Epileptic Seizure Detection in EEG Signals Using a Unified Temporal-Spectral Squeeze-and-Excitation Network

Yang Li, Yu Liu, Wei-Gang Cui, Yu-Zhu Guo, Hui Huang, and Zhong-Yi Hu

***Abstract*— The intelligent recognition of epileptic electro-encephalogram (EEG) signals is a valuable tool for the epileptic seizure detection. Recent deep learning models fail to fully consider both spectral and temporal domain representations simultaneously, which may lead to omitting the nonstationary or nonlinear property in epileptic EEGs and further produce a suboptimal recognition performance consequently. In this paper, an end-to-end EEG seizure detection framework is proposed by using a novel channel-embedding spectral-temporal squeeze-and-excitation network (CE-stSENet) with a maximum mean discrepancy-based information maximizing loss. Specifically, the CE-stSENet firstly integrates both multi-level spectral and multi-scale temporal analysis simultaneously. Hierarchical multi-domain representations are then captured in a unified manner with a variant of squeeze-and-excitation block. The classification net is finally implemented for epileptic EEG recognition based on features extracted in previous subnetworks. Particularly, to address the fact that the scarcity of seizure events results in finite data distribution and the severe overfitting problem in seizure detection, the CE-stSENet is coordinated with a maximum mean discrepancy-based information maximizing loss for mitigating the overfitting problem. Competitive experimental results on three EEG datasets against the state-of-the-art methods demonstrate the effectiveness of the proposed framework in recognizing epileptic EEGs, indicating its powerful capability in the automatic seizure detection.**

Manuscript received October 8, 2019; revised January 17, 2020; accepted January 30, 2020. Date of publication February 12, 2020; date of current version April 8, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1809209, Grant 61671042, Grant 61403016, Grant 61573356, and Grant 61702376, in part by the Beijing Natural Science Foundation under Grant L182015 and Grant 4172037, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LSZ19F020001, and in part by the Major Project of Wenzhou Natural Science Foundation under Grant ZY2019020. *(Corresponding author: Yu Liu.)*

Yang Li is with the Department of Automation Sciences and Electrical Engineering, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100083, China (e-mail: liyang@buaa.edu.cn).

Yu Liu, Wei-Gang Cui, and Yu-Zhu Guo are with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100083, China (e-mail: sy1803113@buaa.edu.cn; cuiweigang94@foxmail.com; yuzhu.guo@sheffield.ac.uk).

Hui Huang and Zhong-Yi Hu are with the Intelligent Information Systems Institute, Wenzhou University, Wenzhou 325035, China (e-mail: huanghui@wzu.edu.cn; hujunyi@163.com).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.2973434

***Index Terms*— EEG, deep learning, multi-domain feature extraction, squeeze & excitation, maximum mean discrepancy, seizure detection.**

I. Introduction

EPILEPSY is the second most common neuronal disease of the brain, which is characterized by transient and sudden abnormal disturbance of brain neurons and affects over 50 million people worldwide [1]. Generally, manual seizure detection requires long-duration visual inspection of EEG signals, whose efficacy could be deteriorated by latencies including contaminations and discrepant clinical standards of different physicians [2]. The intelligent recognition of epileptic EEGs has demonstrated to be a promising alternative for manual seizure detection. However, it is not an easy task due to the severe nonstationary and stochastic characteristic of EEG signals [3].

In order to realize automatic epileptic seizure detection efficiently, many machine learning techniques have been proposed [2]–[4]. Most of these studies firstly extracted discriminative features in two categories, namely spectral-domain [3], [4] and temporal-domain [5]. However, features extracted in a fixed manner ignore the heterogeneity among patients, which potentially undermines the effectiveness of these pre-defined features for EEGs recognition [6].

Recently, deep learning methods can tackle above problems to some extent, aiming to exploit hierarchical spectral or temporal patterns alternatively [7]. For example, as for scalogram-based convolutional neural network (CNN), Thodoroff *et al.* [8] converted EEGs into image-based representations with spectral property and electrode montage knowledge preserved, which were fed into CNN and bidirectional recurrent neural network for seizure detection. Yuan et al. [9] further combined convolutional autoencoder with CNN for multi-view spectrogram representations learning based on the short time Fourier transform (STFT) method. However, studies aforementioned only consider prior representations of raw EEGs, which may lead to neglecting temporal patterns embedded in original EEG signals and ignoring stochastic nature in EEGs consequently [6]. On the contrary, end-to-end deep learning architectures explore hierarchical temporal patterns from raw EEGs directly without hand-engineered representations. For instance,

Schirrmeister *et al.* [6] proposed a deep CNN for EEG decoding. A 13-layer deep CNN was used for discriminating normal, inter-ictal and ictal EEG segments [10]. However, a conventional convolution layer essentially equates to a low-pass filter while discarding high-frequency components [7]. This constraint suggests the merely partial spectral-domain analysis in an end-to-end CNN architecture and undermines the nonstationary of EEGs in nature. In aggregate, recent deep learning approaches achieve a limited success in EEG recognition due to their failure to reconcile spectral and temporal analysis simultaneously [11]. Therefore, a unified seizure detection framework, which benefits from complementary information from both spectral and temporal domain simultaneously, should be highly required.

In order to explore supplementary information in spectral-temporal domain for EEG recognition, most of recent studies merely construct independent multi-stream architectures [12], [13] or extract features in a domain-agnostic manner [14]. However, it is far from satisfactory to only concatenate parallel branches for multi-domain analysis or mix them blindly. Recently, an architectural unit, termed the squeeze & excitation (SE) block, was investigated to selectively highlight informative channel-wise feature responses by explicitly modeling interdependencies between channels of its intermediate features, which has gained encouraging success in image classification and segmentation [15]. Based on this motivation, in this paper, we aim at leveraging the high recognition performance of the SE block to supplementary spectral-temporal pattern learning by boosting discriminative spectral-temporal representations unifiedly.

Additionally, the scarcity of seizure events usually results in the highly imbalance characteristic of seizure datasets [1]. As a consequence, the severe overfitting problem hinders the application of deep learning algorithms in a prevalent clinical scenario, *i.e.*, seizure onset detection. Unfortunately, conventional data augmentation techniques may not work well for epileptic EEGs due to the severe nonstationary and stochastic property of EEG signals [16]. Encouragingly, the use of amortized variational inference, which refers to replace instance-specific local inference with global, *i.e.*, amortized, inference networks, has demonstrated to be a feasible solution to avoid overfitting problems caused by finite data distribution [17]. However, few similar attempts have been made in the context of supervised recognition of epileptic EEGs.

To address aforementioned issues, in this paper, we propose a novel end-to-end deep learning framework, called channel-embedding spectral-temporal squeeze-and-excitation network (CE-stSENet) with a maximum mean discrepancy-based information maximizing loss. Specifically, the CE-stSENet includes three subnetworks as follows. The channel-embedding spectral-temporal net (CE-Spectral-Temporal net) firstly takes raw EEGs as inputs, which is represented as a set of dynamic filter-band components, and further employs multi-level spectral analysis and multi-scale temporal analysis parallelly. Secondly, the spectral-temporal squeeze-and-excitation net (ST-SENet) consists of a series of parallel subnetworks, each of which takes a certain group of spectral-temporal representations as inputs and captures multi-domain feature responses independently while exaggerating discriminative ones jointly. The classification net is finally implemented for epileptic EEG recognition. Additionally, these three subnetworks are coordinated with two types of loss: cross-entropy loss for prediction and maximum mean discrepancy-based information maximizing loss for mitigating the overfitting problem. Our proposed framework is evaluated on three public epileptic EEG datasets with several independent classification cases for automatic seizure detection. Experimental results demonstrate that the proposed method can capture discriminative features embedded in EEG recordings efficiently and achieve a better classification performance against the state-of-the-art methods on these three datasets.

Main contributions of our study can be summarized as follows: (1) A novel CE-stSENet is proposed for automatic seizure detection. To the best of our knowledge, although many CNN-based researches have been studied for automatic seizure detection, this is the first attempt to integrate fully spectral and temporal domain embeddings into one end-to-end deep learning model simultaneously. Besides, we innovatively realize single-level spectral analysis by a generalized convolutional operator, termed wavelet convolution (waveConv) layer. (2) A variant of the SE block, termed group convolution squeeze & excitation (gcSE) unit, is introduced to explore spectral-temporal embeddings unifiedly. (3) A joint training objective including a maximum mean discrepancy-based information maximizing loss is used to mitigate overfitting problems caused by scarcity of seizure events and further produce a better seizure detection performance. As a result, a consistent improvement with different classification cases is achieved by our proposed scheme for detecting epileptic EEG signals effectively.

II. Methodology

In this section, we firstly define the notations used in this paper. Secondly, basic subnetworks in the CE-stSENet including the CE-Spectral-Temporal net, the ST-SENet and the classification net are described. Next, we present a joint training objective including a maximum mean discrepancy-based information maximizing loss. Finally, implement details are given. Additionally, the model with superscription * indicates that it is trained with maximum mean discrepancy-based information maximizing loss, for example CE-stSENet*.

A. Definitions and Notations

For the i-th classification case conducted in this paper, we define the epileptic dataset as $D_i = \{(X_1, y_1), \cdots, (X_{N_i}, y_{N_i})\}$, where N_i is the total number of labeled EEG segments, the matrix $X_j \in \mathrm{R}^{E \times T}$ is an EEG segment with E channels and T sampling indexes at a sampling rate of 2^{f_i} Hz. f_i is equal to $\lfloor \log_2(s_i) \rfloor$ where s_i is the original sampling rate in the i-th classification case and $\lfloor\ \rfloor$ is the rounding down operation respectively.

B. Channel-Embedding Spectral-Temporal Squeeze-and-Excitation Network

The layout of the proposed CE-stSENet* is given in Fig. 1. Specifically, the CE-Spectral-Temporal net firstly embeds

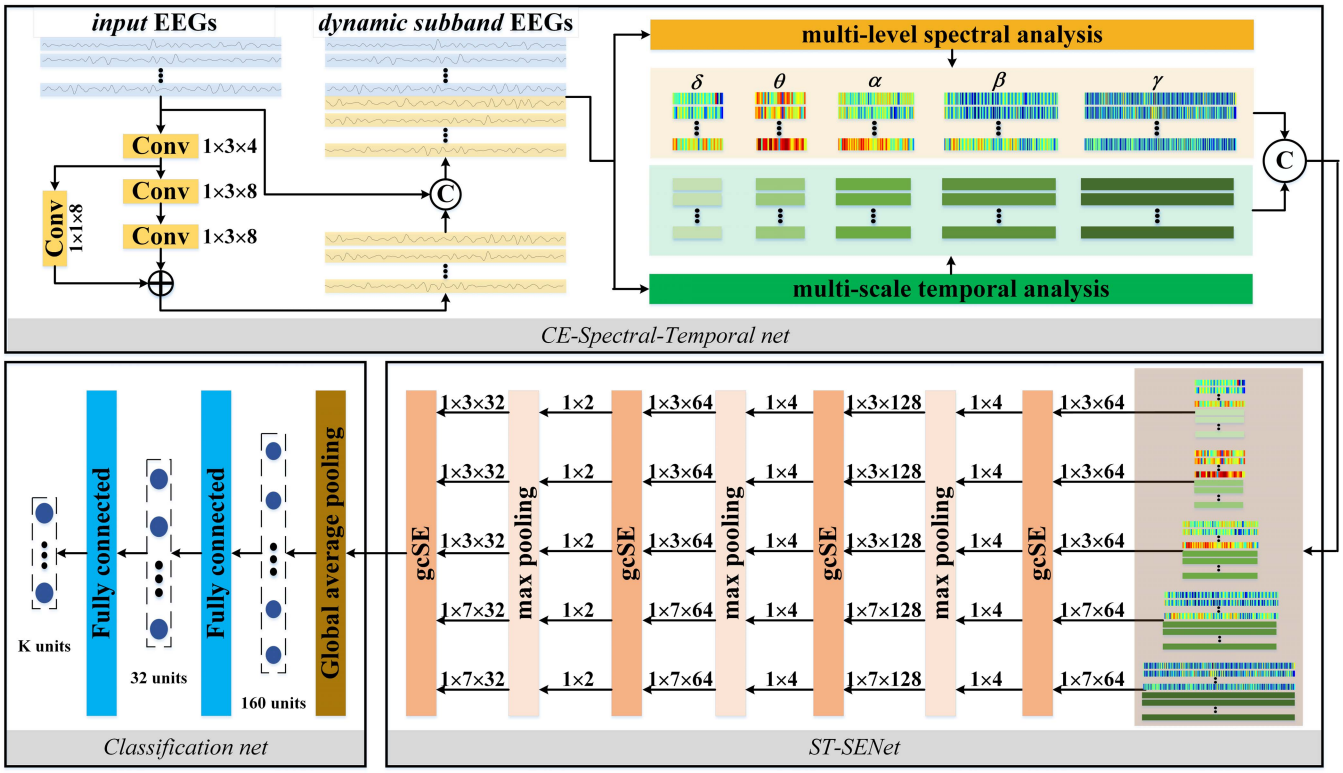

Fig. 1. The proposed CE-stSENet* framework, where 3-D feature maps are flatted channel-wisely for visualization, the ST-SENet block involves 5 independent subnets with the same configuration and different parameters, which are denoted parallelly.

input EEG segments and employs multi-level spectral analysis and multi-scale temporal analysis simultaneously. Then the ST-SENet explores spectral-temporal features unifiedly by using the gcSE unit for boosting cross-domain discriminative representations. The classification net is finally used to map input features to final epileptic seizure detection.

1) The Design of the CE-Spectral-Temporal Net: In order to adaptively construct an optimal filter-band for the subsequent spectral-temporal analysis, successive convolution and batch normalization operations are employed before the multi-level spectral and multi-scale temporal analysis block. As a result, the CE-Spectral-Temporal net firstly maps EEG segment with a size of $E \times 1 \times T$ into a set of time-series-like embeddings of $8 \times 1 \times T$. All convolution filters share the same size of 1×3 with a stride of 1 and padding of 1, which ensures output feature maps with the same size as input EEGs. Moreover, we adopt skip-connection to avoid gradient vanishing and accelerate convergence [18]. With no activation function inserted, we stack input EEGs and output embeddings via a channel-wise concatenating function, thus forming a dynamic subband EEGs matrix $M \in \mathrm{R}^{C\times 1\times T}$ for subsequent networks where C is equal to $E+8$. Compared with applying the spectral-temporal analysis at the beginning of the model, subband EEGs matrix M provides the subsequent network with different representations of input EEGs at different scales and the original EEG time series as well.

Secondly, in order to integrate fully spectral-domain analysis into an end-to-end deep learning model, the multi-level periodic-padding wavelet decomposition is implemented by using a generalized convolution operator iteratively, named waveConv layer. Given a 1-D input x with N elements, the waveConv layer is defined by:

$$x_p = x(N - \frac{R}{2} + 1), \cdots, x(N-1)|x(0), \cdots, x(N-1)|x(0), \cdots, x(\frac{R}{2} - 2) \quad (1)$$

$$y_A(i) = (x_p \otimes g)(i) = \sum_{r=0}^{R} x_p(s \times i - r) \times g(r)$$

$$y_D(i) = (x_p \otimes h)(i) = \sum_{r=0}^{R} x_p(s \times i - r) \times h(r) \quad (2)$$

where $|$ is a concatenating operator, $x(i)$ is the i-th element of input x, $\otimes$ is a convolution operator, g and h represent a pair of scaling and wavelet filter, and R, s are the parameter *kernel size* and *stride* in convolution operator, respectively. Compared with other extension modes including zero-padding, periodic-padding operator can effectively avoid border distortions and ensure a minimal wavelet decomposition length as well [19]. However, a convolution operation with a *stride* of 2, which essentially retains odd-numbered elements after the convolution [18], is not compatible with downsampling operation in wavelet decomposition, since the downsampling operation saves even-numbered components instead [19]. The modified periodic-padding operator in (1) can resolve this issue and enable downsampling operation in wavelet decomposition replaced by the parameter *stride*, *i.e.*, $s = 2$ in (2), since it pads $\lfloor R/2 - 1 \rfloor$ elements rather than $\lfloor R/2 \rfloor$. Similarly, a single-level spectral analysis for dynamic subband EEGs matrix M can be deployed by a waveConv layer with the parameter *group* in the convolution operator set to C. Consequently, fully spectral-domain analysis can be inserted into

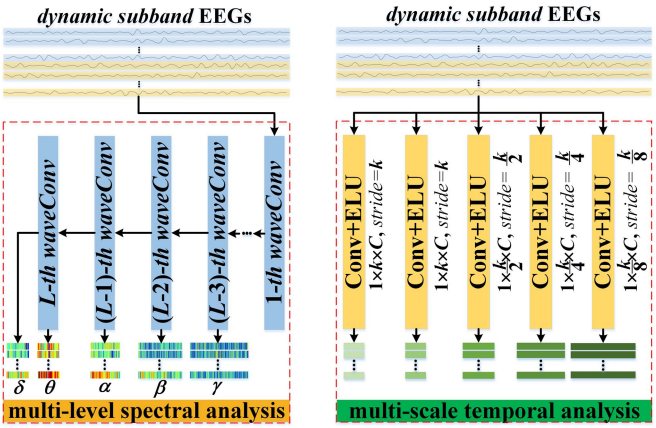

Fig. 2. A schematic illustration of the spectral-temporal analysis. (a) multi-level spectral analysis, (b) multi-scale temporal analysis.

an end-to-end deep learning model for the better EEG recognition. Specifically, the waveConv layer iteratively processes approximation coefficients generated by upper level, denoted as y_A in (2), until the level L is reached. In order to obtain wavelet coefficients corresponding to 32-64 Hz (γ rhythm), 16-32Hz (β rhythm), 8-16Hz (α rhythm), 4-8Hz (θ rhythm) and 0-4Hz (δ rhythm), which satisfy the clinical interests [20], the level L is equal to $f_i - 3$, which is determined by sampling rate. We select Daubechies order-4 (Db4) wavelet for our experiments [3], which indicates no learnable parameters involved in a waveConv layer. Note that any pair of quadrature mirror filters is theoretically qualified as an alternative for Db4 wavelet, which also implies the negligible possibility of a conventional convolution layer to employ spectral analysis as the waveConv layer do, due to its data-driven training mode while ignoring biorthogonal constraint on learnable filters. The structure of the multi-level spectral analysis is illustrated in Fig. 2(a).

Meanwhile, considering the variation of seizures inter- and intra-patients, interested temporal patterns may be embedded in origin epileptic EEGs at multiple scales. Therefore, we implement 5 independent convolution, batch normalization and exponential linear unit (ELU) operations with different receptive fields to capture multi-scale temporal representations with *kernel size* empirically set to be $\{k, k, k/2, k/4, k/8\}$ where $k = 2^{f_i-3}$. The parameter *Stride* in convolution operators are determined to generate coarser temporal representations with a consistent size as spectral ones above. The structure of the multi-scale temporal analysis is given in Fig. 2(b).

In order to guarantee a compact design of the subsequent model architecture, we further concatenate above spectral and temporal representations channel-wisely. As a result, the CE- Spectral-Temporal net captures preliminary multi-domain embeddings effectively, resulting in five groups of spectral-temporal representations.

2) The Design of the ST-SENet: Motivated by the high independence between coefficients of Db4 wavelet transform, the ST-SENet branches into 5 independent subnets, each of which takes a specific group of spectral-temporal representations generated by the CE-Spectral-Temporal net as inputs. Within each subnet, we further explore hierarchical multi-domain patterns unifiedly. Therefore, a variant of the SE unit is constructed by integrating group convolution operator into the SE block, which is called the gcSE unit.

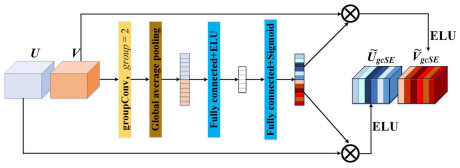

Fig. 3. A diagram of the proposed gcSE unit.

The gcSE unit is composed of group convolution, SE block, batch normalization and ELU operations, successively. Considering the heterogeneity of multi-domain features in nature, we deploy group convolution on spectral-temporal embeddings with the parameter *group* of 2, which reduces computational complexity at a large margin as well [21]. However, although no information interflow is allowed between spectral and temporal branches, both of them can be interpreted as a collection of local descriptors whose statistics are discriminative for the whole input EEGs [15]. Therefore, channel-wise feature responses boosting, *i.e.*, the SE block, should not be restricted within single-domain, but qualified and recalibrated jointly. Considering input spectral representations $U = [u_1, u_2, \cdots, u_n]$ with $u_i \in \mathrm{R}^{1\times w}$ and temporal ones $V = [v_1, v_2, \cdots, v_n]$ with $v_i \in \mathrm{R}^{1\times w}$, the SE block firstly embeds the global distribution of channel-wise feature responses as a statistic $z \in \mathrm{R}^{1\times 1\times (C_n+C_n)}$ by aggregating u_i and v_i across temporal dimension, termed squeeze operation. The k-th element of z is defined by:

$$z_k = \frac{1}{w}\sum_{i=1}^{w} H(U, V)(k, 1, i) \tag{3}$$

where $H(\cdot)$ is a channel-wise concatenating function. Then an excitation operation models the channel interdependence through a self-gating mechanism $\tilde{z} = \sigma(W_2\delta(W_1 z))$ with $W_1 \in \mathrm{R}^{((C_n+C_n)/r)\times(C_n+C_n)}$ and $W_2 \in \mathrm{R}^{(C_n+C_n)\times((C_n+C_n)/r)}$ being weights of two fully connected layers and $\delta(\cdot)$ and $\sigma(\cdot)$ referring to the ELU and Sigmoid activation respectively. The choice of reduction ratio r will be discussed in section III-C. Finally, the SE block recalibrates input U and V to boost multi-domain feature discriminability jointly:

$$\begin{aligned}\tilde{U}_{gcSE} &= [\tilde{z}_1 u_1, \tilde{z}_2 u_2, \cdots, \tilde{z}_{C_n} u_{C_n}]\\ \tilde{V}_{gcSE} &= [\tilde{z}_{C_n+1} v_1, \tilde{z}_{C_n+2} v_2, \cdots, \tilde{z}_{C_n+C_n} v_{C_n}]\end{aligned} \tag{4}$$

where $\tilde{z}_i$ is the i-th element of $\tilde{z}$.

Subsequent batch normalization and ELU operations are adopted to further enhance robustness and nonlinearity activations. A diagram of the proposed gcSE unit is shown in Fig. 3.

As a result, each branch of the ST-SENet adopts successive gcSE unit and maxpooling operator to generate coarser multi-domain representations with discriminative ones dynamically exaggerated. As in Fig. 1, total 160 multi-domain feature maps are transported to the following classification net.

3) The Design of the Classification Net: The classification net finally serves as a multilayer perceptron (MLP) for the final recognition of epileptic EEGs. The output of the ST-SENet is firstly compressed by a global pooling layer and vectorized to a 1-D vector before fed into the first fully connected layer of the classification net. As a result, the label with maximum probability is decided as the final classification result of the CE-stSENet*.

C. Joint Training Objective for Overfitting Resistance

Generally, the recognition of epileptic EEGs aims to learn a map function $f(Xi;\theta)$: $\mathrm{R}^{E\times T} \rightarrow L$, which is defined by θ, to assign label y_i to EEG segment X_i correctly. Most recent learning-based EEG recognition methods optimize θ through maximizing log likelihood in the following form:

$$\arg\max_{\theta} \sum_{(X,y)\in D} \log p(y|X) \tag{5}$$

where the set D is the epileptic dataset in Section II-A. However, due to the scarcity of seizure events, *i.e.*, $p(x)$ is a finite distribution, matching $p(x)$ too closely may lead to severe overfitting [17]. Variational autoencoders (VAE) [22] mitigate empirical distribution caused overfitting problems in unsupervised learning by regularizing the latent space [17]. In other words, VAE encodes inputs as a distribution over the latent space rather than a set of individual points, which is commonly referred as amortized variational inference (AVI). Specifically, the VAE assumes a prior distribution $p(z)$ over latent variables z, *i.e.*, features, and optimizes both the marginal likelihood of input data and the quality of the AVI, commonly referred as $q_\phi(z|x)$, simultaneously. Inspired by the VAE, we innovatively introduce a maximum mean discrepancy-based information maximizing loss to alleviate the overfitting problem in supervised recognition of epileptic EEGs. Specifically, we treat the input of the classification net as latent variables in the VAE, denoted as $z = [z_1, z_2, \cdots, z_5]$ with $z_i \in \mathrm{R}^{1\times 32}$. Considering two main properties that can express the regularity of the latent space, *i.e.*, continuity and completeness, both the covariance matrix and the mean of the distribution of latent variables need to be regularized [22]. In practice, this procedure is accomplished by enforcing the covariance matrix to be close to identity, which prevents punctual distributions, and the mean to be close to 0, which prevents encoded distributions to be too far apart from each other [22]. Therefore, we impulse the Gaussian normal distribution $p(z_i)$ on z_i, which is the same as the implementation in the VAE [17]. We further define an AVI $q_\varphi(z|x)$ and alternate log likelihood in (5) with a lower bound:

$$\begin{aligned} L_{\mathrm{ELBO}}(x) = &-D(q_\phi(z|x) \parallel p(z)) \\ &+ E_{q_\phi(z|x)}[\log p(y|z)] + \log(p(y|x)/p(y)) \end{aligned} \tag{6}$$

where $D(\cdot)$ is a divergence matrix that measures the discrepancy between distribution Q_ϕ and P, which is selected as Kullback–Leibler divergence in the VAE. Note that it can be replaced by any strict divergence. We adopt maximum mean discrepancy (MMD) due to its compensation for AVI failures encountered by the conventional VAE [17]:

$$\begin{aligned} D_{\mathrm{MMD}}(q_\phi||p) = &\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N} k(x_i, x_{i'}) \\ &-\frac{2}{N\times M}\sum_{i=1}^{N}\sum_{j=1}^{M} k(x_i, y_j) \\ &+\frac{1}{M^2}\sum_{j=1}^{M}\sum_{j'=1}^{M} k(y_j, y_{j'}) \end{aligned} \tag{7}$$

where $\{x_i\}_{i=1}^{N}$ and $\{y_j\}_{j=1}^{M}$ are latent variables and data sampled from the Gaussian normal distribution respectively, $k(\cdot,\cdot)$ is any positive definite kernel. The Gaussian radial basis kernel function is adopted in this paper due to its good generalization properties in modeling analysis. Moreover, the second term in (6) can be optimized via minimizing cross-entropy loss (CE) between labels and predictions:

$$\arg\min_{\theta}\Big(\sum_{i=1}^{N}\sum_{k=1}^{K} -\log(p_k)\delta(y_i = l_k) + \alpha \parallel\theta\parallel\Big) \tag{8}$$

where p_k is the k-th conditional probability that the model outputs, l_k is the label of the k-th class from the set L, α is the trade-off regularization weight and δ is the indicator function.

As a consequence, the joint training objective is given by:

$$\begin{aligned} l_{\mathrm{total}} &= \lambda_1 \times l_{\mathrm{CE}} + \lambda_2 \times l_{\mathrm{REG}} \\ &= \lambda_1 \times \Big(\sum_{i=1}^{N}\sum_{k=1}^{K} -\log(p_k)\delta(y_i = l_k) + \alpha \parallel\theta\parallel\Big) \\ &\quad +\lambda_2 \times \sum_{i=1}^{5}(D_{\mathrm{MMD}}(q_i||p_i)) \end{aligned} \tag{9}$$

where l_{CE} and l_{REG} are the loss function for prediction accuracy and regularization on latent space, q_i and p_i denote $q_\phi(z_i|x)$ and $p(z_i)$, $\lambda_i (i = 1, 2)$ is a set of scaling parameters for the tradeoff between the recognition accuracy and divergence constraint, respectively.

In detail, we employ Adam optimizer with a batch size of 256 and learning rate of 10^{-2} for Bonn dataset, 128 and 10^{-3} for TUSZ and CHB-MIT dataset, respectively. For the hyper-parameters, we empirically set $\lambda_1 = 1$, $\lambda_2 = 0.1$ for Bonn dataset and $\lambda_1 = 0.1$, $\lambda_2 = 4$ for both TUSZ and CHB-MIT dataset, since Bonn dataset is merely slightly imbalanced while TUSZ and CHB-MIT dataset are severely imbalanced. Considering the difference in sampling rate and sampling indexes among Bonn, TUSZ and CHB-MIT datasets, maxpooling layers in the CE-stSENet* are removed for TUSZ and CHB-MIT dataset. Our model is implemented with the Pytorch framework [16]. In training stage, we additionally adopt *WeightedRandomSampler* function in *sampler* module to balance the volume of each class.

III. Experimental Results and Discussions

A. EEG Dataset and Preprocessing

The effectiveness of the proposed CE-stSENet* is evaluated on three public epilepsy EEG datasets.

TABLE I
TOTAL COUNTS OF DIFFERENT TYPES OF SEIZURES IN TUSZ DATASET

Seizure type	Count
Focal non-specific seizure (FN)	988
Generalized non-specific seizure (GN)	413
Simple partial seizure (SP)	44
Complex partial seizure (CP)	342
Absence seizure (AB)	76
Tonic seizure (TN)	67
Tonic clonic seizure (TC)	50
Myoclonic seizure (MY)	3

1) Bonn Dataset: Bonn dataset [23], which was obtained from the University of Bonn in Germany, consists of five subsets denoted as A, B, C, D and E. All EEG signals were acquired from the same 128-channel amplifier at a sampling rate of 173.61 Hz. Each subset contains 100 single-channel EEG segments of 23.6s duration (4096 sampling indexes). Concretely, subset A and B were recorded extracranially from five healthy volunteers with eyes open and closed respectively, while set C, D and E were collected intracranially from five epileptic patients. Signals in subset C and D were recorded from hemisphere hippocampal formation and the epileptogenic zone respectively when the patients were in seizure-free intervals, while Set E only contains intracranial EEG signals that correspond to seizure attacks. Note that the part of experimental paradigm in Bonn dataset, such as single-channel EEGs with labels assigned to long-time EEG segments, was replaced by multi-channel recordings and framewise detection [9]. Original EEG signals have already been denoised with a band-pass filter of 0.53–40 Hz, therefore only detrend procedure is performed before EEG signals are fed into the CE-stSENet*.

2) TUSZ Dataset: Temple University Hospital EEG Seizure Corpus (TUSZ) [24], which is a portion of Temple University Hospital EEG Corpus, is one of the world's largest publicly available corpus of annotated data for epileptic EEG recognition. TUSZ dataset contains a rich variety of seizure morphologies, which protects the corpus from bias and thus makes it an extremely challenging task for machine learning systems. Table I shows a summary of the distribution of seizures in TUSZ dataset in terms of different seizure types. We discard seizure events that last less than 4s since in subsequent experiments the length of sliding window is empirically set to be 4s. The preprocessing procedure of EEG data includes three steps. Firstly, considering the difference in sampling rate inter-patients, we resample EEG recordings to 128 Hz. Then all of EEG recordings are transformed into 20 common channels. The 20 common channels are listed in the Supplementary Materials A. Finally, we employ baseline removal and detrend operations.

3) CHB-MIT Dataset: CHB-MIT dataset [25], which was collected at Children's Hospital Boston, consists of long-duration multi-channel EEG recordings from 23 pediatric patients with intractable seizures. Recordings at a sampling rate of 256 Hz are grouped into 24 cases, each of which contains continuous EEG signals from a single patient. In this study, EEG recordings which contain 21 common channels and at least one seizure event have been considered. Details about CHB-MIT dataset are described in Table II. We fail to read some channel data for patients 6, 12 and 16, thus these three patients have been removed [26]. The preprocessing step includes baseline removal, detrend and a lowpass filter (0-64 Hz) for denoising and artifacts removal.

TABLE II
DATA INFORMATION OF CHB-MIT DATASET

ID-Gender -Age	Total .edf files with seizures	Seizure events (T_{min}-T_{max}) in seconds	Total seizure time(s)	Total seizure-free time(s)
1-F-11	7	7 (27-101)	442	23483
2-M-11	3	3 (9-82)	172	7987
3-F-14	7	7 (47-69)	402	24798
4-M-22	2	2 (49-111)	160	38097
5-F-7	5	5 (96-120)	558	17442
7-F-14.5	3	3 (86-143)	325	32212
8-M-3.5	5	5 (134-264)	919	17081
9-F-10	3	4 (62-79)	276	34223
10-M-3	7	7 (35-89)	447	50017
11-F-12	3	3 (22-752)	806	9253
13-F-3	7	10 (17-70)	440	24760
14-F-9	7	8 (14-41)	169	25053
15-F-16	14	20 (31-205)	1992	48442
17-F-12	3	3 (88-115)	293	10531
18-F-18	6	6 (30-68)	317	19957
19-F-19	3	3 (77-81)	236	10310
20-F-6	6	8 (29-49)	294	19742
21-F-13	4	4 (12-81)	199	13591
22-F-9	3	3 (58-74)	204	10596
23-F-6	3	7 (20-113)	424	31830
24-NR-NR	12	15 (16-70)	511	42689

M and F represent male and female patients respectively, T_{min} and T_{max} in denote the minimum and maximum seizure duration of each patient.

For TUSZ and CHB-MIT dataset, continuous EEG recordings are pre-organized into half-overlapping 4s epochs, corresponding to 512 and 1024 sampling indexes respectively, before fed into the CE-stSENet*.

B. Experimental Settings

In order to verify the effectiveness and robustness of the proposed CE-stSENet*, we conduct several experiments in the context of seizure event and onset detection respectively.

1) Epileptic Seizure Event Detection: For Bonn dataset, three kinds of different cases are designed as follows. In Case I, sets A, B, C, and D are grouped together as the normal class whereas the set E is considered as seizure class. Case II, corresponding to sets A and B versus C and D versus E, is related to discriminate healthy subjects, inter-ictal epilepsy patients and ictal ones. Finally, Case III, set A versus B versus C versus D versus E, is further considered to demonstrate the effectiveness of the proposed CE-stSENet*, since it is the most challenging classification task based on Bonn dataset. As a result, Case I, II and III correspond to classifying the EEG signals into two, three and five categories respectively.

In order to further evaluate the generalizability of the proposed CE-stSENet*, we investigate Case IV on TUSZ dataset, which aims at discriminating different kinds of seizure types. Among all the seizure types summarized in Table I, we only ignore Myoclonic seizure (MY) owing to its minimal occurrence frequency, thus forming a seven-class classification

TABLE III
THE OVERALL COMPARISON OF CLASSIFICATION PERFORMANCE ON BONN AND TUSZ DATASET

Method	Bonn Dataset						TUSZ Dataset	
	Case I		Case II		Case III		Case IV	
	F1	ACC(%)	F1	ACC(%)	F1	ACC(%)	F1	ACC(%)
EMD+SVM	0.9656±0.0051	97.80±0.40	0.9282±0.0353	93.20±3.31	0.7104±0.0461	72.00±3.95	0.8685±0.0087	85.69±0.73
CWT+SVM	0.9746±0.0134	98.40±0.80	0.9507±0.0169	94.60±1.62	0.7602±0.0161	76.20±1.72	0.9190±0.0070	90.89±0.35
ResNet18	0.9610±0.0065	98.44±0.27	0.9732±0.0039	97.20±0.36	0.8438±0.0120	84.72±1.18	0.7536±0.0151	79.58±0.68
Deep ConvNet	0.9505±0.0287	96.80±1.94	0.8901±0.0532	87.80±5.15	0.8154±0.0651	81.60±6.47	0.8940±0.0041	88.03±0.68
CE-stSENet*	**0.9969±0.0000**	**99.80±0.00**	**0.9930±0.0011**	**99.36±0.08**	**0.9459±0.0078**	**94.60±0.78**	**0.9369±0.0033**	**92.00±0.15**

where bold values indicate best results.

case. Additionally, each seizure event is classified into a type by taking a vote of the predictions by each epoch generated from this seizure event [27].

In order to obtain an unbiased evaluation of the classification performance, five-fold cross validation is adopted. Means and standard deviations of F1 score (F1) and accuracy (ACC) are recorded for the performance evaluation.

2) Epileptic Seizure Onset Detection: To facilitate the timely epileptic seizure alarm, real-time seizure onset detection, rather than event detection, is needed. Therefore, a continuous epileptic EEG database, namely CHB- MIT dataset, is used for the patient-specific latency study. Chang *et al.* [28] conducted experiments on CHB-MIT dataset and found that adaptively selected four to six channels were good enough for epileptic seizure detection. Therefore, considering the tradeoff between the classification performance and computational complexity, top 5 channels that correspond to the highest variance between ictal and inter-ictal classes in training set are selected. Selected 5 channels in each loop for each patient are given in the Supplementary Materials B. Additionally, to solve the class imbalance problem, in each iteration we randomly select 1000 inter-ictal epochs and all of ictal epochs for training, where pre-ictal epochs, defined as inter-ictal epochs with the same number of ictal epochs before seizure onset marker, are contained.

For the performance evaluation, we apply leave-one-record-out cross validation (LOOCV). Specifically, for each patient, supposing that there are N EEG recordings (.edf file) that contain seizures, in each fold, one EEG recording is used for testing while other N-1 recordings are used as the training set. Compared with k-fold cross-validation, the LOOCV not only ensures that the testing procedure covers all the seizures and alleviates the overfitting problem but also is more similar to the real clinical application. The sensitivity (SEN), specificity (SPE), accuracy (ACC), latency and sensitivity by onsets are adopted for the performance evaluation, where the latency is the time delay between the beginning of the seizure detected by the algorithm and the seizure onset marked by an expert, while the sensitivity by onsets is the percentage of seizures correctly detected in the total number of seizures. We define a seizure event as at least three consecutive seizure epochs detected by the algorithm during the calculation of latency and sensitivity by onsets.

C. Experimental Results on Seizure Event Detection

1) Overall Comparison: With aforementioned four classification cases in seizure event detection, the performance of the proposed CE-stSENet* is evaluated among approaches as follows:

1) The empirical mode decomposition (EMD) based temporal and spectral analysis is a commonly adopted baseline method to recognize epileptic EEGs and a support vector machine (SVM) classifier is adopted for the classification [12].
2) The continuous wavelet transform (CWT) combined with the gray level co-occurrence matrix (GLCM) descriptor is a widely used EEG feature extraction method. A SVM classifier is used for the classification [29].
3) The ResNet18 is a popular deep learning model that has achieved immense success in computer vision and biomedical signal processing [18]. We retrain the first and final layer for recognizing epileptic EEG signals.
4) Deep ConvNet is one of state-of-the-art deep learning models designed for the EEG decoding [6].

Table III gives the classification performance derived by the EMD+ SVM, CWT+SVM, ResNet18, Deep ConvNet and the proposed CE-stSENet* respectively. For Bonn dataset, the proposed method can reach a maximal classification accuracy of 99.80% in Case I due to the significant discrimination between ictal EEG segments and normal ones. The CE-stSENet* can also distinguish ictal, interictal and healthy EEG segments effectively with an accuracy of 99.36% in Case II. Furthermore, experimental results on Case III with a classification accuracy of 94.60% is even more encouraging considering the subtle differences among five categories. Meanwhile, for the TUSZ dataset, the proposed CE-stSENet* can still recognize different types of epileptic seizures efficiently with an accuracy of 92%. Compared with conventional approaches, our proposed method can gain a maximum increase of 3.00%, 11.56%, 22.6% and 12.42% on Case I, II, III and IV respectively. These experimental results indeed demonstrate the capability of the proposed CE-stSENet* for epileptic seizure event detection effectively.

2) Classification Performance Depending on Spectral-Temporal Analysis: Based on overall classification performance above, the proposed method can handle seizure event detection cases effectively in the presence of the spectral-temporal analysis. In this section, we further exploit the efficacy of the spectral-temporal analysis by comparing the CE-stSENet* with two simpler architectures as follows:

1) The CE-tSENet* is constructed by removing wave-Conv layers and corresponding spectral feature maps in the ST-SENet subnetwork. Therefore, it is a mix-scale

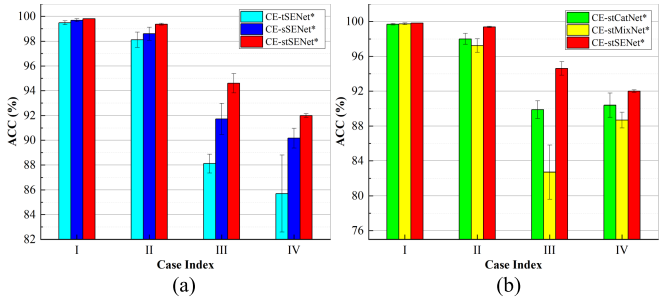

Fig. 4. Comparison of classification performance by (a) spectral-temporal analysis. (b) gcSE unit.

TABLE IV
ACCURACY COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT PARAMETER r

parameter r	ACC(%)
2	94.0
4	94.8
8	**96.0**
16	93.6
32	95.0

where bold fonts indicate the best results.

model without abilities of capturing robust spectral features.

2) The CE-sSENet* is implemented by removing temporal feature maps in the ST-SENet subnetwork, which can only extract spectral representations while cannot learn morphology ones.

The classification performance is shown in Fig. 4(a) with the average accuracy and standard error given. We can observe that the proposed CE-stSENet* outperforms other two baseline descriptors in all classification cases, which demonstrates the effectiveness of the spectral-temporal analysis intuitively. By inserting waveConv layers into a conventional CNN structure, the CE-sSENet* can gain higher accuracy than the CE-tSENet* ranging from 0.20% to 4.47%. However, due to the negligible discrimination involved in Case III and IV, it is not sufficient to learn spectral representations solely. Therefore, classification accuracies of the proposed CE-stSENet* further increase 2.88% and 1.83% against the CE-sSENet* due to the employment of the spectral-temporal analysis in Case III and IV respectively. These results indicate that spectral-temporal analysis can indeed improve the recognition performance of epileptic EEGs.

3) Efficacy of gcSE Unit: Apart from the spectral-temporal analysis, the gcSE unit can also affect classification results. As for the parameter r in the gcSE unit, which indicates the bottleneck in self-gate mechanism, it is set to 16 for image classification [15], while we conduct experiments with different values of $r = 2, 4, 8, 16, 32$ to search for its optimal setting for epileptic EEG recognition via the cross validation. We select Case III for the evaluation purpose. Table IV shows results of the proposed CEstSENet* with different parameter r, which indicates that the CE-stSENet* can achieve higher classification accuracy when r is set to 8.

In order to further evaluate the effectiveness of the gcSE unit in multi-domain feature extraction, two baseline methods are designed as follows:

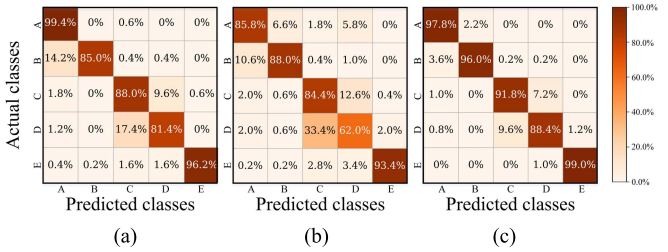

Fig. 5. Confusion matrixes on Case III. (a) CE-stCatNet*, (b) CE-stMixNet*, (c) CE-stSENet*.

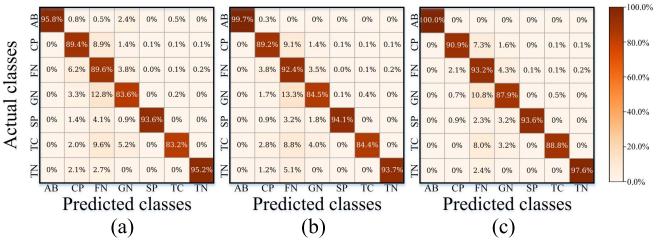

Fig. 6. Confusion matrixes on Case IV. (a) CE-stCatNet*, (b) CE-stMixNet*, (c) CE-stSENet*.

1) The CE-stCatNet* explores spectral and temporal representations independently, namely no SE block is implemented.

2) The CE-stMixNet* mixes multi-domain representations, namely the gcSE unit is replaced by regular convolution, batch normalization and ELU operations.

In Fig. 4(b), improvements of the classification accuracy manifest that the gcSE unit indeed helps to overcome the heterogeneity of multi-domain features. From Fig. 4(b), maximum accuracy increases of 0.12%, 2.12%, 11.88% and 3.31% are achieved in four cases respectively. Moreover, we exploit the efficiency of the gcSE unit for two specific classification cases, Case III and IV, by confusion matrixes and experimental results are shown in Fig. 5 and Fig. 6, respectively. Comparing between Fig. 5(a) and (b), we can see that classification accuracies of set A and set D are obviously improved by 13.6% and 19.4% respectively with the implement of group convolution operators. However, merely paralleling multi-domain feature extraction is not sufficient for a challenging task as Case III. Therefore, from Fig. 5(a) and (c), the SE block, which boosts discriminative multi-domain features jointly in a highly class-specific manner, contributes to the effectiveness of the proposed gcSE unit. Similarly, in Fig. 6, accuracies of all classes except SP are improved when applying the gcSE unit.

D. Experimental Results on Seizure Onset Detection

1) Overall Performance: Patient-specific confusion matrixes and evaluated performance parameters on CHB-MIT dataset are listed in Table V and Table VI respectively. With LOOCV, the average of patient-specific classification sensitivity, specificity, and accuracy achieve 92.41%, 96.05% and 95.96% respectively, which indicates the effectiveness of the proposed method on managing multi-channel continuous dataset.

TABLE V
PATIENT-SPECIFIC CONFUSION MATRIX FOR PROPOSED METHOD ON CHB-MIT DATASET

ID	Truly detected seizure (TP)	Misclassified seizure (FN)	Truly detected non-seizure (TN)	Misclassified non-seizure (FP)
1	210	2	11543	183
2	77	5	3879	107
3	189	3	12167	216
4	77	0	8956	2927
5	265	8	8651	57
7	144	15	15702	397
8	406	48	8035	493
9	130	3	16651	451
10	202	13	24352	639
11	374	26	4585	34
13	160	48	11573	786
14	65	11	12168	327
15	933	36	23587	592
17	135	8	5044	215
18	124	28	9371	591
19	104	10	4953	194
20	117	19	9573	281
21	85	10	6738	46
22	95	4	5262	27
23	198	5	15693	209
24	185	16	18177	1371

TABLE VI
PATIENT-SPECIFIC EVALUATED PARAMETERS FOR PROPOSED METHOD ON CHB-MIT DATASET

ID	Parameters for seizure events detection			Parameters for seizure onsets detection	
	SEN(%)	SPE(%)	ACC(%)	Average latency(s)	Sensitivity by onsets(%)
1	99.06	98.44	98.45	0.00	100
2	93.90	97.32	97.25	0.67	100
3	98.44	98.26	98.26	0.00	100
4	100.00	75.37	75.53	0.00	100
5	97.07	99.35	99.28	1.60	100
7	90.57	97.53	97.47	0.67	100
8	89.43	94.22	93.98	4.00	100
9	97.74	97.36	97.37	0.00	100
10	93.95	97.44	97.41	0.00	100
11	93.50	99.26	98.80	0.00	100
13	76.92	93.64	93.36	2.00	90.0
14	85.53	97.38	97.31	0.29	87.5
15	96.28	97.55	97.50	1.80	100
17	94.41	95.91	95.87	1.33	100
18	81.58	94.07	93.88	4.00	100
19	91.23	96.23	96.12	4.67	100
20	86.03	97.15	97.00	3.75	100
21	89.47	99.32	99.19	0.00	100
22	95.96	99.49	99.42	2.67	100
23	97.54	98.69	98.67	1.14	100
24	92.04	92.99	92.98	0.53	100
Aver	92.41	96.05	95.96	1.39	98.93

The proposed method performs with high specificity and satisfactory sensitivity for most of patients, especially for the patient 1 and 3 with all three evaluated parameters over 98%. The maximum sensitivity of 100% is achieved by the patient 4 while the highest specificity of 99.49% is reported for the patient 22, respectively. Moreover, our method reaches a maximum accuracy of 99.42% for the patient 22, which demonstrates the efficacy of our proposed method again. For the patient 4, the unsatisfying performance may be incurred by the relatively limited coverage of training data (merely two .edf files involved each loop).

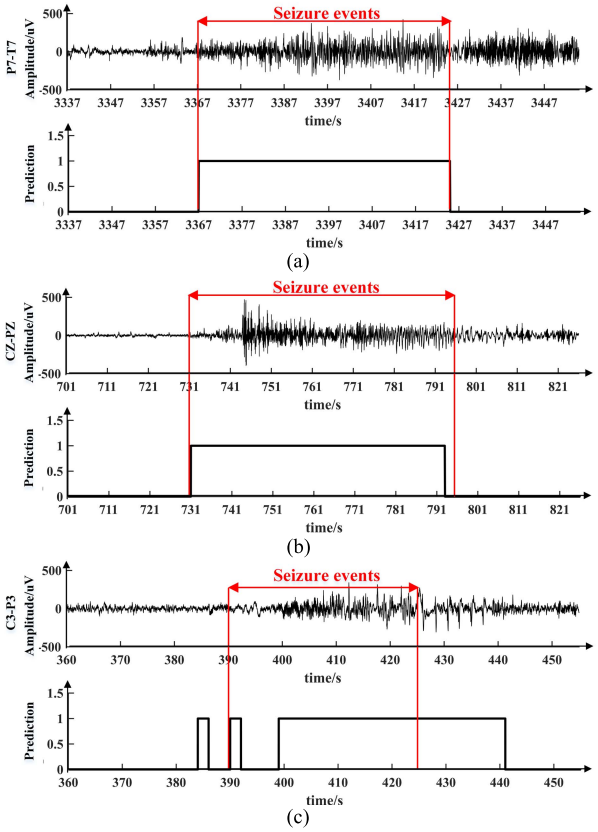

Fig. 7. Time plot of three segments of .edf file and corresponding predictions by the proposed method. (a) chb22_20, (b) chb03_02, (c) chb20_15.

Moreover, in seizure onset detection scenario, we evaluate latency and sensitivity by onsets and exhibit average latency and sensitivity by onsets in Table VI. Note that the average latency refers to an average of latencies of all the seizures in the corresponding patient. In detail, latencies for every seizure event of each patient are listed in the Supplementary Materials C. Specifically, zero latency is achieved by the patients 1, 3, 4, 9, 10, 11 and 21. All seizures except one seizure in patient 13 and 14 respectively have been detected. Additionally, plots of three typical classification results of the proposed method over time are shown in Fig. 7. The latency corresponding to Fig. 7(a), (b) and (c) is 0s, 0s and 8s respectively. As a result, the proposed CE-stSENet* achieves a negligible latency with a satisfactory detection rate of seizure onsets, which demonstrates its potential to handle seizure onset detection tasks intuitively.

2) Overfitting Resistance Depending on Maximum Mean Discrepancy-Based Information Maximizing Loss: Based on overall performance above, the proposed CE-stSENet* can serve as an automatic seizure onsets detection system effectively in the presence of maximum mean discrepancy-based information maximizing loss. In this section, we exploit the advantage of information maximizing loss in overfitting problem mitigation. Therefore, we evaluate the performance of the CE-stSENet coordinated with or without information

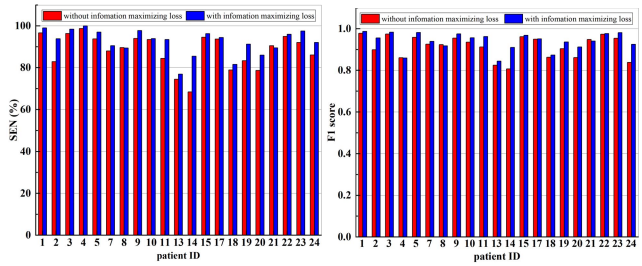

Fig. 8. Patient-specific parameters comparison of CE-stSENet without and with information maximizing loss. (a) Sensitivity by samples. (b) F1 score.

TABLE VII
COMPARISON OF PARAMETERS AND INFERENCE TIME

Model name	Parameters (million)	Inference time (10^{-2} ms)
Deep ConvNet	0.27	4.29
ResNet18	11.18	24.94
CE-stCatNet*	0.25	11.69
CE-stMixNet*	0.48	4.68
CE-stSENet*	0.29	10.13

maximizing loss, and show the sensitivity by samples and F1 score of each patient in Fig. 8(a) and (b) respectively. Firstly, we explore the efficacy of information maximizing loss in terms of sensitivity by samples, and experimental results are described in Fig. 8(a). Concretely, the sensitivity by samples of almost every patient has been boosted, with a maximum increase of 17.11% achieved by the patient 14. Furthermore, from Fig. 8(b), the F1 score has been constantly improved by using information maximizing loss for almost every patient. Specifically, the information maximizing loss obviously alleviates the overfitting problem, as a higher F1 score growth of 0.0867 and 0.1033 is gained by the patients 24 and 14 respectively. Consequently, the maximum mean discrepancy-based information maximizing loss combats overfitting problem at a large margin, which benefits the better potential clinical usage.

E. Discussion

1) Efficacy of Model Compactness: In order to further evaluate the generalizability and real-time performance of the proposed method, we compare the model size among the CE-stSENet* and existing methods and exhibit in Table VII. Since the real-time performance is one of the most important factors for the online seizure onset detection, we conduct experiments based on CHB-MIT dataset, which indicates the size of input EEGs is $5\times1\times1024$. Despite of the deeper architecture, the CE-stSENet* involves 2.9×10^5 parameters, which is of the same order as that of the Deep ConvNet. Moreover, the proposed method is substantially more compact than ResNet18. The significant compactness of the proposed CE-stSENet* is mainly relied on the adaptation of convolution operations with smaller filter size and the employment of group convolution. Therefore, we further compare the volume of learnable parameters in the CE-stCatNet*, CE-stMixNet* and CE-stSENet*, and list in Table VII. Note that the group convolution operator in the gcSE unit not only contributes to a promising classification performance but also reduces the model complexity by 47.92% against the CE-stMixNet*, while the SE block in the gcSE unit only increases overall network complexity by a very small fraction. Consequently, the CE-stSENet*, which holds a lightweight model structure, can theoretically handle epileptic seizure detection at a significant margin.

Moreover, we evaluate the inference time, defined as the required time for a deep learning model to make a decision for one EEG epoch, and list in Table VII. We perform this procedure on GeForce RTX 2080 GPU with 8 GB memory with a batch size of 256. Despite of the reduction of parameter volumes, group convolution operations cannot contribute to the mitigation of the computational complexity. However, the CE-stMixNet* gains the highest classification accuracy in terms of a tradeoff between the complexity and classification performance, thus it is concluded that the CE-stMixNet* is more competitive than most of the state-of-art deep learning architectures.

2) Performance Comparison of Various Methods for EEG Signal Recognition on Bonn and TUSZ Dataset: Table VIII presents a comparison between the CE-stSENet* and recent studies conducted on Bonn Dataset. Note that the proposed CE-stSENet* illustrates a better classification performance than most other existing methods. For example, as for the classification task sets A, B, C and D versus E, Riaz *et al.* [12] conducted the EMD for temporal and spectral feature extraction, a satisfactory recognition performance with an accuracy of 96% was achieved. Zhang *et al.* [2] employed the local mean decomposition (LMD) for temporal statistical and non-linear feature extraction. A classification accuracy of 98.87% was obtained by a genetic algorithm optimized SVM (GA-SVM). High-resolution time-frequency images were constructed by using multiscale radial basis function (MRBF) network with both a modified particle swarm optimization (MPSO) method and an orthogonal least squares (OLS) algorithm [4], and time-frequency features were extracted by GLCM texture descriptors. They obtained an accuracy of 99.53% with the SVM classifier. In comparison, an accuracy of 99.80% is achieved by our CE-stSENet* in the classification of ABCD-E. Additionally, for the three-class classification case, Riaz *et al.* [12] also classified Case AB-CD-E and provided an accuracy of 83%. Zhang *et al.* [2] conducted a similar case and achieved an accuracy of 98.40%. From results mentioned above, most existing approaches only exhibited a satisfactory performance in particular cases. By contrast, our presented approach is able to handle different cases on premise of promising classification accuracy. As for five-class case, Tzallas et al. [30] adopted time-frequency analysis with an artificial neural network (ANN) and achieved an accuracy of 89%. Furthermore, Türk *et al.* [31] constructed a CNN structure with two-dimensional time-frequency scalograms as inputs, which exploited spectral properties of EEGs efficiently. This approach reached to an accuracy of 93.60%. In comparison, our proposed CE- stSENet* exhibits a better classification performance with an accuracy of 94.60%.

TABLE VIII
EXPERIMENTAL SETTINGS AND PERFORMANCE RESULTS OF EXISTING METHODS ON BONN DATASET

Authors	Methods	Case I	Case II	Case III
Tzallas et al. [30]	Time frequency features	--	--	89
Riaz et.al [12]	Temporal and spectral features in EMD domain	96	83	--
Zhang et.al [2]	LMD based hybrid features with GA-SVM	98.87	98.40	--
Türk et.al [31]	Scalogram based convolutional neural network	--	--	93.60
Li et.al [4]	Multi-scale radial basis function networks and the fisher vector encoding	99.53	--	--
This work	**CE-stSENet***	**99.80**	**99.36**	**94.60**

where bold fonts indicate our proposed method.

TABLE IX
COMPARISON BETWEEN PROPOSED METHOD AND OTHER METHODS USING TUSZ DATASET

Authors	Methods	Seizure types	Training rate	Case IV	
				F1	ACC(%)
Saputro et al. [32]	MFCC +Hjorth +SVM	3	~57% (no CV)	--	91.4
Roy et al. [27]	FFT+KNN	7	80% (5-fold CV)	0.907	--
	FFT+SGD			0.778	--
	FFT+XGBoost			0.844	--
	FFT+Adaboost			0.707	--
	FFT+ResNet50			0.723	--
This work	**CE-stSENet***	**7**	80% (5-fold CV)	**0.9369**	**92.00**

where bold fonts indicate our proposed method.

Additionally, for TUSZ dataset, Table IX shows the comparison of seizure type classification results between the CE-stSENet* and other conventional methods. Currently, few studies have attempted to employ seizure type recognition. For example, Saputro et al. [32] employed Mel Frequency Cepstral Coefficients (MFCC) and Hjorth Descriptor for features extraction. A classification accuracy of 91.4% was achieved by using a cubic SVM classifier. However, only 3 types of seizure events were involved and no cross-validation was adopted in their experiment, which may lead to a bias consequently. Roy *et al.* [27] employed the Fast Fourier Transform (FFT) for the frequency feature extraction. Different classifiers were adopted with the best F1 score of 0.907 achieved by the KNN classifier. Their work is currently public on the website of TUH corpus. On the contrast, the proposed CE-stSENet* captures spectral-temporal features with informative ones boosted adaptively, which results in a F1 score of 0.9369 efficiently.

Consequently, the high classification results in all cases indicate that the proposed CE-stSENet* can theoretically mange epileptic seizure event detection tasks effectively.

3) Performance Comparison of Different Seizure Detection Methods Reported for CHB-MIT Dataset: Table X summarizes the results of some state-of-the-art methods (detecting the entire duration of seizures, not only the onsets) published on CHB-MIT dataset, where the NR stands for not reported values. It is difficult to draw direct comparison between the proposed method and existing studies due to the different experimental settings adopted by these methods, such as the number of selected channels and the number of involved patients. From Table X, existing studies commonly divide training and testing set based on a specific ratio for the evaluation of the classification performance, thus bias may still exist. Meanwhile, compared with the k-fold cross validation, LOOCV, which holds a whole EEG recording (.edf file) for testing in each loop, is more challenging and practical, since this procedure does not break the temporal continuity in testing set. From Table X, the average SEN, SPE and ACC of the proposed method are more promising to most of recent studies. The tradeoff between the sensitivity and specificity can be observed in [26], [33], [34], and [35], while the CE-stSENet* achieves more balancing results with three evaluated parameters all above 92%. Direct comparison to [1] is not feasible due to the synthetic minority oversampling technique adopted, which leads to an incomplete separation of training and test data. However, a conclusion can be drawn that the volume of labeled seizure epochs is a determining factor for seizure detection tasks. The transfer learning can be used to alleviate this problem [3], which leads to a 91.9%, 93.2% and 94.0% of SEN, SPE and ACC respectively. However, this method is restricted by the selection strategy of auxiliary patients, *i.e.*, data in source domain for the transfer learning. Limited empirical patients and a variety of seizure morphologies may result in a negative transfer and suboptimal performance consequently. Our method tackles this problem by using a deep learning network with a lightweight architecture and a joint training objective to combat skew data distribution, which contributes to our outstanding classification performance.

F. Limitations and Future Directions

Although the proposed framework achieves a promising recognition performance, two limitations still remain in current work. Firstly, although priori-selected channels reduce the computational complexity, the handcrafted selection of EEG electrodes may incur information loss and undermine the efficiency of our proposed method. Therefore, in future work, we will focus on integrating an adaptive channel selection mechanism into the proposed method to further improve the classification performance [9]. Secondly, the volume of training data is still limited by scarcity of seizure events in the present work. Although our proposed maximum mean discrepancy-based information maximizing loss mitigates the overfitting problem to some extent, just as claimed in [1], the amount of labeled ictal data is a crucial latency for EEG seizure detection. Therefore, we will consider adopting

TABLE X
EXPERIMENTAL SETTINGS AND PERFORMANCE RESULTS OF EXISTING METHODS ON CHB-MIT DATASET

Authors	Methods	Number of patients-channels	Training rate	Average SEN%-SPE%-ACC%
Rafiuddin et.al [33]	Two wavelet based and statistical based features	23-23	80% (no CV)	NR-NR-80.16
Khan et.al [34]	Relative energy and normalized coefficient of variation	5-NR	80% (no CV)	83.6-100-91.8
Kiranyaz et.al [26]	Time, frequency, time-frequency, non-linear feature, MFCC	21-18	25% (no CV)	89.0-94.7-NR
Zabihi et.al [35]	Seven intersection sequence features	23-23	50% (no CV)	89.1-94.8-94.6
Bhattacharyya et.al [1]	Three multivariate extension features of EWT	23-5	90% (10-fold CV)	97.9-99.6-99.4
Deng et.al [3]	ETTL-TSK-FS	23-23	0% (transfer)	91.9-93.2-94.0
This work	**CE-stSENet***	**21-5**	**LOOCV**	**92.41-96.05-95.96**

where bold fonts indicate the proposed method.

transfer learning approaches in our future work to fully explore the potential of the proposed CE-stSENet*.

IV. CONCLUSION

In this paper, a novel intelligent EEG recognition framework is proposed for automatic seizure detection by means of a CEstSENet aided by a maximum mean discrepancy-based information maximizing loss. Specifically, the CE-stSENet firstly capture fully spectral-temporal information from raw EEGs by the proposed waveConv layer. A variant of the SE block, termed gcSE unit, is further adopted to boost discriminative multi-domain representations unifiedly while suppressing redundant ones. Finally, the classification net is implemented for the epileptic EEGs recognition. Additionally, the proposed maximum mean discrepancy-based information maximizing loss alleviates the overfitting problem caused by the scarcity of seizure events in EEG recordings. The classification performance has been evaluated on three public seizure datasets with several independent classification cases, covering two prevalent clinical scenarios, *i.e.*, seizure event detection and seizure onset detection. Our proposed method produces promising results in terms of measurements SPE, SEN, ACC, latency and sensitivity by onsets. The efficacy of the spectral-temporal analysis and gcSE unit are also validated against two baseline methods respectively. The potential application of the proposed framework as an automatic seizure onset detection system also relies on maximum mean discrepancy-based information maximizing loss. Experimental results confirm that our proposed method can be regarded as an automatic seizure detection system, assisting seizure monitoring and alleviating the burden of clinicians.

REFERENCES

[1] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific eeg seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.

[2] T. Zhang and W. Chen, "LMD based features for the automatic seizure detection of EEG signals using SVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1100–1108, Aug. 2017.

[3] Z. Deng, P. Xu, L. Xie, K.-S. Choi, and S. Wang, "Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 8, pp. 1481–1494, Aug. 2018.

[4] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowl.-Based Syst.*, vol. 164, pp. 96–106, Jan. 2019.

[5] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: A report of four patients," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 603–615, May 2003.

[6] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[7] S. M. Azimi, P. Fischer, M. Korner, and P. Reinartz, "Aerial LaneNet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2920–2938, May 2019.

[8] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 178–190.

[9] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 83–94, Jan. 2019.

[10] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.

[11] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1149–1159, Jun. 2019.

[12] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "EMD-based temporal and spectral features for the classification of eeg signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.

[13] L. Wang *et al.*, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, May 2017.

[14] Z. Gao *et al.*, "EEG-based spatio–temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019.

[15] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[16] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.

[17] S. Zhao, J. Song, and S. Ermon, "InfoVAE: Information maximizing variational autoencoders," 2017, *arXiv:1706.02262*. [Online]. Available: http://arxiv.org/abs/1706.02262

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] G. Strang and T. Q. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA, USA: Wellesley-Cambridge, 1996.

[20] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, and Q. Guo, "Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 386–397, Mar. 2018.

[21] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 510–519.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: http://arxiv.org/abs/1312.6114

[23] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, pp. 116–126, Nov. 2001.

[24] V. Shah, "The Temple University hospital seizure detection corpus," *Frontiers Neuroinform.*, vol. 12, p. 83, Nov. 2018.
[25] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2009.
[26] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term electroencephalography," *J. Biomed. Informat.*, vol. 49, pp. 16–31, Jun. 2014.
[27] S. Roy, U. Asif, J. Tang, and S. Harrer, "Machine learning for seizure type classification: Setting the benchmark," 2019, *arXiv:1902.01012*. [Online]. Available: http://arxiv.org/abs/1902.01012
[28] N.-F. Chang, T.-C. Chen, C.-Y. Chiang, and L.-G. Chen, "Channel selection for epilepsy seizure prediction method based on machine learning," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 5162–5165.
[29] Y. Li, W. Cui, M. Luo, K. Li, and L. Wang, "Epileptic seizure detection based on time-frequency images of EEG signals using Gaussian mixture model and gray level co-occurrence matrix features," *Int. J. Neur. Syst.*, vol. 28, no. 7, Sep. 2018, Art. no. 1850003.
[30] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Comput. Intell. Neurosci.*, vol. 2007, pp. 1–13, Sep. 2007.
[31] Ö. Türk and M. S. Özerdem, "Epilepsy detection by using scalogram based convolutional neural network from EEG signals," *Brain Sci.*, vol. 9, no. 5, p. 115, May 2019.
[32] I. R. D. Saputro, N. D. Maryati, S. R. Solihati, I. Wijayanto, S. Hadiyoso, R. Patmasari, "Seizure type classification on EEG signal using support vector machine," *J. Phys., Conf. Ser.*, vol. 1201, no. 1, 2019, Art. no. 012065.
[33] N. Rafiuddin, Y. U. Khan, and O. Farooq, "Feature extraction and classification of EEG for automatic seizure detection," in *Proc. Int. Conf. Multimedia, Signal Process. Commun. Technol.*, Aligarh, India, 2011, pp. 184–187.
[34] Y. U. Khan, N. Rafiuddin, and O. Farooq, "Automated seizure detection in scalp EEG using multiple wavelet scales," in *Proc. IEEE Int. Conf. Signal Process., Comput. Control*, Jun. 2012, pp. 1–5.
[35] M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M. Gabbouj, and T. Ince, "Analysis of high-dimensional phase space via poincaré section for patient-specific seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 386–398, Mar. 2016.