{-1}(x-\mu_\phi(\mathbf{e})))}{|2\pi\Sigma_\phi(\mathbf{e})|} \tag{2}$$

Here, $\mu_\phi(\mathbf{e}) \in \mathbb{R}^{1\times k}$ is the mean of the multivariate Gaussian distribution, and $\Sigma_\phi(\mathbf{e}) \in \mathbb{R}^{1\times k}$ is the diagonal covariance matrix in the following form:

$$\Sigma_\phi(\mathbf{e}) = \begin{bmatrix} \sigma^2_{\phi 1}(\mathbf{e}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\phi k}(\mathbf{e}) \end{bmatrix} \tag{3}$$

k denotes the dimension of the feature vector z, which is sampled from the multivariate Gaussian distribution, and $\sigma^2_{\phi k}$ denotes the variance of the multivariate Gaussian distribution in the k-th dimension. In the multivariate Gaussian distribution, the mean $\mu_\phi(\mathbf{e})$ and variance $\Sigma_\phi(\mathbf{e})$ are constructed by the trainable deep neural network parameterized by the weights ϕ. This means they can be trained and learned from the training data. The latent space enables the encoded features of different raw ECGs follow a constrained multivariate Gaussian distribution and present smaller morphology variations, which greatly decrease the learning difficulty of the deep models.

C. The Network Structure of the Encoder

Denote the ECG delineation dataset as $\mathbf{T} = \{(\mathbf{e}_n, \mathbf{g}_n), n = 1, 2, \ldots, N\}$, where $\mathbf{e}_n$ denotes the input ECG segment, and $\mathbf{g}_n$ denotes the ground truth label for $\mathbf{e}_n$. $\mathbf{e}_n, \mathbf{g}_n \in \mathbb{R}^{N_w \times N_c}$, N_w and N_c indicates the number of samples in each spatial dimension ($N_w = 600$ and $N_c = 1$ in our ICDIRS dataset). As shown in Fig. 3, the encoder contains 5 convolutional blocks. For each ECG segment $\mathbf{e}_n$, we feed it into the encoder, and extract hierarchical features at different convolutional blocks, including $\mathbf{Conv1} \sim \mathbf{Conv5}$ blocks.

Denote $\mathbf{I}_m, \mathbf{O}_m \in \mathcal{R}^{W\times C}$ as the input and output feature maps of the last convolutional layer for the m-th convolution block ($m \in \{1, \ldots, 5\}$), where W denotes the width of the feature map and C denotes the channel number of the feature map. Let $\omega \in \mathcal{R}^{K\times C}$ denotes a 1D convolution kernel with C channels, and K denotes the kernel size. Each feature channel $\mathbf{O}^{\bar{c}}_{m,\,p} \in \mathcal{R}^{W\times 1}$ ($\bar{c} = \{1, 2, \ldots, C\}$) in $\mathbf{O}_m$ can be computed as:

$$\mathbf{O}^{\bar{c}}_{m,\,p} = \mathcal{F}(\mathbf{I}_m, \omega) = \sum_{c=1}^{C} \sum_{i\in\Omega_K} \omega^c_{i+\frac{K-1}{2}} \mathbf{I}^c_{m,\,i+p} \tag{4}$$

where $\mathcal{F}$ denotes the convolution function in deep learning, p represents the location coordinate in one channel of the input feature map, and Ω defines the convolution area in each convolution operation.

$$\Omega_K = \left\{(i) : i = \left\{-\frac{K-1}{2}, \ldots, \frac{K-1}{2}\right\}\right\} \tag{5}$$

The output feature maps $\mathbf{O}_m$ in each convolution block of the encoder will be used as the input features of the decoder, because we exploit the skip connections to enhance the high layer feature maps and alleviate the vanishing gradients problem.

D. The Network Architecture of the Decoder

The decoder maps the latent feature representation ($\mathbf{z}$) of the input ($\mathbf{e}$) into the target label space using the distribution

$p_\theta(\mathbf{l}|\mathbf{z})$, $\mathbf{l}$ denotes the predicted label for the input $\mathbf{e}$. Specifically, the decoder consists of five up-sampling blocks, including one reshape block and four deconvolution blocks. In the decoder, we first map the latent feature representation $\mathbf{z}$ into a 2D feature map using two fully connected layers and one reshape layer. Then, we construct four deconvolution blocks to up-sample the resolution of the feature maps. For convenience, we define the reshape layer as $\mathbf{Deconv_1}$, and the following four deconvolution blocks as $\mathbf{Deconv_2} \sim \mathbf{Deconv_5}$. The output of the i-th deconvolution block denotes as $\mathbf{D}_i$.

Considering the high layer feature maps lacks low-level detailed features, we adopt the skip connections to combine features from the encoder and decoder to form more potent feature representations. We use channel-wise concatenation to fuse features maps from two sources, and the fused feature maps $\mathbf{F}$ are formulated as

$$F_i = O_{6-i} \oplus D_i, i = \{1, \ldots, 5\} \tag{6}$$

where $\oplus$ denotes the channel-wise concatenation operation, $\mathbf{O}$ and $\mathbf{D}$ are the feature maps from the encoder and decoder, respectively (refer to Fig. 3 for better understanding).

E. The Loss Function

The loss function of our proposed ECGVEDNET consists of two terms, including the reconstruction term and the regularization term. We can write this as

$$L = L_{recon} + \beta L_{KL}(\mathbf{z}, \mathcal{N}(\mu, \Sigma)) \tag{7}$$

The reconstruction term is a binary cross-entropy loss, and can be formulated as:

$$L_{recon} = -\frac{1}{N_w}\sum_{i=1}^{N_w} \mathbf{g}_i log \mathbf{l}_i + (1-\mathbf{g}_i)\log(1-\mathbf{l}_i) \tag{8}$$

where $\mathbf{g}$ and $\mathbf{l}$ are the ground truth label and predicted label, respectively. N_w indicates the number of samples in each ECG segment. The regularization term is a KL Divergence loss, which can be formulated as

$$L_{KL}(\mathbf{z}, \mathcal{N}(\mu, \Sigma)) = \frac{1}{2}\sum_{i=1}^{k}(\sigma_i - \log\,\sigma_i - 1 + \mu_i^2) \tag{9}$$

The KL loss is commonly used to measure of the distance between two distributions, here, it computes the relative entropy between the distribution z and the multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. $\mathbf{z}$, μ and Σ are the sampled latent vector, the mean and the variance of the multivariate Gaussian distribution, respectively.

F. Transfer Learning Framework Between ICDIRS and QTDB

The performance of deep learning networks highly relies on the large-scale datasets. However, most of previous works train the deep delineation networks on small datasets as lacking of large-scale labelled ECG delineation datasets. This limits the deep models' generalisation capability. In this work, we introduce the transfer learning strategy to transfer the knowledge learned by the ECGVEDNET on the large-scale dataset ICDIRS to the ECGVEDNET on the small dataset QTDB. In many scenarios, transfer learning boosts the model performance than training with only a small amount of data. To improve the performance of ECGVEDNET on smaller datasets, with the assistance of the large-scale dataset ICDIRS, we adopt transfer learning to train the ECGVEDNET on the small dataset QTDB. Specifically, we first pre-train our ECGVEDNET on ICDIRS, then fine-tuning it on QTDB.

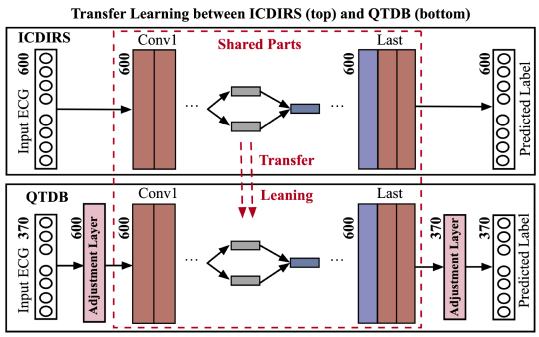

Fig. 4. Transfer learning framework between ICDIRS and QTDB.

However, the proposed ECGVEDNET has a disadvantage that it requires fixed input size and cannot handle varied input sizes. In the datasets preparation stage, we split the long ECG recordings into short ECG segments with fixed length (600 samples on ICDIRS and 370 samples on QTDB). We adopt different input sizes on two datasets mainly because two datasets adopt different sampling frequencies. Sine the input size of ECGVEDNET on QTDB is different from the size on ICDIRS, we cannot adopt the same network architecture on two datasets. To address the size mismatch issue, we add two adjustment layers in ECGVEDNET as shown in Fig. 4, one is added before the $Conv_1$ block and another one is added at the end. These two adjustment layers are convolutional layers, the front adjustment layer accepts data of 370 samples and transfers the input into the output of 600 samples, the back adjustment layer transfers the data size from 600 samples to 370 samples. We use the weights learnt on ICDIRS to initialize the shared parts of two ECGVEDNETs.

IV. Experimental Setup

To demonstrate the effectiveness of the proposed ECGVEDNET, we first collect a large-scale ECG delineation dataset, called ICDIRS, and manually label the QRS onset and T peak points. Then, we conduct comprehensive evaluations on the ICDIRS dataset and the public QTDB dataset [38], [39]. In this section, we first introduce the experimental datasets and evaluation metrics. Then, we describe the implementation details.

A. Datasets

ICDIRS. We collect a larger-scale 12-lead ECG dataset named ICDIRS from 54 patients. The ICDIRS dataset was collected in the University Hospitals of Leicester National Health Service Trust, U.K.. All the 12-leads (I, II, III, aVR, aVL, aVF, V1,

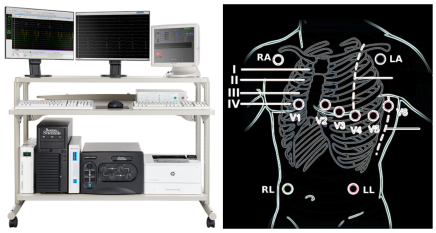

Fig. 5. EP recording system and 12 lead ECG collection sites.

TABLE I
COMPARISONS OF OUR ICDIRS DATASET AND THE QTDB DATASET

Datasets	Patients	Leads	QRS onset	T peak	Frequency
QTDB	105	2	3,623	3,542	250 Hz
ICDIRS	54	12	156,145	156,145	1000 Hz

V2, V3, V4, V5, V6) ECG data are recorded at 1 kHz sampling rate by the same electrocardiograph device LABSYSTE PRO EP Recording System as shown in Fig. 5. We carefully split the 12-leads ECG recordings into ECG segments with fixed length (600 samples) because our ECGVEDNE requires fixed input size. For each QRS onset, we first locate its adjacent T peak on the right, then we use random sliding window to cut off one ECG segment containing the QRS onset and the T peak. The length of 600 can capture only one QRS onset and one T peak. The complete dataset contains 156,145 ECG segments (93,687,000 samples) in total. Finally, we split the dataset into train and test sets with a ratio 4:1 at patient level. In total, the train set contains 124,916 (80%) ECG segments and the test set contains 31,229 (20%) segments. The QRS-onset and T-peak points are manually annotated by doctors in Glenfield Hospital. Ethical approval was granted by the Derbyshire Research Ethics Committee (09/H0401/70) and the study protocol was approved by the Research and Development Office of the University Hospitals of Leicester National Health Service Trust (Leicester, U.K.). All patients gave written, informed consent.

QTDB [38], [39]. QTDB is the most commonly used public dataset for evaluating different delineation algorithms. It contains 2-leads ECG recordings collected from 105 patients. To fairly compare different methods, we split 63 (60%) recordings as the training set, 21 (20%) recordings as the validation set and 21 (20%) recordings as the test set, we train and test all the deep comparison methods using the same data split. The comparisons of QTDB and ICDIRS can be observed in Table I.

We carefully split the 2-leads ECG recordings into ECG segments with fixed length of 370 samples. The ECG segment with 370 samples is sufficiently large to capture one complete heartbeat [40], [41]. We prepare two datasets from QTDB, one for QRS onset and another one for T peak. For each QRS onset, we first locate its adjacent T peak on the right. If the adjacent T peak does not exist within 370 samples, we use random sliding window to cut off one ECG segment containing the QRS onset. If the adjacent T peak exists within 370 samples, we use random sliding window to cut off one ECG segment containing the QRS onset and the T peak. In some cases, the ECG segment may contain two QRS onsets, we use zeros to fill the area around the second QRS onset. We prepare the T peak dataset using the same splitting strategy.

B. ECG Delineation Evaluation Metrics

Mean and standard deviation errors. We adopt the commonly used mean and standard deviation ($m \pm sd$) of the time errors as the delineation evaluation metrics on QTDB. The mean of the time error is calculated as the average time differences between predicted position and ground truth position.

Clinical trails metrics on ICDIRS. The ICDIRS dataset is the part of one UKRI MRC research project which is titled "Development of a successful novel technology for sudden cardiac death risk stratification for clinical use-LifeMap". We devote to apply the proposed framework in clinical trails, hence, on ICDIRS, we also adopt the detection accuracy $Accu_{tr}$ under the accepted tolerances (5 ms and 10 ms) of clinicians as the clinical trial metrics, which can be formulated as

$$Accu_{tr} = \frac{\sum_{i=1}^{N} \mathbb{I}(P_i, GT_i)}{N}, tr = 5\,ms\ or\ 10\,ms \tag{10}$$

where

$$\mathbb{I}(P_i, GT_i) = \begin{cases} 1 & \textbf{if } |P_i - GT_i| < tr, \\ 0 & \textbf{otherwise} \end{cases} \tag{11}$$

N represents the number of testing ECG segments, and tr represents the tolerance, tr =5 ms or 10 ms. The $\mathbb{I}(.)$ function indicates a true positive detection if the time error between predicted position P_i and ground truth position GT_i is lower than 5 or 10 ms. It means a successful detection is achieved when the error between the predicted position and ground truth position is within the tolerance. This evaluation metric helps to achieve satisfactory detection objectives in agreement with the accepted tolerances for clinicians.

C. Implementation Details

All the experiments are conducted on a server with an Intel Xeon CPU @ 2.40 GHz and 1 Nvidia Tesla P100 GPUs. ECGVEDNET is trained with Adam optimization algorithm, weight decay of 0.0001 and momentum of 0.9 are used. We implement the source code using the deep learning platform Keras https://keras.io/getting_started/, the source code and dataset will be public available after acceptation. Moreover, we have conducted extensive experiments to select the best hyper-parameters, including the training epoch, batch size, and learning rate. The details can be found in the Subsection V-A Ablation Experiments.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present and discuss the experimental results and findings. First, we conduct ablation experiments to investigate the influence of different components, including the skip connection, network depth, network hyper-parameters and the transfer learning strategy, on ECGVEDNET. Then, we conduct comparison experiments to compare our ECGVEDNET with several advanced delineation frameworks on ICDIRS and QTDB.

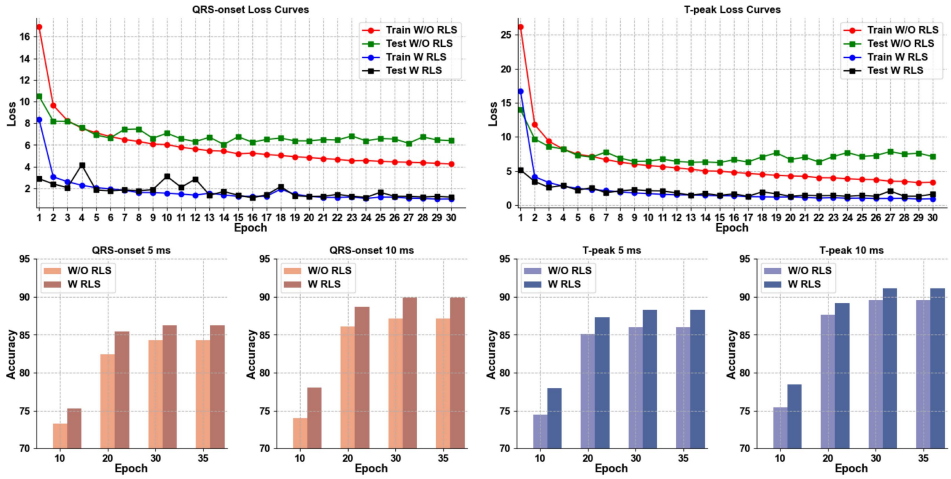

Fig. 6. Loss curves and accuracy of ECGVEDNETs with (W) and without (W/O) regularized latent space (RLS).

A. Ablation Experiments

In this subsection, we conduct ablation experiments on ICDIRS to investigate the influence of each component of ECGVEDNET on the final performance, including the regularized latent space, the skip connections, the network depth, the network hyper-parameters and the transfer learning strategy.

1) Ablation Studies of the Regularized Latent Space (RLS): We first conduct ablation experiments to investigate the influence of RLS on ECGVEDNET. Specifically, we design two types of latent space here, one is the RLS and another one is the common latent space used in most of encoder-decoder structures. The structure comparison of RLS and the common latent space can be found in Fig. 2.

Fig. 6 presents the performance comparisons of ECGVEDNETs with and without RLS, we observe that ECGVEDNETs with RLS converge to a lower loss smoothly and achieve better delineation performance than the ECGVEDNETs without RLS during the complete training stage. This is because the large morphology variations in raw ECGs lead to model overfitting, where the model captures noise in the training data and performs poorly on new, unseen data. The model with the common latent space is likely to overfit on the training data with large variations and becomes very sensitive to the noise. Differently, in ECGVEDNET, RLS translates the unconstrained inputs into a constrained latent distribution so that the large morphology variations across different inputs can be minimized and follow a regularized multivariate Gaussian distribution. In this way, feature representations being robust to morphology variations can be extracted from the latent feature space. The regularized feature representations facilitate the training of the deep models and improve model's generalisation capability on datasets with large morphology variations.

2) Ablation Studies of the Skip Connections and Network Depth: Numerous works show that network depth has the great influence on the performance of deep models, and more convolution layers usually brings higher performance. Hence, we design four network backbones with different depths to investigate the influence of network depth on ECGVEDNET. Specifically, we design network backbones with 3, 4, 5, and 6 convolution blocks in the encoders, and the decoder has corresponding 3, 4, 5, and 6 deconvolution blocks, respectively. However, with the increase of network depth, we found the gradient vanishing problem has become the main hindrance to the accuracy increase. Hence, we introduce skip connections to avoid the performance degradation and investigate the influences of the skip connections on ECGVEDNET.

Fig. 7 presents the performance comparisons of ECGVEDNETs with different depth, skip connection and RLS settings on ICDIRS, from which we observe that the performance of ECGVEDNET without skip connections increases with the depth increase from 3 blocks setting to 4 blocks setting. This is because more convolution layers introduce more non-linearity into the deep network, boosting the feature representation ability. However, a large performance drop has been observed when the network depth increases from 4 to 5 blocks. This is mainly because the increasing depths deteriorate the vanishing gradient problem. The deep network is trained using the gradient-based optimization algorithm, the front layers cannot receive sufficient gradients when more convolution layers are used. Hence, we add the skip connections to address the vanishing gradient problem, as shown in Fig. 7, after the introduction of skip connections, the 5 blocks setting achieves much better performance than 4 blocks setting. However, even though the skip connections can alleviate the vanishing gradient problem, very deep networks are still very hard to be trained. The delineation performance presents a drop trend when the number of blocks in the encoder is greater than 5 blocks.

3) Ablation Study of the Network Parameters: ECGVEDNET is trained with Adam optimization algorithm, the learning rate (LR), batch size and training epochs are three most important hyper-parameters for training deep networks. Hence, we conduct extensive experiments to select the best hyper-parameters empirically. Tables II presents the influence of the hyper-parameters, including the training epoch, batch size, and learning rate training epochs, on the final delineation performance. In the experiments, we find smaller learning rate 0.0001 enables our ECGVEDNET to converge slowly but

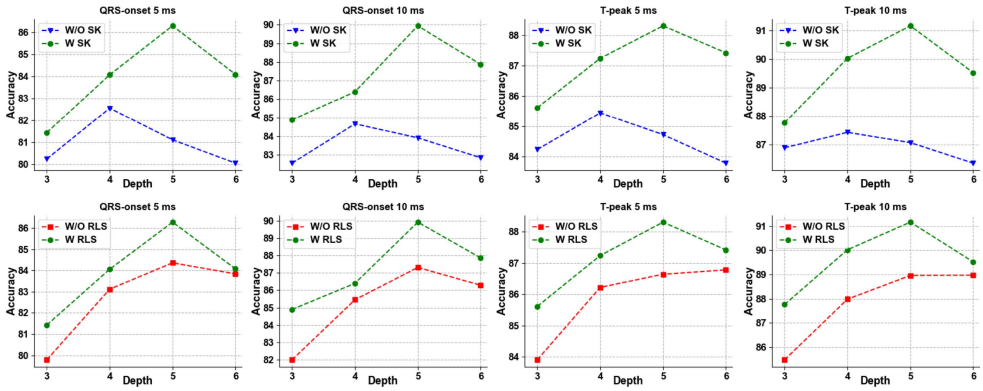

Fig. 7. Performance of ECGVEDNETs with different depth settings on ICDIRS. 'W' and 'W/O' indicate with and without respectively. 'SK' indicates the skip connections and 'RLS' indicates the regularized latent space.

TABLE II
PERFORMANCE OF ECGVEDNET AT DIFFERENT EPOCHS, BATCH SIZE (BATCH), AND LEARNING RATE (LR) ON ICDIRS

Model	Epoch	Batch	LR	5 ms	10ms	Model	Epoch	Batch	LR	5 ms	10ms
QRS onset	20	128	0.001	77.92	79.19	T peak	20	128	0.001	79.53	82.80
	20	128	0.0001	83.44	86.89		20	128	0.0001	85.68	87.12
	20	256	0.001	80.47	83.51		20	256	0.001	82.86	85.31
	20	256	0.0001	85.41	88.69		20	256	0.0001	87.34	89.21
	20	512	0.001	78.42	80.65		20	512	0.001	80.37	83.06
	20	512	0.0001	84.22	86.47		20	512	0.0001	86.25	88.97
	30	128	0.001	79.24	81.88		30	128	0.001	83.06	85.12
	30	128	0.0001	84.90	86.22		30	128	0.0001	86.84	88.36
	30	256	0.001	81.80	84.10		30	256	0.001	83.58	86.72
	30	256	0.0001	**86.28**	**89.94**		30	256	0.0001	**88.31**	**91.16**
	30	512	0.001	82.55	85.06		30	512	0.001	84.24	87.09
	30	512	0.0001	85.63	86.94		30	512	0.0001	87.66	89.52

The bold values indicates the best performed models.

stably to a better optimum, and produces a better performance. Our deep models converge faster if we set a larger learning rate of 0.001, however, the deep models with big learning rate are hard to train as the gradient descent starts to diverge.

We also observe that 30 training epochs are sufficient to train a good deep model, when the number of training epochs is more than necessary, the model performances turn to decease as the deep models start to over-fit on the training data and lose generalisation capacity. In our experiments, a small batch size 128 enables the deep models converge faster while a large batch size 512 make the deep model converge slowly. In practice, the batch size setting chosen is 256, which fits our experimental datasets and the memory requirements of our GPU hardware.

4) Ablation Study of the Transfer Learning Strategy: We adopt different transfer learning strategies to train the ECGVEDNET on QTDB, which helps us to investigate how the model pre-trained on the large ICDIRS dataset influences the model performance on the smaller dataset. It is worth noting that the ECGVEDNET architecture on QTDB is different from the one on ICDIRS. On QTDB, we add two new adjustment layers in the front and back of ECGVEDNET, the new ECGVEDNET on QTDB consists of the three parts: the front adjustment layer, the shared intermediate layers, and the back adjustment layer. Different transfer learning strategies are designed to train the deep models, in all experiments, we use Xavier to initialize two adjustment layers and use the pre-trained model to initialize the shared intermediate layers. During training, we selectively freeze certain parts of the deep models. As shown in Table III, the 1st model has not been trained on QTDB. The 2nd model freezes the front adjustment layer and the intermediate layer during training while the 3 rd model only freezes the intermediate layer, the 4th model fine-tunes all three parts. We also design the 5th model, which is trained from scratch on QTDB without transfer learning, as the reference.

From the experimental results in Table III, we have two important observations: 1) the retraining of the last adjustment layer is of vital importance for the model performance. The 2nd model achieves much better performance than the 1st model since its last adjustment layer has been trained, this is because retraining the last adjustment layer, i.e., the final prediction layer, bridges the domain gaps exist in different datasets; 2) transfer learning is a better strategy than training from scratch. The 3 rd model performs much better than the 5th model since it adopts transfer learning to initialize its intermediate layers. Transfer learning provides a good initialization for the model to be fine-tuned and reuses the features learned on large-scale datasets.

B. Comparison Experiments

In this subsection, we compare our proposed network with several classical and top performing delineation methods, including two signal processing based methods [7], [42] and two advanced deep learning based methods [13], [33]. For the deep learning based methods, we carefully train all the networks on

TABLE III

DELINEATION ERROR (MEAN ± STANDARD DEVIATION) OF ECGVEDNETS WITH DIFFERENT TRANSFER LEARNING STRATEGIES ON QTDB. "FREEZE": KEEP THE WEIGHTS UNCHANGED; "TRANSFER": USE THE MODEL PRE-TRAINED ON ICDIRS TO INITIALIZE THE WEIGHTS; "RETRAINING": TRAIN FROM SCRATCH

Models	Front Adjustment Layer	Intermediate Layers	Back Adjustment Layer	QRS onset (m ± sd)	T peak (m ± sd)
ECGVEDNET	Freeze	Transfer	Freeze	-5.60 ± 20.69	-2.45 ± 21.82
ECGVEDNET	Freeze	Transfer	Retraining	-2.10 ± 9.50	-0.66 ± 14.83
ECGVEDNET	Retraining	Transfer	Retraining	-2.02 ± 9.96	-0.65 ± 14.50
ECGVEDNET	Retraining	Fine-tuning	Retraining	**-1.86 ± 8.02**	**-0.50 ± 12.96**
ECGVEDNET	Retraining	Retraining	Retraining	-3.02 ± 12.83	-1.26 ± 18.66

The bold values indicates the best performed models.

TABLE IV

QUANTITATIVE PERFORMANCE COMPARISONS OF DIFFERENT DELINEATION METHODS ON ICDIRS

Tolerances	Methods	Types	Techniques	QRS onset	T peak	Average Error
5 ms	Martinez et al. [7]	Singnal Processing	Wavelet Transform	75.23	78.18	23.30
	Pilia et al. [42]	Singnal Processing	Wavelet Transform	79.88	81.66	19.23
	Camps [13]	Deep Learning	ConvNet	82.33	84.58	16.55
	Guillermo [33]	Deep Learning	Encoder-Decoder	82.43	83.98	16.80
	Ours	Deep Learning	Variational Encoder-Decoder	**86.28**	**89.94**	**11.89**
10 ms	Martinez et al. [7]	Singnal processing	Wavelet Transform	77.55	78.90	21.78
	Pilia et al. [42]	Singnal processing	Wavelet Transform	82.20	83.98	16.91
	Camps [13]	Deep learning	ConvNet	84.56	87.26	14.09
	Guillermo [33]	Deep learning	Encoder-Decoder	85.10	87.68	13.61
	Ours	Deep learning	Variational Encoder-Decoder	**88.31**	**91.16**	**10.27**

The bold values indicates the best performed models.

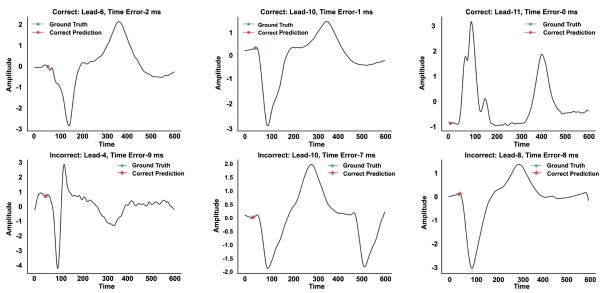

Fig. 8. Visualization of correct (Time Error $\leq$ 5 ms, top row) and incorrect (Time Error $>$ 5 ms, bottom row) QRS onset predictions of our ECGVEDNET.

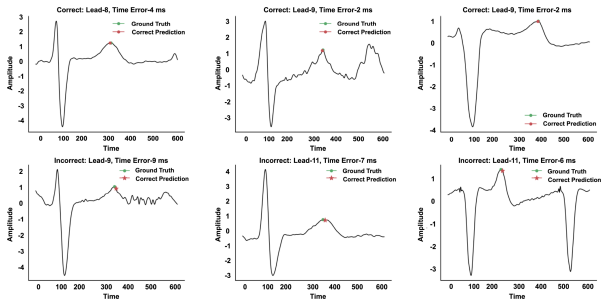

Fig. 9. Visualization of correct (Time Error $\leq$ 5 ms, top row) and incorrect (Time Error $>$ 5 ms, bottom row) T peak predictions of our ECGVEDNET.

the two datasets, choose the best parameters and report the best performance.

1) Comparison Experiments on ICDIRS: Table IV presents the quantitative performance comparisons of different delineation methods on ICDIRS. The accuracy of ECGVEDNET for QRS onset detection are 86.28% and 88.31% at 5 ms and 10 ms, respectively, while the accuracy for T peak detection are 89.94% and 91.16%, respectively. ECGVEDNET performs with a 11.89% average error at 5 ms tolerance and a 10.27% average error at 10 ms tolerance for QRS onset and T peak points detection. It achieves the state-of-the-art performance on ICDIRS, mainly because its encoder encodes the input into a regularized latent space, where the decoder can capture morphology invariant features. The morphology invariant features enables ECGVEDNET to detect the fiducial points in the ECGs of various morphology.

Fig. 8 shows some examples of correct and incorrect QRS onset predictions under 5 ms tolerance, while Fig. 9 shows some examples of correct and incorrect T peak predictions under 5 ms tolerance. From the figures, we observe that most of the incorrect predictions are generated due to the minor time errors (5<time error<10).

The signal processing based delineation methods of Martinez et al. [7] and Pilia et al. [42] are both developed based on the wavelet transform (WT) technique. They first detects the R peaks in order to take the points as the references, then the detection of other fiducial points, such as QRS onsets and T peaks, is performed. However, the wavelet transform based methods require setting some thresholds manually and carefully to achieve good results. Hence, they may not generalise on new datasets and their performances on new datasets are highly relying on the setting of the thresholds. The methods of Camps et al. [13] and Guillermo [33] et al. achieves much better performance than previous two methods, which shows the significant advantages of the deep learning based methods over the signal processing based methods on large-scale datasets. Moreover, deep learning methods have the potential advantage that they can be adapted to new datasets with much less human intervention than signal processing based methods, which require carefully threshold setting.

2) Comparison Experiments on QTDB: We train our ECGVEDNET with a batch size of 128 on QTDB. Adam is employed as the optimizer with a learning rate of 0.0001. The

TABLE V
DELINEATION ERROR (MEAN ± STANDARD DEVIATION) OF DIFFERENT DELINEATION METHODS ON QTDB

Methods	Types	Techniques	QRS onset (m ± sd)	T peak (m ± sd)
Martinez et al. [7]	Signal Processing	Wavelet Transform	4.22 ± 15.20	0.80 ± 13.82
Pilia et al. [42]	Signal Processing	Wavelet Transform	-2.18 ± 11.32	-0.94 ± 16.30
Camps [13]	Deep Learning	ConvNet	-3.76 ± 14.02	-1.38 ± 19.06
Guillermo [33]	Deep Learning	Encoder-Decoder	-4.60 ± 19.51	-1.56 ± 20.20
ECGVEDNET	Deep Learning	Variational Encoder-Decoder	-3.02 ± 12.83	-1.26 ± 18.66
ECGVEDNET-Transfer	Deep Learning	Variational Encoder-Decoder	**-1.86 ± 8.02**	**-0.50 ± 12.96**

The bold values indicates the best performed models.

TABLE VI
DELINEATION ERROR (MEAN ± STANDARD DEVIATION) OF THE DEEP MODELS WITH AND WITHOUT TRANSFER LEARNING

Methods	Transfer	QRS onset (m ± sd)	T peak (m ± sd)
Camps [13]	✗	-3.76 ± 14.02	-1.38 ± 19.06
Camps [13]	✓	-2.82 ± 12.40	-1.10 ± 15.30
Guillermo [33]	✗	-4.60 ± 19.51	-1.56 ± 20.20
Guillermo [33]	✓	-3.75 ± 15.00	-1.20 ± 18.33
ECGVEDNET	✗	-3.02 ± 12.83	-1.26 ± 18.66
ECGVEDNET	✓	**-1.86 ± 8.02**	**-0.50 ± 12.96**

The bold values indicates the best performed models.

TABLE VII
DELINEATION ERROR (MEAN ± STANDARD DEVIATION) OF THE ECGVEDNETS WITH DIFFERENT DEPTHS

Depths	Transfer	QRS onset (m ± sd)	T peak (m ± sd)
3 blocks	✗	**-2.46 ± 10.02**	**-0.98 ± 14.03**
4 blocks	✗	-2.88 ± 11.70	-1.13 ± 16.08
5 blocks	✗	-3.02 ± 12.83	-1.26 ± 18.66
6 blocks	✗	-3.65 ± 14.06	-1.95 ± 18.36
3 blocks	✓	-2.42 ± 10.83	-0.96 ± 13.93
4 blocks	✓	-2.15 ± 9.96	-0.84 ± 13.48
5 blocks	✓	**-1.86 ± 8.02**	**-0.50 ± 12.96**
6 blocks	✓	-1.99 ± 9.37	-0.70 ± 13.20

The bold values indicates the best performed models.

training stops after 30 epochs as the training loss has already converged. The quantitative comparisons of different delineation algorithms are presented in Table V, from which we observe that, among four comparison methods, the signal processing based method of Martinez et al. [7] achieves the best performance for T peak detection while the signal processing based method of Pilia et al. [42] achieves the best performance for QRS onset detection. It's very interesting to observe that these two signal processing based methods achieves better performance than the deep learning based methods of Camps [13] and Guillermo [33]. This is mainly because the deep learning based methods cannot be well trained with limited training data, which greatly hinders the performance.

More interestingly, we find that our shallow ECGVEDNETs performs better than the deep ones. For example, as shown in Table VII, our ECGVEDNET under the 3 blocks setting performs much better than the one under the 5 blocks setting. We assume this is because QTDB is a much smaller dataset than ICDIRS, which cannot provide sufficient data for training the deeper models. The shallow networks requires less training data, and thus achieve better performance than the deep ones on small datasets.

To further verify this assumption, we introduce the transfer learning strategy to train our ECGVEDNET on QTDB and find our ECGVEDNET trained with the transfer learning strategy achieves much better performance than all the comparison methods. We give the details of the transfer learning strategy in the following section.

3) Transfer Learning Between ICDIRS and QTDB: From the experiments above, we assume the QTDB dataset is not large enough to train a good deep learning model, which highly relies on the large-scale datasets. Hence, we employ our transfer learning framework to train our ECGVEDNET, the deep model of Camps et al. [13] and the deep model of Guillermo et al. [33] on QTDB. Table VI presents the performance of three deep learning methods trained with and without our transfer learning strategy, the transfer learning framework greatly improves the performance of all the deep learning methods. The transfer learning strategy takes the pre-trained model from ICDIRS as prior knowledge, greatly improving the models' performance on QTDB. All these evidences show the advantages of the usage of the larger-scale datasets, and highlights that insufficient training data is one of the major hindrances of deep delineation models.

To further verify the effectiveness of the transfer learning framework, we apply it to train our ECGVEDNETs with different depths. As shown in Table VII, the transfer learning framework greatly reduces the delineation error of all the ECGVEDNETs, especially for the deep ones. For example, the ECGVEDNET under 5 blocks setting outperforms the one under 3 blocks setting. This is because the model capacities of the deep networks are higher than the shallow ones and the transfer learning strategy can alleviate the insufficient data problem and take full advantage of the deep networks' high capacity.

VI. CONCLUSION

Deep learning models are data-hungry, requiring a large amount of data for better performance. To promote the development of deep learning in the field of ECG delineation, we collect a large-scale ECG dataset ICDIRS which contains the largest QRS onset and T wave peak annotated database to date. This dataset contains sufficient data (>150,000 manually labelled QRS onsets and T peaks) to train the deep learning models, and also provide a useful a platform for the researchers to evaluate and develop advanced deep ECG delineation models. Moreover, we propose a novel variational autoencoder network, named ECGVEDNET, for QRS onset and T peak detection. The morphology of ECG collected from the same person or

different persons presents large variations, which hinder the performance of deep learning models. To address the morphology variations challenge, our ECGVEDNET encodes the raw ECG data into a regularized latent space, where diverse ECG data can extract regularized latent features. The regularized latent features present much smaller variations than the raw ECG data, greatly benefiting the accuracy of ECG delineation models that fail due to inherent morphology variation.

ECGVEDNET achieves state-of-the-art performance on the large-scale dataset ICDIRS, however, it still cannot achieve satisfactory performance on the small-scale dataset QTDB due to the insufficient training data. Hence, we propose a transfer learning framework to break the limitations of deep learning models on small-scale datasets. The transfer learning framework first learn the general knowledge on the large-scale dataset ICDIRS, then exploiting the learned knowledge to assist the training of deep models on the small-scale datasets, this strategy enables our ECGVEDNET achieve the state-of-the-art performance on the small-scale dataset QTDB. It is also worth noting that the proposed ECGVEDNET requires fixed input size and cannot handle varied input sizes. Hence, we cannot directly perform fine-tuning on a new dataset when the input size changes. Hence, in the further work, an improved deep delineation network, which can accept varied input sizes, is expected.

References

[1] L. Smital et al., "Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 10, pp. 2721–2734, Oct. 2020.

[2] Z. Jiang et al., "Diagnostic of multiple cardiac disorders from 12-lead ECGs using graph convolutional network based multi-label classification," in *Proc. Comput. Cardiol.*, 2020, pp. 1–4.

[3] F. Khosrow-Khavar et al., "Automatic and robust delineation of the fiducial points of the seismocardiogram signal for noninvasive estimation of cardiac time intervals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1701–1710, Aug. 2017.

[4] G. D. Clifford et al., *Advanced Methods and Tools for ECG Data Analysis*, vol. 10. Norwood, MA, USA: Artech House, 2006.

[5] G. D. Lannoy et al., "Supervised ECG delineation using the wavelet transform and hidden Markov models," in *Proc. 4th Eur. Conf. Int. Federation Med. Biol. Eng.*, 2009, pp. 22–25.

[6] S. Graja and J.-M. Boucher, "Hidden Markov tree model applied to ECG delineation," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 6, pp. 2163–2168, Dec. 2005.

[7] J. P. Martínez et al., "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[8] R. Almeida et al., "P wave delineation using spatially projected leads from wavelet transform loops," in *Proc. Comput. Cardiol.*, 2010, pp. 1003–1006.

[9] A. Martínez, R. Alcaraz, and J. J. Rieta, "A new method for automatic delineation of ECG fiducial points based on the Phasor Transform," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, 2010, pp. 4586–4589.

[10] P.-C. Chen, S. Lee, and C.-D. Kuo, "Delineation of T-wave in ECG by wavelet transform using multiscale differential operator," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1429–1433, Jul. 2006.

[11] C. Böck et al., "ECG beat representation and delineation by means of variable projection," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 10, pp. 2997–3008, Oct. 2021.

[12] J. M. Bote et al., "A modular low-complexity ECG delineation algorithm for real-time embedded systems," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 429–441, Mar. 2018.

[13] J. Camps, B. Rodríguez, and A. Mincholé, "Deep learning based QRS multilead delineator in electrocardiogram signals," in *Proc. Comput. Cardiol. Conf.*, 2018, pp. 1–4.

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

[15] L. Chen et al., "Spatio-temporal ECG network for detecting cardiac disorders from multi-lead ECGs," in *Proc. Comput. Cardiol.*, 2021, pp. 1–4.

[16] H. Abrishami et al., "Supervised ECG interval segmentation using LSTM neural network," in *Proc. Int. Conf. Bioinf. Comput. Biol.*, 2018, pp. 71–77.

[17] R. Almeida et al., "Multilead ECG delineation using spatially projected leads from wavelet transform loops," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 1996–2005, Aug. 2009.

[18] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P-and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2840–2849, Dec. 2010.

[19] E. Soria-Olivas et al., "Application of adaptive signal processing for determining the limits of P and T waves in an ECG," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 1077–1080, Aug. 1998.

[20] H. D. Hesar and M. Mohebbi, "A multi rate marginalized particle extended Kalman filter for P and T wave segmentation in ECG signals," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 112–122, Jan. 2019.

[21] A. Martínez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points," *Physiol. Meas.*, vol. 31, no. 11, 2010, Art. no. 1467.

[22] M. Akhbari et al., "ECG fiducial point extraction using switching Kalman filter," *Comput. Methods Prog. Biomed.*, vol. 157, pp. 129–136, 2018.

[23] F. Rincón et al., "Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 854–863, Nov. 2011.

[24] A. Alcaine et al., "A wavelet-based electrogram onset delineator for automatic ventricular activation mapping," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 12, pp. 2830–2839, Dec. 2014.

[25] J. Dumont, A. I. Hernandez, and G. Carrault, "Improving ECG beats delineation with an evolutionary optimization process," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 607–615, Mar. 2010.

[26] Q. Xue, Y. H. Hu, and W. J. Tompkins, "Neural-network-based adaptive matched filtering for QRS detection," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 4, pp. 317–329, Apr. 1992.

[27] R. Poli, S. Cagnoni, and G. Valli, "Genetic design of optimum linear and nonlinear QRS detectors," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 11, pp. 1137–1141, Nov. 1995.

[28] I. Saini, D. Singh, and A. Khosla, "K-nearest neighbour-based algorithm for P-and T-waves detection and delineation," *J. Med. Eng. Technol.*, vol. 38, no. 3, pp. 115–124, 2014.

[29] Z. Chen et al., "Post-processing refined ECG delineation based on 1D-UNet," *Biomed. Signal Process. Control*, vol. 79, 2023, Art. no. 104106.

[30] A. N. Londhe and M. Atulkar, "Semantic segmentation of ECG waves using hybrid channel-mix convolutional and bidirectional LSTM," *Biomed. Signal Process. Control*, vol. 63, 2021, Art. no. 102162.

[31] S. Nurmaini et al., "Beat-to-beat electrocardiogram waveform classification based on a stacked convolutional and bidirectional long short-term memory," *IEEE Access*, vol. 9, pp. 92600–92613, 2021.

[32] A. Peimankar and S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," *Expert Syst. Appl.*, vol. 165, 2021, Art. no. 113911.

[33] G. Jimenez-Perez, A. Alcaine, and O. Camara, "U-Net architecture for the automatic detection and delineation of the electrocardiogram," in *Proc. Comput. Cardiol.*, 2019, pp. 1–4.

[34] X. Liang et al., "ECG_segnet: An ECG delineation model based on the encoder-decoder structure," *Comput. Biol. Med.*, vol. 145, 2022, Art. no. 105445.

[35] J. Wang et al., "A knowledge-based deep learning method for ECG signal delineation," *Future Gener. Comput. Syst.*, vol. 109, pp. 56–66, 2020.

[36] W. Xu et al., "Variational autoencoder for semi-supervised text classification," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 3358–3364.

[37] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture IE*, vol. 2, no. 1, pp. 1–18, 2015.

[38] P. Laguna et al., "A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG," in *Proc. Comput. Cardiol.*, 1997, pp. 673–676.

[39] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[40] B. Tutuko et al., "Short single-lead ECG signal delineation-based deep learning: Implementation in automatic atrial fibrillation identification," *Sensors*, vol. 22, no. 6, 2022, Art. no. 2329.

[41] S. Nurmaini et al., "Electrocardiogram signal classification for automated delineation using bidirectional long short-term memory," *Inform. Med. Unlocked*, vol. 22, 2021, Art. no. 100507.

[42] N. Pilia et al., "ECGdeli-an open source ECG delineation toolbox for matlab," *SoftwareX*, vol. 13, 2021, Art. no. 100639.