Abstract—Freezing of gait (FoG) is one of the most common symptoms of Parkinson’s disease, which is a neurodegenerative disorder of the central nervous system impacting millions of people around the world. To address the pressing need to improve the quality of treatment for FoG, devising a computer-aided detection and quantification tool for FoG has been increasingly important. As a non-invasive technique for collecting motion patterns, the non-invasive technique for collecting motion patterns, the
footstep pressure sequences obtained from pressure sensitive gait mats provide a great opportunity for evaluating FoG in the clinic and potentially in the home environment. In this study, FoG detection is formulated as a sequential modelling task and a novel deep learning architecture, namely Adversarial Spatio-temporal Network (ASTN), is proposed to learn FoG patterns across multiple levels. ASTN introduces a novel adversarial training scheme with a multi-level subject discriminator to obtain subject-independent FoG representations, which helps to reduce the over-fitting risk due to the high inter-subject variance. As a result, robust FoG detection can be achieved for unseen subjects. The proposed scheme also sheds light on improving subject-level clinical studies from other scenarios as it can be integrated with many existing deep architectures. To the best of our knowledge, this is one of the first studies of footstep pressure-based FoG detection and the approach of utilizing ASTN is the first deep neural network architecture in pursuit of subject-independent representations. In our experiments on 393 trials collected from 21 subjects, the proposed ASTN achieved an AUC 0.85, clearly outperforming conventional learning methods.

Index Terms—Adversarial learning, deep learning, freezing of gait detection, footstep pressure, parkinson’s disease.

I. INTRODUCTION

MILLIONS of people around the world are impacted by the Parkinson’s disease (PD) which is a neurodegenerative disease predominantly characterized by its effects on the motor system [1]. Freezing of gait (FoG) is one of the most common PD symptoms, identified by its sudden and brief episodes of cessation of movement despite the intention of a patient to keep walking [2], [3]. With the progression of the disease, FoG happens more and more frequently, becoming a major risk factor for falls [4], [5] and eventually affects the mobility, independence and quality of life of a PD patient and their family [6]. Accurate detection and quantification of FoG are of great importance in clinical practice and could be used for assessing the impact of a treatment [7]. However, current manual annotation of FoG events relies heavily on subjective scoring by well-trained experts, which is extremely time-consuming. Therefore, computer aided intelligent solutions are required for timely, precise and objective FoG detection.
There has been ground-breaking success of deep learning techniques for many tasks such as object detection and human action recognition [20], [21], which provides a great opportunity for developing deep learning based methods to accurately characterize FoG patterns from footstep pressure sequences. In this scheme, pressure mat based FoG detection can be formulated as a sequence to sequence task where a pressure mat sequence is mapped into a FoG output sequence through temporal neural networks such as recurrent neural networks (RNNs). However, considering that the variations among all subjects (inter-subject variance) could be higher than the variance between FoG and non-FoG events (inter-class variance), there could be the potential issues to directly adopt these deep architectures as they were devised for general tasks often involving high inter-class variance. For subject-level clinical applications including FoG detection, such large inter-subject variance of the feature vector (representation) could negatively impact the model performance when applying the trained model to unseen subjects due to the potential of over-fitting risk on subject specified patterns. Thus, it is critical to devise proper mechanisms to reduce the inter-subject variance for deriving subject-independent FoG representations.

In this study, a novel end-to-end deep architecture, namely Adversarial Spatio-Temporal Network (ASTN), is proposed to address the FoG detection problem by utilizing the footstep pressure sequence at three levels. As FoG exhibits both the short-term intrinsic property and the long-term dynamic characteristic, ASTN introduces temporal convolutions and RNN cells together to formulate multi-level temporal FoG patterns from the sequentially organized footstep pressure data. In addition, by introducing an adversarial training scheme with a multi-level discriminator, ASTN is able to reduce the inter-subject variance and formulate subject-independent FoG representations. With the proposed mechanisms, it is expected that the subject-independent representation can be obtained for effective FoG detection.

In summary, the major contributions of this article are four-fold:

- We propose pressure mat based FoG detection as a sequential modelling task, which is one of the first studies utilizing footstep pressure data for FoG detection.
- A novel end-to-end deep architecture ASTN with an adversarial training scheme is proposed to characterise subject-independent FoG representations by reducing the inter-subject variation.
- The proposed architecture sheds light on improving other subject-level clinical modelling tasks, as it is available to integrate the adversarial training scheme with many deep learning architectures.
- A large footstep pressure sequence dataset was created during clinical assessments of 21 subjects to evaluate the effectiveness of our proposed methods.

The rest of the paper is organized as follows. Section II reviews the related works on FoG detection, footstep pressure data analysis, and spatio-temporal deep learning techniques. Section III introduces the details of our proposed method. Section IV presents comprehensive experimental results to
evaluate the effectiveness of our proposed footstep pressure based FoG detection method. Lastly, Section V concludes our study with discussions on our future work.

II. RELATED WORK

In this section, the related studies are reviewed from three aspects. Firstly, existing freezing of gait detection methods are discussed in terms of traditional wearable sensor systems and vision cameras. Secondly, as our method utilized footstep pressure sequence data, key methods of existing research on footstep pressure data analysis are reviewed. Lastly, considering that footstep pressure sequences contain both spatial and temporal patterns, we focus on the potential deep learning based techniques to deal with the combined modalities by the spatio-temporal data.

A. Freezing of Gait Detection for Parkinson’s Disease

In general, existing studies related to automatic FoG detection are based on two major approaches: wearable sensor-based methods and vision-based methods. For traditional sensor-based FoG detection methods [8], [9], [10], [11], [12], [13], [22], sensor systems are designed similar to general gait analysis methods (e.g. [23]) including accelerometers, gyroscopes or by pairing these two categories of sensors in an inertial measurement unit to capture temporal signals. Machine learning methods such as support vector machine, random forest, and deep learning are further utilized to analyze and detect the FoG patterns. Wearable sensor-based methods have gained promising performance for FoG detection with a single or more on-body sensors. But optimal configurations are yet to be validated across centres [24] and placing wearable sensors correctly on a patient's body generally need suitable training and experiences.

Recently, to address the limitations of these traditional sensor-based methods, vision-based methods for FoG detection [14], [15] have been proposed by taking advantage of its natural and non-contact characteristics [25]. These studies collect videos from clinical assessments during which the subject follows instructions to participate in a timed up and go (TUG) gait test. Such TUG tasks assess the functional mobility of a person in a standardized fashion [26]. Deep learning architectures are devised to characterize and detect fine-grained FoG patterns in these studies due to their great success in solving many vision related tasks. Although these studies have achieved promising results, there may be potential issues when applying them in practical environments which are generally more complex compared with the experimental protocols used during clinical data collection. For example, these studies work for cameras with fixed placements and require well-lit conditions, without shadows and occlusions. Additionally, although there might be privacy concerns with video-based methods, privacy-preserving learning for general action recognition has been recently explored [27].

Therefore, footstep pressure sequence data can be attractive for monitoring FoG events in both clinical and home care settings. Recent studies [16], [17], [18], [19] adopted insole pressure data for FoG detection. Particularly, an LSTM (Long Short-Term Memory) based deep learning method was devised to formulate the temporal patterns [17]. In our study, we focus on a different approach by obtaining the footstep pressure sequence data from a pressure mat where both pressure data and gait patterns (e.g., foot locations) are captured. Moreover, the pressure mat data provides the chance for multimodal learning [28] with data collected using other modalities such as videos in pursuit of robust unobtrusive FoG detection.

B. Footstep Pressure Data Analysis

In addition to analyzing the gait of patients for clinical purposes [29], [30], footstep pressure (sequence) data has been used for machine learning applications such as biometrics [31], [32], [33]. Most of them follow a traditional pattern recognition pipeline: extracting hand-crafted features first and then feeding these features into machine learning models such as generalized linear models (GLM), support vector machines (SVM), ensemble learning methods, and hidden Markov models (HMM) to produce prediction outputs. However, few evidences demonstrate that these existing hand-crafted features are well associated with FoG detection.

Very recently, deep learning has been introduced to deal with footstep pressure sequence data. By treating a footstep pressure sequence as 2-dimensional image series, a two-stream architecture is devised to process the spatial and the temporal streams of a footstep pressure sequence [34]. The pipeline is similar to general deep learning models taking image inputs, whilst the spatial information of a footstep sequence at a specified temporal index contains a one-channel heat map instead of 3 channels as in general RGB images. A simplified version of the ResNet-50 is adopted for each stream to learn spatial and temporal patterns, respectively. Note that this study treats the temporal dimension of an input sequence as an image channel, thus the approach may not characterize temporal patterns well.

C. Deep Learning for Spatio-Temporal Data

A footstep pressure sequence contains both spatial and temporal patterns, which can be viewed as a 2-dimensional grey image sequence or a 3-dimensional vision cube. Deep learning architectures devised for the spatio-temporal data have demonstrated great success across a wide range of tasks including action recognition and object detection and tracking. Initially to process 2-dimensional spatial sequences, single-stream [35] and two-stream methods [36] are proposed. The single stream based methods utilize pre-trained 2-dimensional convolution filters frame by frame and a number of temporal fusion strategies are being further investigated. The two-stream based methods take advantage of the appearance and the optical flow features obtained by 2-dimensional convolutions to form spatio-temporal representations.

Based on these pioneering studies, three major types of deep learning methods are currently utilized to process spatio-temporal inputs: convolution neural network (CNN) based methods, RNN based methods and two-stream based methods. The first type in general extends the 2-dimensional CNN architecture to a 3-dimensional counterpart by which the convolution filters are extended to filter 3-dimensional inputs.
such as C3D [37], [38], P3D [39] and I3D [40]. By considering the input as a 2-dimensional spatial sequence, the second type aims to model the temporal structure with RNNs [41], [42], [43], which have been widely used to model sequential patterns. In particular, LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) networks are proposed to address the gradient vanishing and exploding issues of the original RNN by using gate mechanism [44], [45]. GRU involves less computations than LSTM while achieving comparable performance to LSTM, thus improving the efficiency of RNNs. Moreover, GRU has demonstrated better performance on smaller datasets [46]. The last type which is based on the pioneering two-stream approach represents video content with both appearance and motion features [47].

The methods mentioned above mainly address the general spatio-temporal related problems, which involve significant inter-class variations. However, intra-class variations could be dominant in many applications such as FoG detection. Hence, fine-grained architectures are proposed for the applications with subtle inter-class variations. In summary, there are two major approaches of these architectures. Firstly, the patch-based methods (e.g. [48], [49] and our recent one [15]) have been proposed to model the subtle variations by utilizing regional or patch based patterns to characterize inter-class variations. The patch-based approaches are able to further analyze the localized key patterns. However, for footstep pressure sequences, it is not straightforward to identify such patches, considering the data modality as shown in Fig. 1. Secondly, a bilinear pooling approach is proposed to eliminate the need for region related prior knowledge [50]. Nonetheless, the bilinear pooling increases the model complexity, whilst footstep pressure sequence data is sensitive to complex architectures and the FoG detection performance could be compromised by directly following this approach.

III. PROPOSED METHOD

As shown in Fig. 2, our novel adversarial spatial-temporal network (ASTN) consists of two key components: a three-level spatial-temporal feature extractor and a multi-level subject discriminator. Firstly, for each trial, three-level spatio-temporal representations are obtained, including spatial representation, intrinsic FoG representation and dynamic temporal representation. Secondly, a multi-level subject discriminator is introduced with an adversarial training scheme to determine whether the trial representations contain any patterns associated with subject specific distributions. In this section, we first introduce footstep pressure sequence data and its spatial and temporal representations, FoG classifier, and multi-level subject discriminator, followed by adversarial training scheme.

A. Footstep Pressure Sequence Data

Footstep pressure sequences were obtained during the clinical assessment for each subject with a number of trials. Pressure data of each trial was collected from a set of sensors at a specified sampling rate. In total, \( W \times H \) sensors recorded the pressure levels of each frame (i.e., sample) along the temporal indices. By taking the locations of these sensors into consideration, the signals collected from them can be viewed as a 2-dimensional pressure heat map for each frame. A matrix \( X_{t,p}^{m,n} \in \mathbb{R}^{W \times H} \) is used to represent this pressure map, where \( t, p, m \) and \( n \) indicates temporal index (second level), frame index, subject index and trial index, respectively. Specifically, a pair \((t, p)\) indicates the \( p\)-th frame within the \( t\)-th second. The matrix can be viewed as a grey scale image. Hence, a footstep pressure map sequence \( \{X_{t,p}^{m,n}, t = 1, \ldots, T^{m,n}, p = 1, \ldots, P\} \) is collected to illustrate a trial of the clinical assessment, where \( T^{m,n} \) indicates the total length of the trial \( n \) of subject \( m \) and \( P \) indicates the sampling rate per second. Note that the sequence contains both spatial and temporal patterns. In addition, we detect FoG events for each temporal segment and denote \( y_{t}^{m,n} \in \{0, 1\} \) as the binary response to be associated with each temporal index of the spatial pressure map sequence. It indicates whether FoG occurs within the temporal window of index \( t \) or not (i.e., 1 for FoG event if at least one frame is annotated as FoG or otherwise 0 for non-FoG event).
B. Spatial Representation

Given a temporal index $t$ and a frame index $p$, a footstep pressure map is derived from a grid of sensors. Intuitively, a pressure heat map can be interpreted as a grey image and deep learning techniques such as CNNs can be used for learning representation of pressure data based FoG patterns.

Various CNN based methods have been proposed for image representation through a series of convolution filters, pooling layers and activation functions. In general, the recent studies of CNNs tend to construct deeper architectures, which are able to learn complex representations for improved classification or recognition performance. However, footstep pressure maps are quite different from general images: only a small area of sensors precept pressure levels higher than zero and a limited range of pressure levels for each position are obtained to produce a pressure map. In addition, for image related tasks, transfer learning has been widely utilized to fine-tune pre-trained deep networks, whilst it is required to train from scratch for footstep pressure maps due to its unique modality.

Therefore, instead of training a sufficiently deep neural network such as ResNet or DenseNet, AlexNet is utilized to learn spatial patterns from footstep pressure heat maps. Note that the ReLU activation functions of the original AlexNet are altered to the leaky version of ReLU to help transfer gradient across layers during the forward and backward propagations without a hitch:

$$\text{LeakyReLU}(x) = \begin{cases} 
    x & \text{if } x \geq 0, \\
    wx & \text{otherwise,}
\end{cases}$$  

where $w$ is the negative slope controlling the extent of layer patterns to be kept when $x < 0$. By denoting the function of obtaining spatial representation as $g(x)$, for each spatial input $X_{t,p}^{m,n}$, the spatial representation can be written as:

$$s_{t,p}^{m,n} = g(X_{t,p}^{m,n}),$$  

where $s_{t,p}^{m,n} \in \mathbb{R}^S$ and $S$ indicates the dimension of the spatial representation.

C. Temporal Representation

As FoG is a kind of movement related symptoms, temporal patterns are critical as well for precise characterization and identification of FoG events. To characterize temporal patterns, in this study, a two-level strategy was adopted. Firstly, the short-term intrinsic FoG patterns were learned to characterize common FoG characteristics, which were independent of each sequence. Secondly, the global sequential FoG patterns were further learned to characterize FoG dynamics of a particular sequence.

To capture the intrinsic FoG patterns, non-overlapped sliding temporal windows, of which the duration was one second, were applied to each footstep pressure sequence. Within each temporal window, a sub-sequence of $P$ frames was derived and further applied with temporal convolutions. Since the convolution filters are invariant across all sub-sequences obtained for each temporal window, they were expected to learn intrinsic FoG patterns independent of each sequence. In detail, denote

$$s_t^{m,n} = (s_{t,1}^{m,n}, \ldots, s_{t,P}^{m,n})^T \in \mathbb{R}^{P \times S},$$  

to represent the spatial representations within the $t$-th temporal window. Note that $s_t^{m,n}$ can be viewed as a 1-dimensional temporal sequence of $S$ channels. Hence, a series of temporal convolution filters, pooling layers and activation functions can be applied to compute $s_t^{m,n}$. By defining the computation as $h(x)$, the intrinsic temporal representation of each second can be summarized as:

$$g_t^{m,n} = h(s_t^{m,n}),$$  

where $g_t^{m,n} \in \mathbb{R}^{H_1}$ and $H_1$ is the dimension of the intrinsic temporal representation.

In general, FoG patterns vary in these sequences in terms of severity and duration, and it is necessary to characterize such sequence-level dynamic patterns for accurate FoG detection. To this end, RNNs are employed to model such sequential relations by memorizing proper historical states. Recently, gated mechanism introduced to RNNs such as the LSTM and the GRU architecture alleviates the gradient vanishing and exploding issues by controlling the extent to keep patterns from the historical states and to update patterns from the current state. In this study, GRU was utilized to characterize the dynamic temporal patterns due to its lower model complexity compared with LSTM. The computations of GRU take the intrinsic temporal representations of an entire trial $\{g_t^{m,n}, t = 1, \ldots, T^{m,n}\}$ as the input. In detail, the dynamic temporal patterns can be computed within GRU cells as:

$$z_t^{m,n} := \sigma(W_{xz}g_t^{m,n} + W_{zh}g_t^{m,n} + b_z),$$  

$$r_t^{m,n} := \sigma(W_{xr}g_t^{m,n} + W_{hr}g_{t-1}^{m,n} + b_r),$$  

$$h_t^{m,n} := \tanh(W_{xh}g_t^{m,n} + r_t^{m,n} \odot W_{hh}g_t^{m,n} + b_h),$$  

$$\tilde{g}_t^{m,n} := (1 - z_t^{m,n})h_t^{m,n} + z_t^{m,n}g_{t-1}^{m,n},$$  

where $z_t^{m,n}$ and $r_t^{m,n}$ are the reset gate and update gate, respectively; $b_z$, $b_r$ and $b_h$ are the bias terms; $h_t^{m,n}$ is the hidden state of the cell $t$; $\tilde{g}_t^{m,n} \in \mathbb{R}^{H_2}$ is the dynamic temporal representation, of which the dimension is $H_2$. By introducing the gated mechanism, GRU was able to keep the track of proper dependencies between footstep pressure maps across the sequence efficiently.

Note that the computations of (5)–(8) illustrate a forward GRU architecture, of which each state only depends on its predecessors. Therefore, the forward architecture can be naturally deployed for the online FoG prediction purpose. Furthermore, additionally involving a backward GRU architecture to summarize future patterns and concatenating with the patterns of historical states, a bi-directional GRU is adopted for fully modelling sequential FoG patterns.

D. FoG Classifier

The spatial-temporal representation including $s_t^{m,n}$, $g_t^{m,n}$ and $\tilde{g}_t^{m,n}$ has been derived in line with the above discussions. Denote a function $G$ to summarize all computations to generate this
spatio-temporal representation:
\[ \{ s_{t}^{m,n}, \tilde{s}_{t}^{m,n}, \bar{s}_{t}^{m,n} \} = G(\{ X_{t,p}^{m,n} \}). \quad (9) \]

Note that the dynamic temporal pattern \( \tilde{g}_{t}^{m,n} \) is based on the computations of the spatial pattern \( s_{t}^{m,n} \) and the intrinsic pattern \( \bar{s}_{t}^{m,n} \), and thus contains the other two levels of patterns implicitly. Hence, it is reasonable to construct a classifier \( C \) by utilizing the dynamic temporal pattern straightforwardly to detect FoG events along with the temporal index \( t \). In detail, it contains fully connected (FC) layers with proper activation functions (LeakyReLU functions for all FC layers except for the last one which adopts a sigmoid function) to generate the predictions. We denote \( y_{t}^{m,n} \) as the estimation of \( \tilde{y}_{t}^{m,n} \).

As FoG detection can be viewed as a binary classification problem, a binary cross entropy loss function was applied for the optimization purpose:

\[
\min_{C,G} J_{c} = - \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{T-1} \frac{1}{m,n} \sum_{t=1}^{T-1} \log(y_{t}^{m,n}) + (1 - y_{t}^{m,n}) \log(1 - y_{t}^{m,n}),
\]

which optimizes the parameters of \( G \) and \( C \).

In general, optimizing this loss function is able to obtain proper parameters to extract the spatio-temporal representation for FoG detection. However, this representation could contain subject specific information existing in the training set, which could cause the over-fitting risk and negatively impact the detection performance when predicting on the subjects not seen during the training. Considering the inter-subject variance is non-negligible compared with the variance between FoG and non-FoG cases, it is necessary to reduce the inter-subject variance among the representations and to derive subject-independent FoG representations.

### E. Multi-Level Subject Discriminator

In this study, a subject discriminator was proposed to distinguish whether two given spatio-temporal representations are generated from the same subject distribution or not. It worked together with an adversarial training scheme to reduce the inter-subject variance of the spatio-temporal FoG representation and to treat the task as a fine-grained classification problem.

Basically, for two subjects \( m' \) and \( m'' \) and their trials \( n' \) and \( n'' \), the spatio-temporal representations can be summarized as \( G(\{ X_{t,p}^{m',n'} \}) = \{ s_{t}^{m',n'}, \tilde{s}_{t}^{m',n'}, \bar{s}_{t}^{m',n'} \} \) and \( G(\{ X_{t,p}^{m'',n''} \}) = \{ s_{t}^{m'',n''}, \tilde{s}_{t}^{m'',n''}, \bar{s}_{t}^{m'',n''} \} \), respectively. The discriminator \( D(G(\{ X_{t,p}^{m',n'} \}), G(\{ X_{t,p}^{m'',n''} \})) \) differentiates whether the two representations are from the same subject distribution (\( n' = n'' \)) or not (\( n' \neq n'' \)).

In detail, the second-order difference of the two representations are computed as illustrated in (11)–(13), where \((\cdot)^{2} \) represents element-wise square operator. The second-order difference can be viewed as a special case of the bilinear computation, which has been proven as an efficient approach for fine-grained classification. Hence, it is expected that the second-order difference can be beneficial for generating subject-independent representation.

#### Algorithm 1: End-to-End Adversarial ASTN Training.

**Input:** population of the footstep pressure sequences.

**Output:** parameters of \( G, C \) and \( D \).

1. **while** not stopping criteria **do**
2. **sample** four sequences for the current batch: \( \{ X_{t,p}^{m_{1},n_{1}}, X_{t,p}^{m_{2},n_{2}}, X_{t,p}^{m_{3},n_{3}} \} \) and \( X_{t,p}^{m_{4}} \).
3. **minimize** \( J_{C} \) (10) and update the parameters of \( G \) and \( C \) over the sampled sequences.
4. **compute** spatio-temporal representations \( G(\{ X_{t,p}^{m_{1},n_{1}} \}), G(\{ X_{t,p}^{m_{2},n_{2}} \}), G(\{ X_{t,p}^{m_{3},n_{3}} \}) \) and \( G(\{ X_{t,p}^{m_{4}} \}) \) using the updated \( G \).
5. **minimize** \( J_{D} \) (14) and update the parameters of \( D \) over the two pairs: \( G(\{ X_{t,p}^{m_{1},n_{1}} \}), G(\{ X_{t,p}^{m_{2},n_{2}} \}) \) and \( G(\{ X_{t,p}^{m_{3},n_{3}} \}) \), while the parameters of \( G \) and \( C \) are frozen.
6. **maximize** \( J_{A} \) (15) over the latter pair \( G(\{ X_{t,p}^{m_{3},n_{3}} \}) \) to update the parameters of \( G \), while keeping the parameters of \( D \) frozen.
7. **end while**

Note that the computations take the spatial, intrinsic temporal and dynamic temporal representations into account together, although the temporal patterns implicitly contain the spatial pattern. The idea is similar to the consideration of using identity block in ResNet, which alleviates gradient vanishing issues and helps adversarial training to transfer and update gradients from the early layers.

\[
\Delta s_{t}^{m',n',m'',n''} = (s_{t}^{m',n'} - s_{t}^{m'',n''})^{2}, \quad (11)
\]
\[
\Delta \tilde{s}_{t}^{m',n',m'',n''} = (\tilde{s}_{t}^{m',n'} - \tilde{s}_{t}^{m'',n''})^{2}, \quad (12)
\]
\[
\Delta \bar{s}_{t}^{m',n',m'',n''} = (\bar{s}_{t}^{m',n'} - \bar{s}_{t}^{m'',n''})^{2}. \quad (13)
\]

By concatenating the third two-order difference vectors in (11)–(13), a neural network based discriminator can be constructed, which consists of one FC layer and a sigmoid activation function for binary outputs (i.e., 0 for \( m' = m'' \) and 1 for \( m' \neq m'' \)). This simple architecture later helps to efficiently refine the parameters of the spatio-temporal representation generator \( G \) during the adversarial training. Equation (14) illustrates the binary cross entropy loss function for optimizing the discriminator. Note that the parameters of \( G \) is frozen during the optimization of \( J_{D} \):

\[
\min_{D} J_{D} = - \sum_{m''=m'} \log(D(G(\{ X_{t,p}^{m',n'} \}), G(\{ X_{t,p}^{m'',n''} \})))
- \sum_{m'' \neq m'} \log(1 - D(G(\{ X_{t,p}^{m',n'} \}), G(\{ X_{t,p}^{m'',n''} \})). \quad (14)
\]

### F. Adversarial Training

To let the FoG classifier and the subject discriminator work together, the adversarial training, which was originally proposed for generation tasks [51], was introduced so that the training...
was not limited to the two independent loss functions proposed as $J_C$ and $J_D$. During the adversarial training, the classifier continuously attempts to generate better FoG representations for improving the FoG classification performance related to the training data, whilst the subject discriminator was trained to become a better detector for correctly judging whether two representations are uniquely distributed or not. The equilibrium of this game is achieved when the classifier is able to detect FoG patterns accurately, and the subject discriminator is left to always randomly guess at 50% confidence for the representations. In detail, the adversarial training optimizes the loss function $J_A$ in (15), which combines (10) and (14) to play a minimax game.

$$\max_G J_A = - \sum_{m' \neq m''} \log(1 - D(G(X_{t,p}^{m'}, n'), G(X_{t,p}^{m'', n''})).$$  

Maximizing $J_A$ optimizes the parameters of $G$ to produce subject-independent spatio-temporal representations, which confuses the discriminator. Note that the parameters of the discriminator $D$ is frozen during the optimization of $J_A$. With the above discussions, now an end-to-end adversarial training scheme for the footstep pressure sequence can be derived. Algorithm 1 illustrates the key steps of our adversarial training.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Dataset

The FoG footstep pressure dataset used in this study consists of 21 subjects who participated in our clinical assessment. All the subjects were required to finish a number of TUG gait trials. In each trial, the subjects followed the instruction to walk within specified scenarios (i.e. either across an elevated plank that varied in width, or across a ground plank that varied in width) presented in virtual reality through an HTC VIVE head mounted display. In total, 393 trials were recorded by video and a Zeno pressure sensor gait mat (Protokinetics) simultaneously. Videos were collected for the purpose of FoG annotation by well-trained experts at the video frame level. To collect pressure sensing data, 500 $\times$ 50 sensors, which collected 10 levels of force response were placed in a Zeno gait mat, sampling at 120 Hz during the trial. In total, 16,716 seconds of pressure signals were obtained. For FoG detection, these pressure recordings were divided into 1-second long periods of non-overlapping temporal segments, and each segment was classified as either FoG or non-FoG class. If any frame of a segment was annotated as FoG frames according to the ground truth, the pressure data segment was labelled as FoG; otherwise, it was labelled as non-FoG. Among these segments, 3,816 seconds contained FoG patterns confirmed by the video data and were marked as FoG events, while the rest of the trials were denoted as non-FoG events. The event rate is 22.8%, which indicates the dataset is imbalanced as one would expect that FoG happens during the minority of the patient’s gait.

B. Experimental Settings

It is common that machine learning based clinical studies are devised in line with a number of trials collected from each subject. To optimize and evaluate these machine learning models, the trials are usually split into training and test partitions. Note that under this protocol the trials of one unique subject had the chance to appear in both the training and the test stages. Hence, it is possible to over-fit the potential high subject-variance and over-estimate the model performance for a dataset which is not large enough. Thus, it makes sense to follow a subject-level split manner, i.e., trials of one specified subject appear either in the training partition or in the test partition, which helps objectively estimate the effectiveness of the modelling for unseen subjects. Fig. 3(a) and (b) illustrate the sample split protocols of the subject-level and Fig. 3(c) and (d) are for the trial-level.

Therefore, in this study, three types of architectures are constructed to demonstrate the effectiveness of the proposed method. The details of the three architectures are as follows:

- **Subject-level model** is trained and tested following the subject split manner. The subject split protocol helps to verify the effectiveness of the modelling method for the assessment of unseen subjects. Note that this architecture does not contain the discriminator and it only optimizes the loss function $J_C$ defined in (10).

- **Trial-level model** shares the same architecture as the subject-level model. However, it follows the trial split protocol. As subject-level related patterns could be already revealed during the training stage, it is expected that optimistic performance can be obtained compared to the subject-level model.

- **Subject-level model with discriminator (ASTN)** and adversarial training scheme is introduced to reduce the intersubject variance of the spatio-temporal representation. This model is expected to be robust for unseen subjects and to gain better performance than the subject-level model. Moreover, forward and bi-directional counterparts are implemented for all these three architectures by altering the direction of the GRU part. To comprehensively evaluate the performance of the proposed method, for each architecture, 5 random

---

![Illustration of the two protocols for the training-test split. Each colored shape represents a trail and the color identifies a particular subject. For the subject-level split protocol, trials belonging to one subject either appear in the training or the test partition. For the trial-level split protocol, trials from one subject can appear in both the training and test partitions.](image-url)
TABLE I
FOG DETECTION PERFORMANCE OF OUR MODELS IN TERMS OF DIFFERENT METRICS

| Model                                      | AUC  | Youden’s J | Sens. | Spec. | FPR  | FNR  | LR+  | LR−   | Accuracy |
|--------------------------------------------|------|------------|-------|-------|------|------|------|-------|----------|
| **Subject-level**                          |      |            |       |       |      |      |      |       |          |
| Forward Model                              | 0.754| 0.40       | 68.6% | 71.7% | 28.3%| 31.4%| 2.42 | 0.44  | 70.6%    |
| Forward Model w/ Discriminator            | 0.780| 0.45       | 71.4% | 73.2% | 26.8%| 28.6%| 2.66 | 0.39  | 70.0%    |
| Bi-directional Model                      | 0.798| 0.47       | 76.5% | 70.6% | 29.4%| 23.5%| 2.60 | 0.33  | 71.8%    |
| Bi-directional Model w/ Discriminator     | **0.847**| **0.56**   | **83.4%**| 72.9%| 27.1%| 16.6%| **3.08**| **0.23**| **75.7%**|
| **Trial-level**                            |      |            |       |       |      |      |      |       |          |
| Forward Model                              | 0.769| 0.43       | 74.1% | 68.5% | 31.5%| 25.9%| 2.35 | 0.38  | 70.3%    |
| Bi-directional Model                      | 0.819| 0.50       | 73.9% | 76.0% | 24.0%| 26.1%| 3.08 | 0.34  | 75.2%    |

The bold values indicate the best metric among the methods in the table.

As expected, for both the forward and the bi-directional architectures, the performances of the subject-level models drop compared to those of the trial-level models in terms of AUC: 0.754 for the forward subject-level model vs. 0.769 for the forward trial-level model; 0.798 for the bi-directional subject-level model vs. 0.819 for the bi-directional trial-level model. These results confirm the assumptions that subject-level patterns could lead to an over-fitting issue in a trial-level model, in which a subject can appear in both training and test splits. Such settings can inflate the performance and reduce the reliability for unseen subjects who possess different subject-level patterns. The proposed discriminator and adversarial training strategy help to alleviate this issue and increase the performances of the subject-level models with all kinds of GRU architectures (forward and bi-directional). Particularly, with the adversarial training strategy, the performances of the subject-level models achieve superior performance even compared to the inflated performances of the trial-level models.

In particular, the sensitivity, specificity and accuracy related to a threshold $\theta$ are also adopted for evaluation, which maximizes the following Youden’s J statistics [52]:

$$\hat{\theta} = \arg \min_\theta \text{sensitivity} + \text{specificity} - 1.$$  

By maximizing the statistic, a threshold can be derived to treat sensitivity and specificity with equal importance. The evaluation metrics and associated J statistics are also listed in Table I. In terms of AUC, for the best model, sensitivity, specificity and accuracy values related to this threshold achieve 83.4%, 72.9% and 75.7%, respectively.

Table II shows the comparisons between ASTN and other spatio-temporal learning methods. Note that these methods gained promising performance for different tasks and datasets,
TABLE II
COMPARISON WITH THE STATE-OF-THE-ART METHODS AND ASTNs WITH DIFFERENT BACKBONES

| Methods                              | AUC    |
|--------------------------------------|--------|
| Conv + Bi-directional LSTM [17, 41]  | 0.543  |
| Conv + Bi-directional GRU [41]       | 0.798  |
| R3D [53]                             | Not converged |
| ViViT [54]                           | 0.590  |
| Conv-VTN [55]                        | 0.562  |
| ViT-VTN [55]                         | 0.584  |
| ViViT-VTN [54, 55]                   | 0.749  |
| ASTN (LSTM backbone)                 | 0.681  |
| ASTN (GRU backbone)                  | 0.847  |
| ASTN (Conv-VTN backbone)             | 0.567  |
| ASTN (ViViT-VTN backbone)            | 0.766  |

The bold values indicate the best metric among the methods in the table.

whilst none of them performed well for sequential pressure mat data. For example, LSTM based methods performed well for insole pressure based FoG detection [17], but are not directly transferable for pressure mat data. Considering that the ASTN learning scheme for spatio-temporal data can be adopted for different backbone networks, we also took these spatio-temporal methods jointly with ASTN for evaluation. It can be observed that our proposed ASTN can improve other backbones as well. Particularly, a ViViT-VTN method, which addresses the short-term spatio-temporal patterns with a ViViT network [54] and long-term temporal patterns with a temporal transformer [55], achieved an AUC 0.749. With our ASTN learning strategy, the AUC improved to 0.766. This observation suggests that between different modalities, additional efforts should be made for the backbone models, whilst the subject-level invariance can steadily help improve the performance.

As illustrated in Fig. 5(a)–(d), the FoG detection results of four selected trials shown in (a)–(d), which indicate the effects by introducing the adversarial training scheme. Compared with the ground truth data, both the forward and the bi-directional ASTNs characterize FoG events better than their counterparts without a subject discriminator involved. Fig. 6. Spatial latent feature $s_{m,n}^{(t)}$ PCA analysis. The temporal indices highlighted in red represent an FoG event.

D. Impact of Discriminator

Fig. 6. Spatial latent feature $s_{m,n}^{(t)}$ PCA analysis. The temporal indices highlighted in red represent an FoG event.

The discriminator determines whether any two representations belong to a unique subject distribution or not. It is expected that it would be increasingly challenging to distinguish the representations from distinct subject distributions along the adversarial training. Fig. 7 shows the AUC values of the discriminator ROC curves against the iteration steps. Along the iteration, both the discriminators of the forward and the bi-directional ASTNs tend to achieve an AUC around 0.5, which indicate that it is approximately random in discriminating subject distributions of the input pair with their spatio-temporal representations.
TABLE III

| Discriminator type                  | AUC  |
|-------------------------------------|------|
| Forward ASTN architecture           |      |
| First-order difference discriminator | 0.756|
| Absolute first-order difference discriminator | 0.776|
| Second-order difference discriminator | 0.780|
| Concatenated discriminator           | 0.760|
| Bi-directional ASTN architecture    |      |
| First-order difference discriminator | 0.800|
| Absolute first-order difference discriminator | 0.825|
| Second-order difference discriminator | 0.847|
| Concatenated discriminator           | 0.846|

The bold values indicate the best metric among the methods in the table.

further implies that the inter-subject variation of the spatio-temporal representation has been significantly reduced, which could be beneficial for the generalization of the classifier for undertaking FoG detection on unseen subjects. Moreover, it can be observed that the bi-directional architecture converges with fewer iterations than the forward counterpart. These findings suggest that the adversarial optimization of both forward and backward patterns increase the effectiveness to eliminate the subject associated redundancy.

Table III further presents the performance comparisons of different discriminator architectures. In addition to the second-order difference discriminator, the first-order difference discriminator, the absolute first-order difference discriminator and the general concatenated discriminator are involved for the comparison. It can be observed that the proposed second-order difference discriminator achieves the best performance in terms of AUC for both the forward and the bi-directional ASTNs. Note that most of these discriminators outperform the corresponding trial-level models, which indicates the necessity and effectiveness to integrate the discriminator with an adversarial training scheme for subject-level clinical studies.

Note that the discriminator adopts spatial, intrinsic temporal and dynamic temporal representations simultaneously, which helps to transfer and update gradients within the layers near the input layer. This multi-level representation design plays a similar role to the mechanism of ResNet and DenseNet. To verify this design, we implement the counterparts for both forward and bi-directional ASTNs with discriminators only involving the dynamic temporal representations, which is close to the output layer instead of the input layer. As shown in Table IV for the performance of these two strategies, the discriminator based on the multi-level representation outperform the one based on single dynamic temporal representation in terms of AUC.

Moreover, Fig. 8 illustrates the AUC of the bi-directional model with different adversarial scales.

Fig. 8. Illustration of the performance of the bi-directional model with different adversarial scales.

TABLE V

| Observing window duration | AUC   |
|---------------------------|-------|
| 0.5s                      | 0.840 |
| 1s                        | 0.847 |
| 1.5s                      | 0.845 |
| 2s                        | 0.836 |

The results suggest that the default setting is a reasonable choice for the FoG detection task.

E. Temporal Resolution

Our proposed ASTN is able to provide different temporal resolutions for FoG detection. In addition to the 1-second based detection, Table V lists the AUC for different temporal resolutions from 0.5 s–2 s. Generally, it is expected that an algorithm is able to differentiate FoG patterns in a more precise manner and action localization task has been studied for this purpose. Recent FoG detection methods started to explore this direction [56], which adopted both smooth term and the binary classification loss to formulate accurate boundaries. Considering the recent findings (e.g., [57]) in the action localization field, it is reasonable to study FoG events by formulating their starting and ending points with a spatio-temporal learning architecture as a future direction.

F. Analysis of Representation Distribution

To further understand the improvements of the representation space regularized by the adversarial training scheme with the multi-level subject discriminator, principal components analysis (PCA) is conducted on the three levels of representations individually, including the spatial representation, the intrinsic temporal representation and the dynamic temporal representation. As shown in Fig. 9, each row illustrates the representation at a specific level with and without the adversarial training.
scheme. For each training scheme, the first two components of the PCA results are visualized in a 2D space. To clearly present the PCA results, we use different markers for the data points in line with their ground truth labels and predicted labels: the red and the blue point markers denote the ground truth labels of the positive cases and the negative cases, respectively; the red and the blue circle markers denote the predicted positive cases and the predicted negative cases, respectively.

For the spatial representation space, the model with a discriminator tends to generate more separable data distribution for the two classes, whilst the model without a discriminator showed dispersed data points of each class over an extensive area. In terms of the predicted results, the separable data distribution helps to derive more discriminant regions to distinguish the positive cases from the negative cases. In contrast, without the discriminator, the regions for the two classes tend to be undifferentiated, which is the cause of the decreased performance of the subject-level model. Similar results can be observed from the intrinsic and dynamic temporal representation spaces. Thus, the adversarial training with the multi-level subject discriminator contributed to the robustness of the representation space, and thus the model performance was superior compared with others.

V. CONCLUSION

To investigate footstep pressure based FoG detection for the first time, this article presents a novel architecture ASTN to process the data modality of footstep pressure sequences. ASTN involves a classifier to learn the spatio-temporal FoG representation from pressure sequences and a subject discriminator to regularize the representation to be subject-independent using the adversarial training procedure. The subject-independent characteristic is critical for clinical studies to ensure the effectiveness of the modelling methods to be deployed for unseen subjects. Note that this adversarial training scheme for the subject-independent representation can be integrated with many existing neural network architectures, and thus may improve other subject-level clinical studies. Experimental results on a large in-house dataset highlights the superior performance of ASTNs compared to those analytic counterparts not exploiting the subject independent FoG representation. In our future research, we aim to further improve the spatio-temporal learning architecture to better FoG detection performance, as there is still a gap between our pressure mat based detection method and the methods using other sensors. It would also be helpful to apply the multi-level adversarial scheme to multi-modal FoG data to generate subject-independent cross-domain representations for more accurate characterization of FoG patterns.

ACKNOWLEDGMENT

We thank our patients who participated into the data collection and they all provided written informed consent. The ethical approval of this research was obtained from the University of Sydney Human Ethics Board (#2014/255). We would like to acknowledge and thank Kristin Economou for conducting and scoring the freezing of gait assessments.

REFERENCES

[1] E. R. Dorsey et al., “Global, regional, and national burden of Parkinson’s disease, 1990–2016: A systematic analysis for the global burden of disease study 2016,” *Lancet Neurol.*, vol. 17, no. 11, pp. 939–953, 2018.

[2] M. A. Hely, W. G. J. Reid, M. A. Adena, G. M. Halliday, and J. G. L. Morris, “The Sydney multicenter study of Parkinson’s disease: The inevitability of dementia at 20 years,” *Movement Disorder*, vol. 23, no. 6, pp. 837–844, 2008.

[3] M. Macht et al., “Predictors of freezing in Parkinson’s disease: A survey of 6,620 patients,” *Movement Disorder*, vol. 22, no. 7, pp. 953–956, 2007.

[4] B. R. Bloom, J. M. Hausdorff, J. E. Visser, and N. Giladi, “Falls and freezing of gait in Parkinson’s disease: A review of two interconnected, episodic phenomena,” *Movement Disorder*, vol. 19, no. 8, pp. 871–884, 2004.

[5] S. J. G. Lewis and R. A. Barker, “A pathophysiological model of freezing of gait in Parkinson’s disease,” *Parkinsonism Related Disorders*, vol. 15, no. 5, pp. 333–338, 2009.

[6] J. D. Schaafsma, Y. Balash, T. Gurevich, A. L. Bartels, J. M. Hausdorff, and N. Giladi, “Characterization of freezing of gait subtypes and the response of each to Levodopa in Parkinson’s disease,” *Eur. J. Neurol.*, vol. 10, no. 4, pp. 391–398, 2003.

[7] S. Donovan et al., “Lasershoe cues for gait freezing in Parkinson’s disease: An open-label study,” *Parkinsonism Related Disorders*, vol. 17, no. 4, pp. 240–245, 2011.

[8] M. Bachlin et al., “Wearable assistant for Parkinson’s disease patients with the freezing of gait symptom,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 436–446, Mar. 2010.

[9] E. E. Tripolity et al., “Automatic detection of freezing of gait events in patients with Parkinson’s disease,” *Comput. Methods Prog. Biomed.*, vol. 110, no. 1, pp. 12–26, 2013.

[10] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, “A deep learning approach to on-node sensor data analytics for mobile or wearable devices,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 56–64, Jan. 2017.

[11] D. Rodríguez-Martín et al., “Home detection of freezing of gait using support vector machines through a single waist-worn triaxial accelerometer,” *PLoS One*, vol. 12, no. 2, 2017, Art. no. e0171764.

[12] G. Prateek, I. Skog, M. E. McNeely, R. P. Duncan, G. M. Earhart, and A. Nehorai, “Modeling, detecting, and tracking freezing of gait in Parkinson disease using inertial sensors,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 10, pp. 2152–2161, Oct. 2018.

[13] J. E. Thorp, P. G. Adamszycz, H.-L. Ploeg, and K. A. Pickett, “Monitoring motor symptoms during activities of daily living in individuals with Parkinson’s disease,” *Front. Neurol.*, vol. 9, 2018, Art. no. 1036.
[14] K. Hu et al., “Vision-based freezing of gait detection with anatomic directed graph representation,” IEEE J. Biomed. Health Inform., vol. 24, no. 4, pp. 1215–1225, Apr. 2020.

[15] K. Hu et al., “Graph sequence recurrent neural network for vision-based freezing of gait detection,” IEEE Trans. Image Process., vol. 29, pp. 1890–1901, 2020.

[16] A.Marcante et al., “Foot pressure wearable sensors for freezing of gait detec-
tion in Parkinson’s disease,” Sensors, vol. 21, no. 1, 2020, Art. no. 128.

[17] G. Shalin, S. Pardoel, E. D. Lemaire, J. Nantel, and J. Kofman, “Prediction and detection of freezing of gait in Parkinson’s disease from plantar pres-
sure data using long short-term memory neural-networks,” J. Neuroeng. Rehabil., vol. 18, no. 1, pp. 1–15, 2021.

[18] S. Pardoel, G. Shalin, J. Nantel, E. D. Lemaire, and J. Kofman, “Early detection of freezing of gait during walking using inertial measurement unit and plantar pressure distribution data,” Sensors, vol. 21, no. 6, 2021, Art. no. 2246.

[19] S. Pardoel, J. Nantel, J. Kofman, and E. D. Lemaire, “Prediction of freezing of gait in Parkinson’s disease using unilateral and bilateral plantar-pressure data,” Front. Neurol., vol. 13, 2022, Art. no. 831063, doi: 10.3389/fneur.2022.831063.

[20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[21] B. Qian, Y. Wang, R. Hong, M. Wang, and L. Shao, “Diversifying inference path selection: Moving-mobile-network for landmark recognition,” IEEE Trans. Image Process., vol. 30, pp. 4984–4904, 2021.

[22] J. O’Day et al., “Assessing inertial measurement unit locations for freezing of gait detection and patient preference,” J. Neuroeng. Rehabil., vol. 19, no. 1, pp. 1–15, 2022.

[23] Q. Zou, L. Ni, Q. Wang, Q. Li, and S. Wang, “Robust gait recognition by integrating inertial and RGBD sensors,” IEEE Trans. Cybern., vol. 48, no. 4, pp. 1136–1150, Apr. 2018.

[24] S. T. Moore et al., “Autonomous identification of freezing of gait in Parkin-
son’s disease from lower-body segmental accelerometry,” J. Neuroeng. Rehabil., vol. 10, no. 1, pp. 1–11, 2013.

[25] Z. Zhang, J. Chen, Q. Wu, and L. Shao, “GHI representation-based cross-
view gait recognition by discriminative projection with list-wise con-
straints,” IEEE Trans. Cybern., vol. 48, no. 10, pp. 2935–2947, Oct. 2018.

[26] A. Shamway-Cook, S. Brauer, and M. Woollacott, “Predicting the prob-
ability for falls in community-dwelling older adults using the timed up &
go test,” Phys. Ther., vol. 80, no. 9, pp. 896–903, 2000.

[27] S. Pentyala, R. Dowsley, and M. De Cock, “Privacy-preserving video classification with convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., 2021, pp. 8487–8499.

[28] Y. Wang, “Survey on deep multi-modal data analytics: Collaboration, rivalry, and fusion,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 17, no. 1s, pp. 1–25, 2021.

[29] S.-H. Lee et al., “Gait performance and foot pressure distribution during wearable robot-assisted gait in elderly adults,” J. Neuroeng. Rehabil., vol. 14, no. 1, 2017, Art. no. 123.

[30] G. Guo, K. Guffey, W. Chen, and P. Pergami, “Classification of normal and pathological gait in young children based on foot pressure data,” Neuroinformatics, vol. 15, no. 1, pp. 13–24, 2017.

[31] O. Costilla-Reyes, P. Scully, and K. B. Ozanyan, “Temporal pattern recogni-
tion in gait activities recorded with a footprint imaging sensor system,” IEEE Sensors J., vol. 16, no. 24, pp. 8815–8822, Dec. 2016.

[32] R. Vera-Rodriguez, J. S. Mason, J. Fierrez, and J. Ortega-Garcia, “Compar-
ative analysis and fusion of spatiotemporal information for footstep recogni-
tion,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 4, pp. 823–834, Apr. 2013.

[33] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon, “A floor sensor system for gait recognition,” in Proc. IEEE Workshop Autom. Identification Adv. Technol., 2005, pp. 171–176.

[34] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, “Analysis of spatio-temporal representations for robust footstep recogni-
tion with deep residual neural networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 285–296, Feb. 2019.

[35] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-
Fei, “Large-scale video classification with convolutional neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1725–1732.