IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 September 2023, accepted 20 November 2023, date of publication 12 December 2023,
date of current version 29 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341418

RESEARCH ARTICLE

Symbolism and Directivity of Joint Keypoints in Temporal and Spatial Dimensions in Human Pose Prediction With GCN-Based Model

JINHUI LI, JIANYING HUANG, AND HOON KANG, (Member, IEEE)
School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 06974, Republic of Korea

Corresponding author: Hoon Kang (hkang@cau.ac.kr)

This work was supported in part by the Chung-Ang University Research Scholarship Grant, in 2022; and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant 2021R1A2C1009735.

ABSTRACT A wide variety of methods have been developed to predict the posture of the human body at a given point in time based on data on previous movements. More recently, prediction models based on deep learning have become a topic of active research and development. In this study, we adopt the strategy of separating spatial and temporal information based on an existing STGCN model to extract features effectively in both space and time, and we analyzed the effects of signed or unsigned and directed or undirected forecasts of the positions of human joints with this approach. We propose a method using an encoder based on a modified graph adjacency matrix in a graph convolutional network model and focus especially on the terms of the signs and directions of data on the locations of the joints in space and time. We also introduce a global residual block. The results of an experimental evaluation of our proposed method showed that we obtained better performance by applying the signed and directed features independently to the spatial and temporal adjacency matrices. The proposed model exhibited noticeable improvements in several aspects. In future research, we expect these features of the modified adjacency matrix to help learning models understand the correlation between symbols and directions for various actions and poses.

INDEX TERMS Graph convolutional networks (GCN), spatial temporal graph convolutional networks (STGCN), 3D datasets, human pose prediction, directional & symbolic method, human joints key points.

I. INTRODUCTION

Human pose prediction techniques have been developed to predict future poses based on previous movements. These methods have been adopted for various applications such as human-computer interaction, autonomous driving, and medical teleoperation. Conventional methods such as hidden Markov models (HMMs) [21] and Gaussian process dynamical models (GPDMs) [2] provide good results for simple actions, and deep learning models using recurrent neural networks (RNNs) [1], [3] have shown better results for complex actions. However, RNN models represent 3D data as vectors, and errors tend to accumulate with deeper sequential processing, which limits their effectiveness for pose prediction tasks; RNN models can overfit the input data.

In human posture prediction, interactions between joints and the feature information of each joint play a key role [7], and the adjacency matrix and feature matrix used by graph convolutional network (GCN) models can respectively contain the information on relationships between joints and the feature information of each joint well [7]. In this study, we proposed a new model based on a spatiotemporal GCN [5] to predict human poses from 3D datasets by separating spatial and temporal information [4] with a directional and symbolic adjacency matrix to extract spatial and temporal features followed by a temporal convolutional network (TCN) model to predict future poses. Given the nature of information on the positions of nodes representing human joints in space and time contained in a sequence of image frames [5], [7], we apply a spatiotemporal adjacency matrix to represent spatiotemporal features. To explore the interconnectivity between joints represented by the spatiotemporal

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang.

adjacency matrix, we performed directional [15] and symbolic [14] exploration of the information on relationships between the human joints contained in some benchmark datasets and experimentally evaluated our proposed directional and symbolic method. This method is inspired by the signed [14] and directed [15] approach mentioned by Sofianos et al. in their work on a space-time separable GCN (STS-GCN) model [4], and combined with the method of using separate spatial and temporal adjacency matrices, our proposed method can separate the spatiotemporal adjacency matrix into temporal and spatial adjacency matrices to resist spatiotemporal crosstalk well [4]. The initial intention of subdividing the spatiotemporal adjacency matrix into separate matrices for temporal and spatial dimensions is the same as that of adopting directionality and significance for temporal and spatial dimension adjacency matrices; both subdivide a higher-dimensional adjacency matrix into two lower-dimensional adjacency matrices to predict the positions of joint nodes representing the human body. We trained the proposed model using the Human3.6M [6] and AMASS datasets, and the results were verified using 3DPW. We found that the multidimensional correlation of human joints in time and space and the variation process of spatiotemporal adjacency matrices in the model were not clearly comprehensible in the mathematical forms provided in previous related works. Hence, we collated graph theories on this topic provided in the relevant literature [7], [10], [11], [20] the correlations between them along with related mathematical inference methods to describe the variation process of spatial-temporal adjacency matrices on the 3D dataset. The results of an experimental evaluation of our proposed model showed that it exhibited better performance on the Human3.6M and AMASS datasets and approached the performance of a space-time separable GCN on the 3DPW dataset.

II. RELATED WORK

Almost all prior research on predicting the posture of the human body has focused on representing poses using 3D coordinates, and many related works have encoded a representation of the degree of joint connectivity with GCN models [7], [8]. However, basic GCN models cannot represent spatiotemporal dimensions sufficiently with a basic adjacency matrix. Hence, to represent the information about the degree of joint connectivity of a human pose in spatial-temporal dimension, STGCN [5] adopted a spatiotemporal adjacency matrix to represent the positions of the joints in spatial and temporal dimensions and showed good performance. Because STGCN models fuse the temporal and spatial dimensions into a single spatiotemporal dimension represented by a spatiotemporal adjacency matrix, they cannot distinguish the features of poses in the two dimensions clearly. To solve this problem, Sofianos et al. proposed an STS-GCN [4] model to separate a spatial-temporal adjacency matrix into temporal and spatial adjacency matrices. Their results showed that subdividing a complex higher-dimensional problem into its constituent dimensions was effective in pose prediction tasks.

Pose prediction methods provided in the literature may be categorized as either relying on traditional algorithms or adopting new approaches based on deep learning. HMMs and GPDMs have exhibited better performance than other such methods for these tasks. HMMs are a Kalman filter-like method for describing a process state using discrete random variables. GPDMs use a Gaussian process to simulate time-series data to predict future poses.

Among methods based on deep learning, RNN and long short-term memory (LSTM) models have shown better performance on traditional time series prediction tasks. RNNs process data with temporal relationships by simultaneously considering information from all time steps in a given sequence and making predictions at the current time step based on previous information. However, for longer durations, data from the previous time steps may not convey sufficient information to make predictions accurately, which is known as the problem of long-term dependencies. LSTM models are a type of RNN designed to address this problem. CNN models have also been applied for pose prediction tasks [19], and TCN models have shown better performance than methods based on RNNs [16]. Based on these findings, the proposed approach is based on a simplified TCN model.

III. BACKGROUND

A. BACKGROUND ON GNNS

Many types of GNN [10] models based on graph theory have been developed. In this study, we consider GNNs in terms of recurrent graph neural networks (RecGNNs) [7], convolutional graph neural networks (ConvGNNs) [7], and spatiotemporal graph neural networks (STGNNs) [7] to illustrate GNNs briefly.

RecGNNs assume that the nodes within a graph continually exchange messages with a single neighbor until a balanced state is established. In contrast, ConvGNNs convert convolutional operations on grid data into convolutional operations on graph data, and STGNNs take both spatial and temporal relationships into account, as does the STGCN model used in the proposed approach. That is, graph convolution captures spatial relationships, whereas RNN and CNN models are used to handle temporal dependencies [5], [7].

B. BACKGROUND ON GCNS

ConvGNNs can be categorized as spectral [7], [11] or spatial [7], [11] methods. Graph convolutions in spectral approaches use filters [11] inspired by graph signal processing that use convolutions to remove noise from graph signals effectively.

Graph convolutions in spatial approaches adopt concepts from RecGNNs [7] to perform graph convolutions through the propagation of information. Specifically, similar to the conventional CNN convolution operation on images, the spatial approach defines the convolution of the graph based

on the spatial relation of the nodes. A grid image can be considered as a particular kind of graph specified with each pixel serving as a node and connected to its neighboring pixels. The convolution of a 3×3 filter on the image takes a weighted average of the values of the central node and that of its neighbors.

Similarly, the spatial graph convolution operation involves convolving the representation of a central node with that of its neighbors into an updated representation of the central node. This process effectively propagates information on the nodes through the edges [10].

GCN models have been successfully adopted in previously independent approaches based on spectral and spatial techniques [11] in which a graph convolutional operation is used to propagate information between nodes in the graph and extract features from the graph structure of data for non-Euclidean structures. Approaches based on spatial characteristics have been developed and widely adopted for their better performance. Compared to data representing Euclidean structures as used in conventional CNNs, GCN models retain structural information from non-Euclidean data [7].

From the following derivation of mathematical relations of GCN models given in Appendix B, we note that in methods based on spectral characteristics, GCN models can also be interpreted as spatial methods.

C. BACKGROUND ON STGCNS

STGNN models can be used to forecast the values of future nodes or predict spatiotemporal graph labels.

STGNN models may be roughly divided into two categories including methods based on RNN and CNN models [7]. STGCN models apply a CNN architecture for action recognition based on the positions of the joints and extend graph convolutional networks to spatiotemporal graph patterns by capturing both the spatial and temporal dependencies of a graph. The GCN model plays a crucial role both mathematically and theoretically in correlating spectral and spatial techniques.

Compared to GCN models that represent the local interactions of the graph data, STGCN models add a temporal dimension to model spatiotemporal interactions. Variations are reflected in an adjacency matrix, i.e., the adjacency matrix $\overline{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$ in GCN becomes A^{st} in an STGCN model on the basis of the temporal dimensions added to $\overline{A}$, Thus, the mathematical representation of the graphical convolution process in STGCN is given as follows.

$$H^{(l+1)} = \sigma(A^{st-l}X^l W^l) \tag{1}$$

IV. PROPOSED MODEL

The proposed model is inspired by the method provided by Sofianos et al. [4], in which spatial and temporal dimensions are separated as proposed by Sofianos et al., which splits an adjacency matrix fused into spatiotemporal dimensions into two adjacency matrices A^{t-l} in the time dimension and A^{s-l} in the spatial dimension.

In our proposed approach, an STS-GCN model is adopted to process data structured as a graph that contains spatial and temporal features. This type of data comprise two key aspects, including spatial and temporal relationships [5]. While traditional STGCN models mainly deal with the relationship between temporal and spatial dimensions, STS-GCN models solve this problem by separating the spatial and temporal aspects to limit crosstalk [4] between data in the temporal and spatial dimensions. This separation allows an STS-GCN to capture both spatial and temporal features to model spatial-temporal relationships more accurately.

Among methods to implement graph convolution, traditional GCN models perform graph convolution as given in Eq.(1). In the proposed STS-GCN, because the input data are multidimensional, i.e., $NCtV$ for N samples, an adjacency matrix is used to convolve the CtV dimensions by Einstein summation [4].

The process of Einstein summation differs for different adjacency matrices, i.e., the spatial-temporal and spatial-temporal separated adjacency matrices [4]. This process is described in detail below for the different adjacency matrices.

In the Einstein summation process $NCtV, WVtq \rightarrow NCqW$, $NCtV, Vtq \rightarrow NCqV$ and $NCtV, tVW \rightarrow NCtW$ for the space-time, time, and space dimensions, respectively.

The dimensions of $NCtV, NCqW, NCqV, NCtW$ are $[N^d, C^d, T^d, V^d]$, and the dimensions of $WVtq$ are $[V^d, V^d, T^d, T^d]$, the dimensionality of Vtq is $[V^d, T^d, T^d]$, and the dimensions of tVW are $[T^d, V^d, V^d]$. Each element $N_a C_b q_c W_d$ in the $NCqW$ matrix is obtained using Eq.(2).

$$N_a C_b q_c W_d = \sum_{i=1}^{T^d} \sum_{j=1}^{V^d} N_a C_b t_i V_j \times W_d V_j t_i q_c \tag{2}$$

Each element $N_a C_b q_c V_d$ in the $NCqV$ matrix is obtained by using Eq.(3) as given below.

$$N_a C_b q_c V_d = \sum_{i=1}^{T^d} N_a C_b t_i V_d \times V_d t_i q_c \tag{3}$$

Each element $N_a C_b t_c W_d$ in the $NCtW$ matrix is obtained by using Eq.(4).

$$N_a C_b t_c W_d = \sum_{j=1}^{V^d} N_a C_b t_c V_j \times t_c V_j W_d \tag{4}$$

where a takes values from 1 to N^d, b takes values from 1 to C^d, c takes values from 1 to T^d, and d takes values from 1 to V^d.

In this work, $[N^d, C^d, T^d, V^d] = [256, 3, 10, 22]$. It should be clear that the graph convolution process performed by Einstein summation can convolve the input features with the corresponding dimensions in the adjacency matrix more effectively even for an adjacency matrix comprising multiple dimensions because Einstein summation can represent modeling spatiotemporal adjacency matrices as well as modeling separated temporal and spatial adjacency matrices. Moreover, it is not limited to these three cases and can also be used to model higher-dimensional relationships. Thus, Einstein summation can be used more systematically in 3D or higher-dimensional modeling [4].

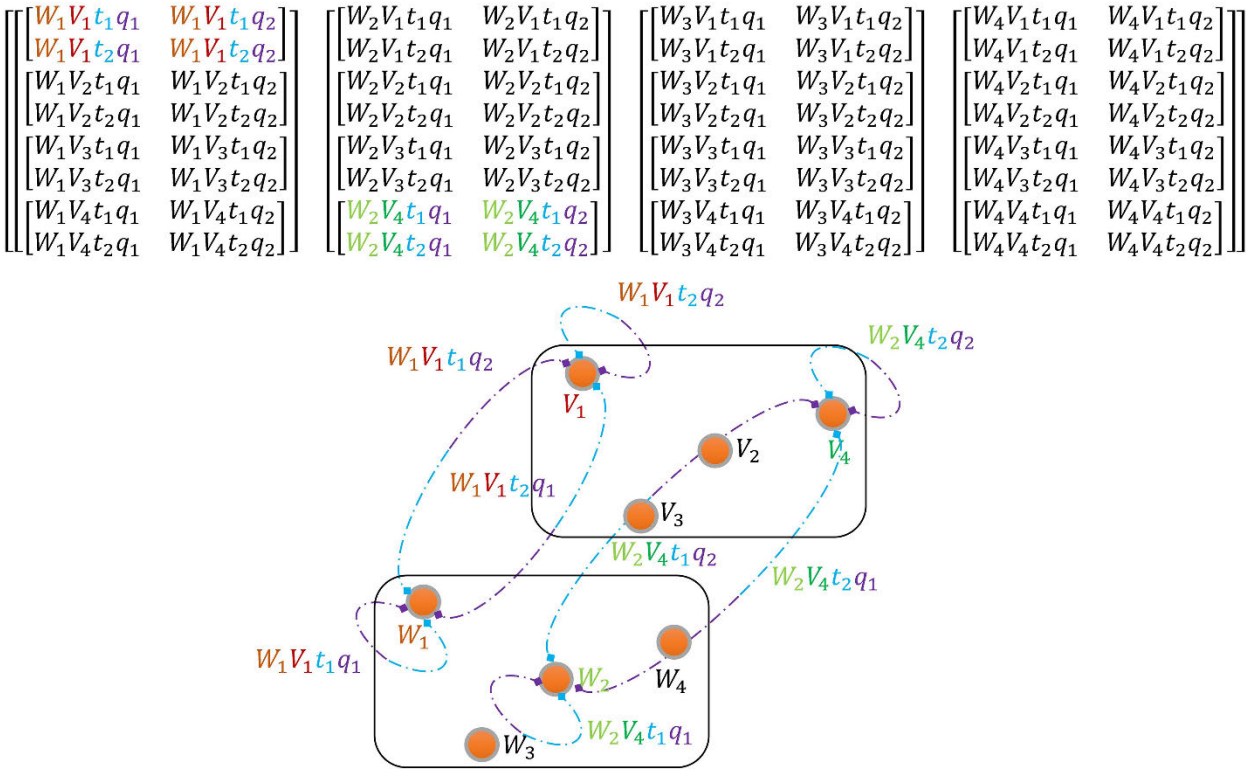


FIGURE 1. The spatial-temporal adjacency matrix in symbols and its representation on the joints of the figure.

In terms of the adjacency matrix, our proposed directional and symbolic method further explores the relationship between keypoints on a skeleton model representing the human body and an adjacency matrix in the two separated dimensions to limit spatiotemporal crosstalk between the temporal and spatial dimensions.

Because the dimensions of the input matrix $[N^d, C^d, T^d, V^d] = [256, 3, 10, 22]$ and the dimensions of the spatiotemporal adjacency matrix $[V^d, V^d, T^d, T^d] = [22, 22, 10, 10]$ are so large, providing an example of a figure to explain the implications of the spatial-temporal adjacency matrix on the human skeleton is relatively inconvenient. Therefore, we assume the dimensions as $[N^d, C^d, T^d, V^d] = [2, 3, 2, 4]$, $[V^d, V^d, T^d, T^d] = [4, 4, 2, 2]$ and explain the implication of the spatiotemporal adjacency matrix on the human skeleton as a method to deduce the relationship between the human skeleton and the spatial-temporal adjacency matrix in data with high dimensionality. Assuming $[N^d, C^d, T^d, V^d] = [2, 3, 2, 4]$, $[V^d, V^d, T^d, T^d] = [4, 4, 2, 2]$, taking the parts of W_1V_1 and W_2V_4 as examples, the spatial-temporal adjacency matrix and its representation on the joints of the figure are as shown in Fig.1.

In this method, the spatial-temporal adjacency matrix can represent not only the information associated with nodes at different times but also that between different nodes in time, i.e., the complex spatiotemporal information between nodes is contained in the spatial-temporal adjacency matrix.Assuming $[N^d, C^d, T^d, V^d] = [2, 3, 2, 4]$, $[V^d, V^d, T^d, T^d] = [4, 4, 2, 2]$, taking the part of V_1 as an example, the temporal adjacency matrix and its representation by symbols on the joints in the figure are shown in Fig 2.1 of Fig 2, and the spatial adjacency matrix and its representation by symbols on the joints in the figure are shown in Fig 2.2 of Fig 2, assuming $[N^d, C^d, T^d, V^d] = [2, 3, 2, 4]$, $[V^d, V^d, T^d, T^d] = [4, 4, 2, 2]$, and taking the part of V_1 as an example.

It may be observed from the figure that the temporal adjacency matrix can still indicate whether nodes are associated at different times, and their degree of association can be indicated by the temporal adjacency matrix A^t. However, in contrast to the spatiotemporal adjacency matrix A^{st}, A^t can only have association information on the same node at different times, whereas different nodes can be associated at different times in the A^{st} part, taking the part of W_2V_4 as an example.

For the spatial adjacency matrix, the nodes can only have associated information simultaneously, and the spatial adjacency matrix reflects the information on whether they are associated and their degree of association. At the level of the spatial adjacency matrix, this method of associating the information between nodes is similar to the method of associating information between the joints of a traditional GCN. It may be noted that modeling the adjacency matrix by separating spatial and temporal information can indeed restrain crosstalk between temporal and spatial information. This countermeasure to prevent information crosstalk [4], that is, modeling

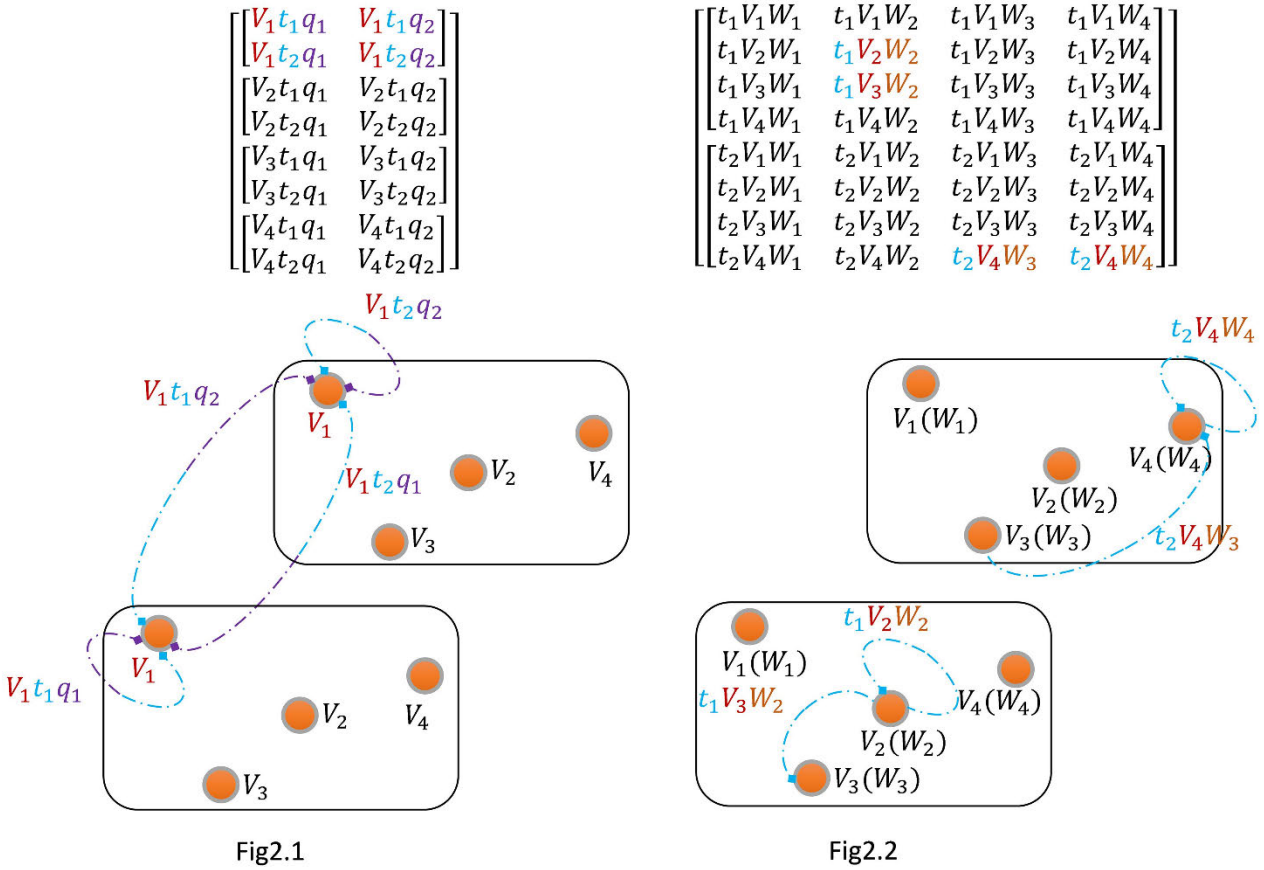


FIGURE 2. The temporal adjacency matrix and an equivalent representation by symbols on the joints are shown in Fig 2.1, and a similar illustration of the spatial adjacency matrix is shown in Fig 2.2.

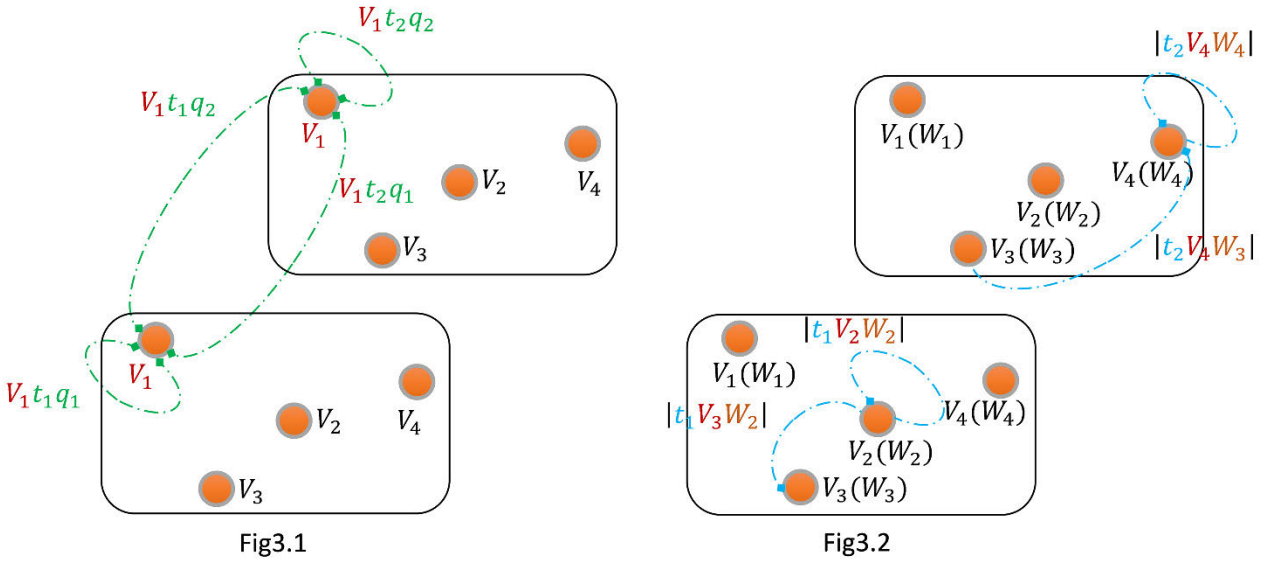


FIGURE 3. The signed and undirected temporal adjacency matrix in Fig 3.1 and the unsigned and directed spatial adjacency matrix in Fig 3.2.

the adjacency matrices sequentially by separation, can enable the model to learn information associated with joints more effectively.

Thus, our proposed method is designed to allow the model to learn information associated between joints more effectively.

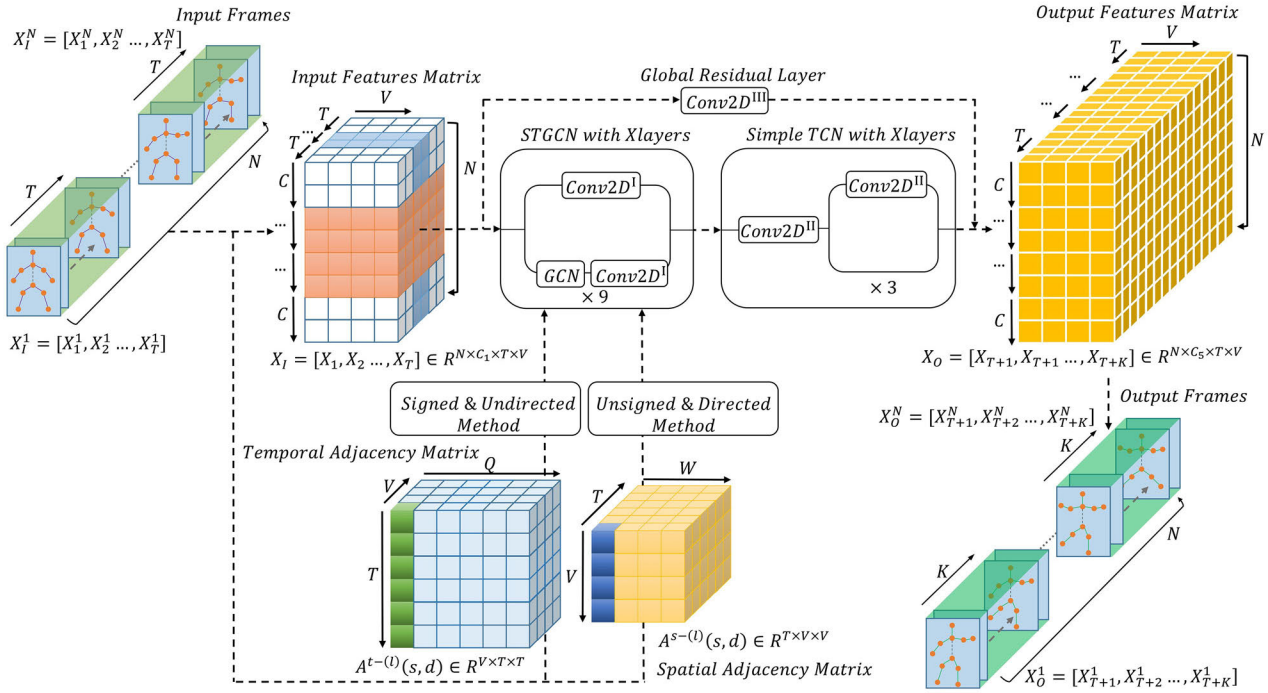


FIGURE 4. The overall pipeline of the proposed model.

Suppose that V_3 is a joint location on the arm and the feature information of V_3 is unchanged before and after a given moment t_2, and suppose V_4 is the location of a joint on the hand.

If the hand being modeled is performing a pushing motion and its associated information $t_2V_3W_4$ is assumed to be positive at the moment t_2 from V_3 to W_4, then, at the moment t_2, V_4 to V_3, its associated information $t_2V_4W_3$ are negative in the simplest case, and the value of $t_2V_3W_4$ is the same as the value of $t_2V_4W_3$. Moreover, the only difference is that the positivity and negativity are reversed. In contrast, if the hand being modeled is performing a pulling motion, based on these assumptions and this logic, the value of $t_2V_3W_4$ would be negative and $t_2V_4W_3$ would be positive.

Conversely, if we assume that $t_2V_3W_4$ is negative for a pushing motion, all the resulting positivity and negativity are the opposite of the result of those implied by these assumptions.

Given this complexity, for $t_2V_3W_4$ and $t_2V_4W_3$, although their values may differ owing to the complex association between the joints in complex space-time and the positivity and negativity of their values may be equivalent, it may be determined relatively easily that the positivity and negativity of the value of the degree of association between joints are the same even though the joints have a complex association in complex space-time. Even with such complex associations, the positive and negative values of the degree of association vary with the sequence of the joints, which means that the positive and negative values represent the same information as the sequential values; that is, by removing the positive and negative values and retaining only the sequential values, information between the joints can be retained without any distortion and redundancy can be eliminated. The benefit of removing the positive and negative values and retaining the sequential values is that overfitting resulting from redundancy is obviously reduced in the learning process.

After applying this idea to process the spatial adjacency matrix, we obtain unsigned and directed spatial adjacency matrices as shown in Fig 3.2.

For the time dimension, as an example, in Fig 3.1, after directionless processing, the symmetry of the data for $V_1t_1q_2$ and $V_1t_2q_1$ is maintained, which is represented in the figure as removed directionality, i.e., the joint-to-joint information is independent of the order of the starting and ending points in the elements of the matrix, except for the diagonal. We handle the data this way because the information between the joints in the previous frame and those of the next frame should intuitively give the same impression of taking the same size as the information between the joints of the next frame and the joints of the previous frame in the simplest case. The results were found to be meaningful and exceeded the original SOTA results.

As shown in Fig. 4, the proposed model consists of a signed and undirected method, unsigned and directed method, an STGCN with X layers, a Simple TCN with X layers, and a global residual layer [13].

Scenario	Walking (HUMAN3.6M_MPJPE)								Eating (HUMAN3.6M_MPJPE)								Smoking (HUMAN3.6M_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	17.7	33.5	56.3	63.6	72.2	77.7	80.9	82.3	11.0	22.4	40.7	48.4	61.3	72.8	81.8	87.1	11.6	22.8	41.3	48.9	60.0	69.4	77.2	81.7
LTD-10-10 [18]	11.1	21.4	37.3	42.9	53.1	59.9	66.2	70.7	7.0	14.8	29.8	37.3	51.1	62.5	72.9	78.6	7.5	15.5	30.7	37.5	49.4	59.2	66.9	71.8
DCT-RNN-GCN [17]	10.0	19.5	34.2	39.8	47.4	52.1	55.5	58.1	6.4	14.0	28.7	36.2	50.0	61.4	70.6	75.5	7.0	14.9	29.9	36.4	47.5	56.6	64.4	69.5
STS-GCN [4]	10.7	16.9	29.1	32.9	40.6	45.0	48.0	51.8	6.8	11.3	22.6	25.4	33.9	40.2	46.2	52.4	7.2	11.6	22.3	25.8	33.6	39.6	45.4	50.0
Ours	**9.5**	**15.4**	**25.6**	**30.1**	**37.7**	**42.3**	**44.3**	**48.3**	**6.2**	**10.6**	**19.8**	**24.4**	**32.2**	**39.1**	**46.0**	**51.5**	**6.5**	**10.9**	**20.0**	**24.4**	**32.2**	**38.3**	**43.1**	**49.7**

Scenario	Discussion (HUMAN3.6M_MPJPE)								Directions (HUMAN3.6M_MPJPE)								Greeting (HUMAN3.6M_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	17.1	34.5	64.8	77.6	98.1	112.9	123.0	129.3	13.5	29.0	57.6	69.7	86.6	99.8	109.9	115.8	22.0	45.0	82.0	96.0	116.9	130.7	142.7	147.3
LTD-10-10 [18]	10.8	24.0	52.7	65.8	88.1	104.4	115.5	121.6	8.0	18.8	43.7	54.9	72.2	86.7	98.5	105.8	14.8	31.4	65.3	79.7	103.7	120.6	134.7	140.9
DCT-RNN-GCN [17]	10.2	23.4	52.1	65.4	86.6	102.2	113.2	119.8	7.4	18.5	44.5	56.5	73.9	88.2	100.1	106.5	13.7	30.1	63.8	78.1	101.9	118.4	132.7	138.8
STS-GCN [4]	9.8	16.8	33.4	40.2	53.4	63.6	72.3	78.8	7.4	13.5	29.2	34.7	47.6	56.5	64.5	71.0	12.4	21.8	42.1	49.2	64.8	76.3	85.5	91.6
Ours	**8.6**	**15.7**	**30.9**	**38.7**	**51.9**	**62.6**	**71.4**	**77.8**	**6.5**	**12.7**	**26.7**	**33.2**	**45.8**	**55.4**	**64.0**	**70.9**	**11.2**	**20.1**	**38.5**	**46.9**	**62.3**	**74.0**	**83.9**	**91.1**

Scenario	Phoning (HUMAN3.6M_MPJPE)								Posing (HUMAN3.6M_MPJPE)								Purchases (HUMAN3.6M_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	13.5	26.6	49.9	59.9	77.1	92.1	105.5	114.0	16.9	36.7	75.7	92.9	122.5	148.8	171.8	187.4	20.3	41.8	76.5	89.9	111.3	129.1	143.1	151.5
LTD-10-10 [18]	9.3	19.1	39.8	49.7	67.8	83.0	96.4	105.1	10.9	25.1	59.1	75.9	107.6	136.1	159.5	175.0	13.9	30.3	62.2	75.9	98.3	115.1	130.1	139.3
DCT-RNN-GCN [17]	8.6	18.3	39.0	49.2	67.4	82.9	96.5	105.0	10.2	24.2	58.5	75.8	107.6	136.8	161.4	178.2	13.0	29.2	60.4	73.9	95.6	110.9	125.0	134.2
STS-GCN [4]	8.2	13.7	26.9	30.9	41.8	51.1	59.3	66.1	9.9	18.0	38.2	45.6	64.3	79.3	94.5	106.4	11.9	21.3	42.0	48.7	63.7	74.9	86.2	93.5
Ours	**7.4**	**12.7**	**24.1**	**29.7**	**40.2**	**49.5**	**58.2**	**65.0**	**8.4**	**16.3**	**34.5**	**43.0**	**60.5**	**77.5**	**92.3**	**102.1**	**10.7**	**20.4**	**38.5**	**45.8**	**61.5**	**74.3**	**85.3**	**93.0**

Scenario	Sitting (HUMAN3.6M_MPJPE)								Sitting Down (HUMAN3.6M_MPJPE)								Taking Photo (HUMAN3.6M_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	13.5	27.0	52.0	63.1	82.4	98.8	112.4	120.7	20.7	40.6	70.4	82.7	106.5	125.1	139.8	150.3	12.7	26.0	52.1	63.6	84.4	102.4	117.7	128.1
LTD-10-10 [18]	9.8	20.5	44.2	55.9	76.4	93.1	106.9	115.7	15.6	31.4	59.1	71.7	96.2	115.2	130.8	142.2	8.9	18.9	41.0	51.7	72.5	90.9	105.9	116.3
DCT-RNN-GCN [17]	9.3	20.1	44.3	56.0	76.4	93.1	107.0	115.9	14.9	30.7	59.1	72.0	97.0	116.1	132.1	143.6	8.3	18.4	40.7	51.5	72.1	90.1	105.5	115.9
STS-GCN [4]	9.1	15.1	29.9	35.0	47.7	57.0	67.4	75.2	14.4	23.7	41.9	47.9	63.3	73.9	86.2	94.3	8.2	14.2	29.7	33.6	47.0	57.4	67.2	76.9
Ours	**8.3**	**14.3**	**27.4**	**33.9**	**46.0**	**56.4**	**66.4**	**74.8**	**14.3**	**22.7**	**39.4**	**47.0**	**61.4**	**72.6**	**84.9**	**94.2**	**7.6**	**13.4**	**26.6**	**32.4**	**44.3**	**56.3**	**66.9**	**74.6**

Scenario	Waiting (HUMAN3.6M_MPJPE)								Walking Dog (HUMAN3.6M_MPJPE)								Walking Together (HUMAN3.6M_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	14.6	29.7	58.1	69.7	87.3	100.3	110.7	117.7	27.7	53.6	90.7	103.3	122.4	133.8	151.1	162.4	15.3	30.4	53.1	61.2	72.0	77.7	82.9	87.4
LTD-10-10 [18]	9.2	19.5	43.3	54.4	73.4	88.2	99.8	107.5	20.9	40.7	73.6	86.6	109.7	122.8	139.0	150.1	9.6	19.4	36.5	44.0	55.7	61.3	66.4	69.8
DCT-RNN-GCN [17]	8.7	19.2	43.4	54.9	74.5	89.0	100.3	108.2	20.1	40.3	73.3	86.3	108.2	120.6	135.9	146.9	8.9	18.4	35.1	41.9	52.7	57.8	62.0	64.9
STS-GCN [4]	8.6	14.7	29.6	35.2	47.3	56.8	66.1	72.0	17.6	29.4	52.6	59.6	74.7	85.7	96.2	102.6	8.6	14.3	26.5	30.5	38.9	44.0	48.2	51.1
Ours	**7.4**	**13.2**	**26.9**	**32.8**	**44.5**	**53.8**	**62.3**	**69.0**	**15.9**	**27.5**	**48.8**	**56.7**	**71.2**	**83.5**	**93.1**	**101.2**	**7.7**	**13.1**	**23.2**	**28.1**	**35.9**	**41.8**	**46.2**	**50.8**

FIGURE 5. **The results in terms of MPJPE on the Human3.6M dataset in 15 scenarios.**

Scenario	Average (HUMAN3.6M_MPJPE)								Average (AMASS_MPJPE)								Average (3DPW_MPJPE)							
milliseconds (ms)	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000	80	160	320	400	560	720	880	1000
ConvSeq2Seq [19]	16.6	33.3	61.4	72.7	90.7	104.7	116.7	124.2	20.6	39.6	59.7	67.6	79.0	87.0	91.5	93.5	18.8	32.9	52.0	58.8	69.4	77.0	83.6	87.8
LTD-10-10 [18]	11.2	23.4	47.9	58.9	78.3	93.3	106.0	114.0	10.3	19.3	36.6	44.6	61.5	75.9	86.2	91.2	12.0	22.0	38.9	46.2	59.1	69.1	76.5	81.1
DCT-RNN-GCN [17]	10.4	22.6	47.1	58.3	77.3	91.8	104.1	112.1	11.3	20.7	35.7	42.0	51.7	58.6	63.4	67.2	12.6	23.1	39.0	45.4	56.0	63.6	69.7	73.7
STS-GCN [4]	10.1	17.1	33.1	38.3	50.8	60.1	68.9	75.6	10.0	12.5	21.8	24.5	31.9	38.1	42.7	45.5	**8.6**	12.8	21.0	24.5	30.4	35.7	39.6	42.3
Ours	**9.1**	**15.9**	**30.1**	**36.4**	**48.5**	**58.5**	**67.2**	**74.3**	**9.8**	**11.5**	**19.9**	**23.9**	**30.2**	**36.1**	**40.1**	**44.6**	9.4	**12.5**	**20.9**	**23.8**	**30.2**	**35.6**	**39.6**	42.5

FIGURE 6. Average results in terms of MPJPE on Human3.6M datasets (Average (HUMAN3.6M_MPJPE)), the average results based on MPJPE metric on AMASS datasets (Average (AMASS_MPJPE)) and the average results based on MPJPE metrics trained with the Human3.6M datasets and validated on the 3DPW datasets(Average (3DPW_MPJPE)).

To simplify the identity notation, the notations N^d, C^d, T^d, V^d describing the dimensions are simplified to N, C, T, V, respectively. As shown in Fig. 4 and described in the pseudocode provided in Appendix A, N input frames with T frames are input into the proposed model, and we can obtain the input features matrix with dimension NC_1TV from the N input frames with T frames. Then, the signed and undirected temporal adjacency matrix with dimension VTT and the unsigned and directed spatial adjacency matrix [14] with dimension TVV are obtained by the signed and undirected method and the unsigned and directed method [15], respectively.

Combined with the input feature matrix above, the operation is performed in the STGCN model, and the process of this operation is shown in detail in the pseudocode presented in Appendix A. Specifically, in the feedforward process, steps 4-7 comprise the operation of the global residual layer, and steps 10-15 comprise the processing method we adopted for the temporal and spatial adjacency matrix, i.e., the signed undirected and unsigned directed methods. Steps 16-20 indicate the operations of the STGCN layer, i.e., the signed and undirected temporal adjacency matrix obtained after the processing is graphically convolved with the input feature matrix and then graphically convolved with the unsigned and directed spatial adjacency matrix.

As described in Appendix A, steps 4-7 are a global residual process. In Fig. 4, corresponding to the parts of a single STGCN layer, the dimensional transformation process is shown in detail in Appendix C, except for the temporal and spatial graph convolution operation on the input features matrix, a residual module is added at the stage of the architecture in which the input feature matrix enters the STGCN layer, and the common convolution operation is performed in the residual module, i.e., $Conv2D^I$, and $H^{(l)}_{NC_{l+1}QW}$ are obtained. The common convolution operation $Conv2D^I$ is also performed after the input features matrix completes the temporal and spatial graph convolution operation, i.e., when $H^{(l+1)}_{NC_lQW}$ is obtained.

Then, $H^{(l)}_{NC_{l+1}QW}$ and $H^{(l+1)}_{NC_lQW}$ are calculated by the residual module and summed to obtain the output $H^{(l+1)}_{NC_{l+1}QW}$, which is used as the input of the next layer of the STGCN model. These operations are repeated until the loop ends.

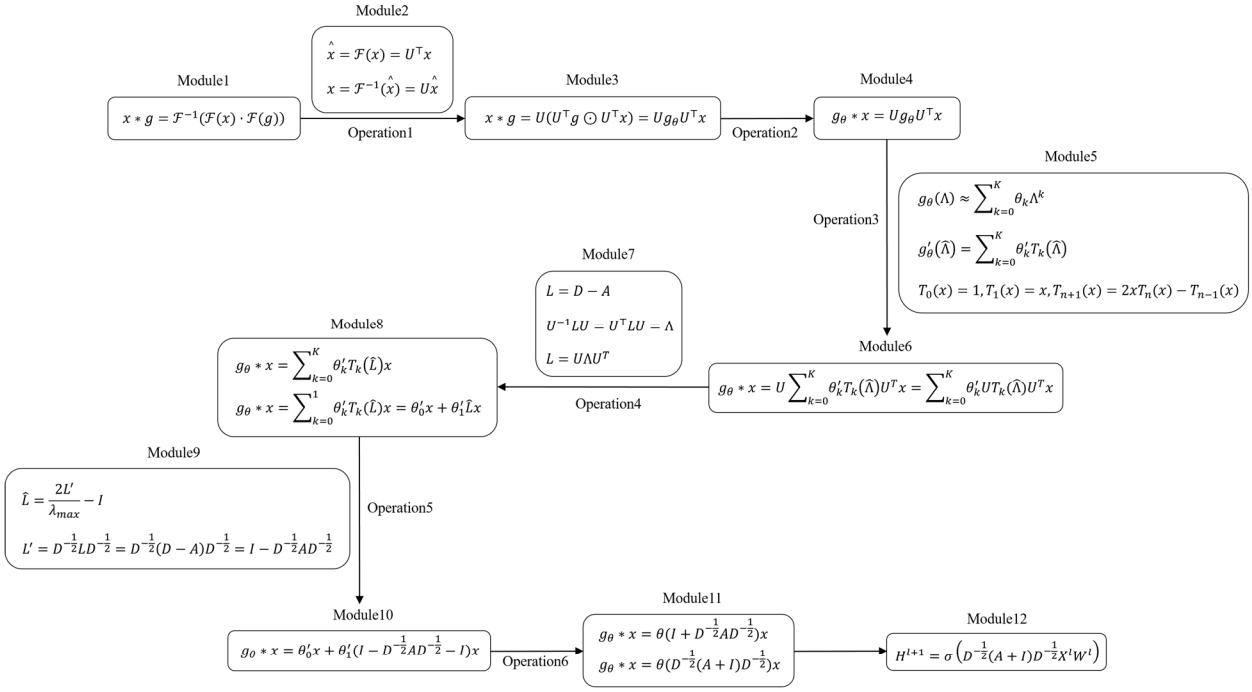
Thus, steps 10-20 encode the signal. In the remaining decoding process, a simple TCN is utilized to predict the next T_{l+1} frames. The first layer of the TCN part of the architecture comprises a $Conv2D^{II}$ convolutional layer with a residual module added to each of the remaining three layers, which are composed of $Conv2D^{II}$ convolutional layers. Thus, steps 23-29 perform decoding.

Finally, the output X_O and $H^{(l)-global}_{NCQW}$ are obtained from the results of the decoding steps and the global residual process, respectively.

We utilized the Human3.6M and AMASS benchmark datasets to train the model based on the mean per-joint position error (MPJPE) coordinate-based joint prediction metric and evaluated the performance of our proposed method on the 3DPW dataset.

V. EXPERIMENTS

After training our proposed model according to the strategy described above, we conducted experiments on the Human3.6M dataset for the human skeleton forecasting task. For a fair comparison with existing methods, we adopted the conventional approach to categorize the human body in motion with a skeleton in common forecasting scenarios. The results shown in Figs. 5 and 6 demonstrate that our proposed approach performed better compared to an STS-GCN model on long- and short-term human pose prediction tasks in each section in terms of the evaluation metric MPJPE on both the Human3.6M and AMASS datasets. We validated the effect of the 3DPW dataset in the model trained on the AMASS dataset and found that it exceeded the original results on sections 160-720 and showed a close approximation to the original model in the other parts. This shows that the matrix processing in our proposed model based on the significance and directionality of the adjacency matrices was able to extract features necessary for prediction with the separated temporal and spatial adjacency matrices, which allowed the model to learn associations between nodes in the temporal and spatial dimensions more effectively. This enabled it to predict human poses based on positions of the joints more accurately than previous methods.



Module1:The Fourier transform of the graph convolution operation

Module2:Fourier transform of x

& Fourier inverse transform of $\hat{x}$

Module3:The Matrix form of Fourier transform of the graph convolution operation

Module4:Express graph operation of Module3 with g_{θ}

Module5:Using the polynomial to approximate the $g_{\theta}(\Lambda)$, especially using the Chebyshev polynomial $g'_{\theta}(\hat{\Lambda})$, and take$K = 1$

Module6:The graph operation after substituting Module5 into Module4

Module7:Utilizing the degree matrix D and the adjacency matrix A of the graph to express the Laplacian matrix L

& Utilizing the Fourier transform and Fourier inverse transform to diagonalize the the Laplacian matrix L into Λ

& Also the Laplacian matrix L could be expressed as $L = U\Lambda U^T$ by simple calculation from the above

Module8:The graph operation represented with $\hat{L}$

Module9:The disassembled Laplace matrix L'

& The normalized and regularized Laplace matrix $\hat{L}$, where $\lambda_{max} = 2$

Module10:The graph operation after Operation5

Module11:The graph operation after the computational parameters-reducing operation

& The graph operation after Utilizing the technique of the "Normalisation Trick"

Module12:Representing Module11 in matrix form as Module12

i.e., input a whole matrix of features X^l, multiplied by the weights W^l corresponding to each feature, and after Graph Filter, the value of H^{l+1} is returned after a nonlinear activation σ.

Operation1:Represent Module2 as Module3 according to the Fourier transform and Fourier inverse transform in Module1

Operation2:Express the the process of graph convolution operation in Module3 as Module4

Operation3:Substituting Chebyshev polynomials in Module5 into Module4

Operation4:Utilizing $L = U\Lambda U^T$ in Module7 and take $K = 1$ to simplify the Module6 as Module8

Operation5:Substituting L' into $\hat{L}$ of Module9,and substituting $\hat{L}$ of Module9 into Module8

Operation6:In order to reduce the number of computational parameters, use θ instead of θ'_0, θ'_1

& In order to achieve the final result of a feature x corresponding to a θ, utilizing the technique of the "Normalisation Trick"

FIGURE 7. **Mathematical Derivation of GCN.**

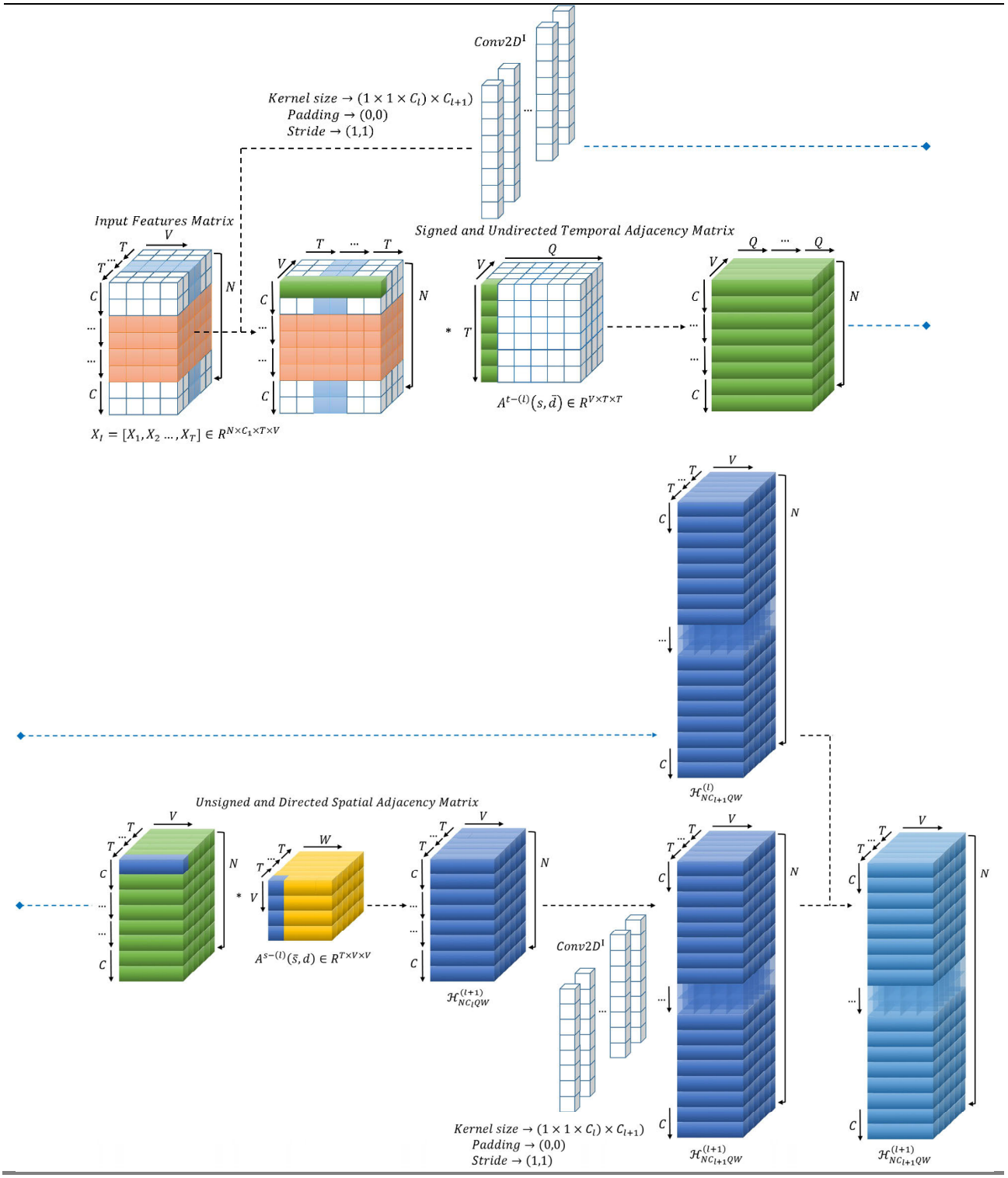


FIGURE 8. The Process of Transformation of single STGCN Layer.

VI. CONCLUSION

In this study, we have proposed an approach based on a GCN model for human pose prediction tasks that focuses on the symbolism and directionality of keypoints representing human joints in the temporal and spatial dimensions in the task of human pose prediction. Our experimental results show that the undirected processing in the temporal dimension and unsigned processing in the spatial dimension of the adjacency matrix enabled our proposed model to perform better than previous methods. That is, our method extracted more

TABLE 1. Table of Notations used in Paper.

Notation	Description
A^{st-l}	The spatial temporal adjacency matrix at layer l
X^l	Input feature matrix at layer l
W^l	The weight matrix corresponding to the feature matrix at layer l
$H^{(l+1)}$	The hidden feature matrix at layer $l+1$
A	Adjacency matrix
I	Identity matrix
D	Degree matrix
A^{st}	The spatial temporal adjacency matrix
A^{t-l}	The temporal adjacency matrix
A^{s-l}	The spatial adjacency matrix
$NCtV, WVtq \rightarrow NCqW$	The Einstein summation process for the space-time dimension
$NCtV, Vtq \rightarrow NCqV$	The Einstein summation process for the time dimension
$NCtV, tVW \rightarrow NCtW$	The Einstein summation process for the space dimension
N	The number of samples
C	The number of features
t, q	The number of frames
V, W	The number of joints
$X_1 = [X_1, X_2 \cdots, X_T] \in R^{N \times C_1 \times T \times V}$	The input feature matrix at layer 1
$Permute_{[0,2,1,3]}(a)$	Permute the dimension of a from $[0,1,2,3]$ to $[0,2,1,3]$
$H^{(1)-global}_{NT_1CV}$	The input matrix in global residual part
$H^{(2)-global}_{NT_2CV}$	The output matrix in global residual part
$uniform(a,b)$	Relocate the elements in $A^{t-(l)}$ or $A^{s-(l)}$ to random values following a uniform distribution in the range $[a,b]$
$abs(A^{s-(l)}(s,d))$	Unsigned and Directed method
$(A^{t-(l)}(s,d) + A^{t-(l)}(s,d)^T)/2$	Signed and Undirected method
$A^{s-(l)}(s,d) \in R^{T \times V \times V}$	Signed and Directed spatial adjacency matrix
$A^{t-(l)}(s,d) \in R^{V \times T \times T}$	Signed and Directed temporal adjacency matrix
$A^{s-(l)}(\bar{s},d) \in R^{T \times V \times V}$	Unsigned and Directed spatial adjacency matrix
$A^{t-(l)}(s,\bar{d}) \in R^{V \times T \times T}$	Signed and Undirected temporal adjacency matrix
$X_O = [X_{T+1}, X_{T+2} \cdots, X_{T+K}] \in R^{N \times C_9 \times T \times V}$	The output feature matrix
$\mathrm{Re}sidual(a,b)$	Residual part
$C = [C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9] = [3,8,16,32,64,128,64,128,3]$	The number of features from layer 1 to layer 9
$T = [T_1, T_2, T_3, T_4, T_5] = [10,25,25,25,25]$	The number of frames from layer 1 to layer 4
σ	Non-linearity
$Conv2D^I$	Convolution operation with kernel size $(1 \times 1 \times C_l) \times C_{l+1}$,padding $(0,0)$ and stride $(1,1)$
$Conv2D^{II}$	Convolution operation with kernel size $(3 \times 3 \times T_l) \times T_{l+1}$,padding $(1,1)$ and stride $(1,1)$
$Conv2D^{III}$	Convolution operation with kernel size $(1 \times 1 \times T_l) \times T_{l+1}$,padding $(0,0)$ and stride $(1,1)$

Algorithm 1 Algorithm of Feedforward Process

$H^{(1)}_{NC_1TV} \leftarrow X_I$
$C = [C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9]$
$= [3, 8, 16, 32, 64, 128, 64, 128, 3]$,
$T = [T_1, T_2, T_3, T_4, T_5]$
$= [10, 25, 25, 25, 25]$
$L_1 = 9, L_2 = 4$
$H^{(1)-global}_{NT_1CV} \leftarrow Permute_{[0,2,1,3]}(X_I)$
$H^{(2)-global}_{NT_2CV} \leftarrow Conv2D^{III}_{[(1\times1\times T_1)\times T_2]}(H^{(1)-global}_{NT_1CV})$
$H^{(1)-global}_{NTCV} \leftarrow H^{(2)-global}_{NT_2CV}$
$H^{(1)-global}_{NCTV} \leftarrow Permute_{[0,2,1,3]}(H^{(1)-global}_{NTCV})$
for $l = 1$ *to* L_1 *do*
  for $\forall v \in V$ *do*
    $stdv\_t \leftarrow 1 \Big/ \sqrt{T}$
    $A^{t-(l)}(s, d) \leftarrow uniform(-stdv\_t, stdv\_t)$
    $A^{t-(l)}(s, \overline{d}) \leftarrow (A^{t-(l)}(s, d) + A^{t-(l)}(s, d)^T) \big/ 2$
    $stdv\_s \leftarrow 1 \Big/ \sqrt{V}$
    $A^{s-(l)}(s, d) \leftarrow uniform(-stdv\_s, stdv\_s)$
    $A^{s-(l)}(\overline{s}, d) \leftarrow abs(A^{s-(l)}(s, d))$
    $H^{(l+1)}_{NC_lQW} \leftarrow \sigma(A^{s-(l)}_{TVW}(\overline{s}, d)(A^{t-(l)}_{VTQ}(s, \overline{d})H^{(l)}_{NC_lTV})W^{(l)})$
    $H^{(l+1)}_{NC_{l+1}QW} \leftarrow Conv2D^{I}_{[(1\times1\times C_l)\times C_{l+1}]}(H^{(l+1)}_{NC_lQW})$
    $H^{(l)}_{NC_{l+1}QW} \leftarrow Conv2D^{I}_{[(1\times1\times C_l)\times C_{l+1}]}(H^{(l)}_{NC_lQW})$
    $H^{(l+1)}_{NC_{l+1}QW} \leftarrow \mathrm{Re}sidual(H^{(l+1)}_{NC_{l+1}QW}, H^{(l)}_{NC_{l+1}QW})$
    $H^{(l)}_{NC_lTV} \leftarrow H^{(l+1)}_{NC_{l+1}QW}$
  end
end
$H^{(l)}_{NT_lCV} \leftarrow Permute_{[0,2,1,3]}(H^{(l)}_{NC_lTV})$
$H^{(l+1)}_{NT_{l+1}CV} \leftarrow Conv2D^{II}_{[(3\times3\times T_l)\times T_{l+1}]}(H^{(l)}_{NT_lCV})$
$H^{(l)}_{NT_lCV} \leftarrow H^{(l+1)}_{NT_{l+1}CV}$
for $l = 2$ *to* L_2 *do*
  $H^{(l+1)}_{NT_{l+1}CV} \leftarrow Conv2D^{II}_{[(3\times3\times T_l)\times T_{l+1}]}(H^{(l)}_{NT_lCV})$
  $H^{(l+1)}_{NT_{l+1}CV} \leftarrow \mathrm{Re}sidual(H^{(l+1)}_{NT_{l+1}CV}, H^{(l)}_{NT_lCV})$
  $H^{(l)}_{NT_lCV} \leftarrow H^{(l+1)}_{NT_{l+1}CV}$
end
$H^{(l)}_{NTCV} \leftarrow H^{(l)}_{NT_lCV}$
$H^{(l)}_{NCTV} \leftarrow Permute_{[0,2,1,3]}(H^{(l)}_{NTCV})$
$H^{(l)}_{NCTV} \leftarrow \mathrm{Re}sidual(H^{(l)}_{NCTV}, H^{(1)-global}_{NCTV})$
$X_O \leftarrow H^{(l)}_{NCTV}$

useful information about the correlations between joints and learned to predict their positions in space and time more accurately and allows the model to learn this information better. Such refinement opens up a new avenue for further research on human pose prediction, and we hope that our mathematical reasoning and the discussion of GCN and STGCN models in the present work will prove useful as a starting point.

In the future, we plan to continue to improve the proposed model using the strategies described in this study. In particular, we plan to consider integrating a GAT model [12] or other attentional mechanisms [9], [17] in the encoder and adopting a better prediction module in the decoder. To this end, we plan to explore the application of TCN models for pose prediction.

APPENDIX A

See Algorithm 1.

APPENDIX B

See Fig. 7.

APPENDIX C

See Fig. 8.

APPENDIX D

See Table 1.

REFERENCES

[1] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4346–4354.

[2] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 283–298, Feb. 2008.

[3] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4674–4683.

[4] T. Sofianos, A. Sampieri, L. Franco, and F. Galasso, "Space-time-separable graph convolutional network for pose forecasting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 11189–11198.

[5] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 1–10.

[6] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014.

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[8] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," 2017, *arXiv:1709.05584*.

[9] J. B. Lee, R. A. Rossi, S. Kim, N. K. Ahmed, and E. Koh, "Attention models in graphs: A survey," *ACM Trans. Knowl. Discovery From Data*, vol. 13, no. 6, pp. 1–25, Dec. 2019.

[10] F. Scarselli, M. Gori, and A. C. Tsoi, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Dec. 2008.

[11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[12] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[13] L. Dang, Y. Nie, C. Long, Q. Zhang, and G. Li, "MSR-GCN: Multi-scale residual graph convolution networks for human motion prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 11447–11456.

[14] T. Derr, Y. Ma, and J. Tang, "Signed graph convolutional networks," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 929–934.

[15] Z. Tong, Y. Liang, C. Sun, D. S. Rosenblum, and A. Lim, "Directed graph convolutional network," 2020, *arXiv:2004.13970*.
[16] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
[17] W. Mao, M. Liu, and M. Salzmann, "History repeats itself: Human motion prediction via motion attention," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, U.K., Aug. 2020, pp. 23–28.
[18] W. Mao, M. Liu, M. Salzmann, and H. Li, "Learning trajectory dependencies for human motion prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9488–9496.
[19] C. Li, Z. Zhang, W. S. Lee, and G. H. Lee, "Convolutional sequence to sequence model for human dynamics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5226–5234.
[20] J. Zhou, G. Cui, and S. Hu, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
[21] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

JINHUI LI received the bachelor's degree in engineering from the Shandong University of Science and Technology, China, in 2018. He is currently pursuing the master's degree in electrical and electronic engineering with Chung-Ang University, South Korea. In 2020, he was with Wenzhou Changjiang Automotive Electronics Company Ltd., China, as a Hardware Engineer, where he participated in the research and development of vehicle central control panel hardware. His research interests include machine learning and deep learning, humanoid robotics, and UAVs.

JIANYING HUANG received the B.S. and M.S. degrees in electrical and electronic engineering from Chung-Ang University, South Korea, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. His research interests include machine learning and deep learning.

HOON KANG (Member, IEEE) was born in Seoul, South Korea, in 1959. He received the B.S. and M.S. degrees in electronic engineering from Seoul National University, South Korea, in 1982 and 1984, respectively, and the Ph.D. degree in computer-integrated manufacturing systems from the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 1989. From 1989 to 1991, he was a Postdoctoral Fellow and then a Research Associate with the Department of Electrical Engineering, Georgia Tech. In 1992, he joined the School of Electrical and Electronics Engineering, Chung-Ang University, Seoul. As he participated in a number of projects sponsored by the National Science Foundation, Office of Naval Research, Ford Motor Company, and Honeywell Inc., he developed new research ideas on fuzzy logic control, intelligent robotic control, and fault detection and identification. His research interests include artificial intelligence, such as neural network models and deep learning, fuzzy expert systems, evolutionary computations, intelligent humanoid robots, and robot vision. He was the Department Chair and a Financial Secretary with the Korean Intelligent Information Systems (KIIS) and the Institute of Control, Robotics and Systems Engineers (ICROS). He also joined the Institute of Electronics Engineers of Korea (IEEK), South Korea, as a Financial Secretary, an Editorial Board Member, and the Director of General Affairs.