

## Research Article

# Multimodal Semisupervised Deep Graph Learning for Automatic Precipitation Nowcasting

Kaichao Miao,<sup>1</sup> Wei Wang,<sup>2</sup> Rui Hu,<sup>3</sup> Lei Zhang,<sup>1</sup> Yali Zhang,<sup>1</sup> Xiang Wang,<sup>1</sup> and Fudong Nian <sup>4</sup>

<sup>1</sup>Anhui Public Meteorological Service Center, Anhui Meteorological Bureau, Hefei, Anhui, China

<sup>2</sup>Center for Assessment and Demonstration Research, Academy of Military Science, Beijing, China

<sup>3</sup>School of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, China

<sup>4</sup>School of Advanced Manufacturing Engineering, Hefei University, Hefei, Anhui, China

Correspondence should be addressed to Fudong Nian; nianfd@hfuu.edu.cn

Received 21 July 2020; Accepted 21 September 2020; Published 1 October 2020

Academic Editor: Jia-Bao Liu

Copyright © 2020 Kaichao Miao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Precipitation nowcasting plays a key role in land security and emergency management of natural calamities. A majority of existing deep learning-based techniques realize precipitation nowcasting by learning a deep nonlinear function from a single information source, e.g., weather radar. In this study, we propose a novel multimodal semisupervised deep graph learning framework for precipitation nowcasting. Unlike existing studies, different modalities of observation data (including both meteorological and nonmeteorological data) are modeled jointly, thereby benefiting each other. All information is converted into image structures, next, precipitation nowcasting is deemed as a computer vision task to be optimized. To handle areas with unavailable precipitation, we convert all observation information into a graph structure and introduce a semisupervised graph convolutional network with a sequence connect architecture to learn the features of all local areas. With the learned features, precipitation is predicted through a multilayer fully connected regression network. Experiments on real datasets confirm the effectiveness of the proposed method.

## 1. Introduction

The goal of precipitation nowcasting is to predict the future rainfall intensity in a local region over a relatively short period. This method plays an important role in land security and emergency management of natural calamities. Though it is very crucial, precipitation nowcasting still depends on the real-time manual analysis of meteorological observation data by forecasters.

To address this issue, researchers are increasingly trying to replace forecasters with computers, e.g., [1]. In essence, precipitation nowcasting is a spatiotemporal sequence forecasting problem with a sequence of past meteorological observation data as input and a sequence of fixed numbers (usually larger than 1) of future precipitation as output. However, traditional optical flow-based methods have unsatisfactory performance.

Precipitation nowcasting has three long-standing problems. The first challenge is spatiotemporal correlations. Precipitation at any location is influenced by the weather condition of its neighboring region over a relatively short period. The second challenge is diversified observation data. With the progress of technology, we can obtain various meteorological or nonmeteorological observation data of a certain region, e.g., radar echo maps, satellite images, topographic map, temperature, and air humidity. However, handling these data in a unified framework is an open problem. The third challenge is semisupervision. It is generally known that the distribution of meteorological stations for observing precipitation is uneven, implying that precipitation at many locations is unknown.

Since the recent advances in deep neural networks [2, 3], their high capability in various classification and regression tasks has been successfully demonstrated, e.g., image

classification by [4], object detection by [5], video representation by [6], and speech recognition by [7]. There are also some studies on precipitation nowcasting using deep learning techniques. Xingjian et al. [8] and Shi et al. [9] first introduced a convolutional long short-term memory (ConvLSTM) network to capture spatiotemporal correlations, which has been shown to outperform traditional optical flow-based methods for precipitation nowcasting, indicating that deep learning models have a huge potential for solving this problem. Subsequently, Singh et al. [10] and Karevan and Suykens [11] introduced a more complicated LSTM structure, and Shi et al. [12] used convolutional neural networks (CNNs) to extract more efficient representations.

Although the methods above can achieve some progress, their inputs are merely weather radar echo maps, and they only focus on addressing the problem of spatiotemporal correlations, neglecting the latter two problems.

Inspired by the success of Deep Learning Vision for Nonvision Tasks by [13], we propose a novel multimodal semisupervised deep graph learning framework for precipitation nowcasting in this study. In contrast to previous studies, different modalities of observation data (including both meteorological and nonmeteorological data) are modeled jointly, and thus, benefiting each other. All information is converted to an image structure. Then, precipitation nowcasting is deemed as a computer vision task to be optimized. We convert all observation information into a graph structure and introduce a semisupervised graph convolutional network (GCN) with a sequence connect architecture to learn the features of all local areas for handling areas without available precipitation. With the learned features, precipitation is predicted through a multilayer fully connected regression network.

We summarize the contributions of this work as follows:

- (1) We introduce a novel short-term precipitation nowcasting solution by leveraging multimodal observation data (including meteorological and non-meteorological data)
- (2) A unified multimodal semisupervised deep graph learning is proposed, aiming at simultaneously addressing three long-standing problems in precipitation nowcasting, i.e., spatiotemporal correlations, diversified observation data, and semisupervision
- (3) To the best of our knowledge, this is the first attempt to address the precipitation nowcasting problem using a GCN
- (4) The experimental results on the collected large-scale real dataset validate the effectiveness of the proposed solution

The rest of the paper is organized as follows. In Section 2, related work is reviewed. Section 3 presents the problem formulation and introduces the proposed semisupervised deep regression framework in detail. In Section 4, we report and analyze the experimental results. Finally, we conclude the paper and discuss future work in Section 5.

## 2. Related Work

In this section, we briefly review the closely related methods in two folds: deep learning-based precipitation nowcasting methods and graph-based semisupervised learning techniques.

*2.1. Deep Learning-Based Precipitation Nowcasting.* Xingjian et al. [8] formulated precipitation nowcasting as a spatiotemporal sequence forecasting problem and proposed a ConvLSTM model, which extends the LSTM by [14], by adding convolutional structures into both input-to-state and state-to-state transitions to solve the problem. Using radar echo sequences for model training, the authors showed that ConvLSTM was better at capturing spatiotemporal correlations than fully connected LSTM and provided more accurate predictions than the real-time optical flow via variational methods for echoes of radar algorithm by [15]. However, the convolutional recurrence structure in ConvLSTM-based models is location invariant, whereas natural motion and transformation (e.g., rotation) are location variant in general. To address this, Shi et al. [9] proposed a trajectory-gated recurrent unit model that can actively learn a location-variant structure for recurrent connections. To obtain a more robust representation, Karevan and Suykens [11] proposed a two-layer spatiotemporal stacked LSTM model. Unlike LSTM-based methods which were used widely on sequence learning and time series prediction, Shi et al. [12] used recurrent dynamic CNNs to handle the spatiotemporal information of radar data.

Unfortunately, all techniques mentioned above only consider radar echo maps as the input data. In addition to radar echo maps, other types of meteorological observed data, such as satellite images, temperature, and air humidity, and nonmeteorological observation data, such as “topographic maps” are available, and these data were incorporated in this study. Notably, data from different sources are complementary. In this work, the proposed model is capable of not only handling spatiotemporal information from different observation sources simultaneously but also unifying labeled and unlabeled data.

*2.2. Graph-Based Semisupervised Learning.* Existing semisupervised learning methods using graph representations can be roughly be divided into two categories: methods that use some form of explicit graph Laplacian regularization and graph embedding-based approaches.

Typical graph Laplacian regularization techniques include label propagation by [16], manifold regularization by [17], and deep semisupervised embedding by [18]. Typical graph embedding-based methods include DeepWalk by [19], LINE by [20], and node2vec by [21]. Perozzi et al. [19] learned embedding via the prediction of the local neighborhood of nodes, sampled from random walks on a graph. Tang et al. and Grover and Leskovec [20, 21] extend DeepWalk with more sophisticated random walk or breadth-first search schemes.

Although the methods mentioned above have achieved significant progress, a multistep pipeline is included in these frameworks, which means that they cannot be trained in an end-to-end way. Recently, Kipf and Welling [22] proposed a more simplified GCN for semisupervised learning by employing a first-order approximation of spectral filters. The GCN and subsequent variants have achieved state-of-the-art results in various application areas, including multilabel classification by [23], zero-shot learning by [24], social networks by [25], and natural language processing by [26].

In this paper, graph-based semisupervised learning is utilized for precipitation nowcasting. There are two main differences between the proposed method and that by [22] (including its variants). First, as observation data are obtained from different sources, graphs in our framework are multimodal. In contrast, most graphs in conventional graph-based semisupervised learning methods are built by a single modality (image, text, etc.). Second, precipitation nowcasting is a regression problem in our study, whereas previous techniques focus on classification tasks.

### 3. Method

In this section, we first define problems of precipitation nowcasting processed in this paper. Then, we introduce the proposed multimodal semisupervised deep graph learning in details.

**3.1. Problem Definition.** Our goal is forecasting precipitation at a place for the next several hours based on the weather conditions of its neighbors over the past several hours. Precipitation is related to many factors. As illustrated in Figure 1, according to the existing observation conditions, herein, we assume that future precipitation can be predicted via modeling the following factors: radar echo maps, air humidity images, satellite images, temperature images, and available precipitation data over the past several hours with the corresponding topographic map. In addition, as precipitation prediction of every longitude and latitude is impossible, we assume that if we can divide a certain region to many cells (that is, each cell represents a local area) then precipitation at different locations in the same cell can be deemed as almost the same.

As the factors belong to different modalities, precipitation nowcasting can be deemed as a multimodal analysis task. Moreover, from Figure 1, we can observe that precipitation in some cells may be unknown when the target area is divided into many grid cells because precipitation stations are not placed everywhere. Thus, precipitation nowcasting is also a semisupervised learning problem in this paper. Therefore, we present a multimodal semisupervised learning method to solve the precipitation nowcasting problems herein. The problem formulations are detailed below.

Let  $\mathbf{X}_t = [\mathbf{x}_{re}, \mathbf{x}_{si}, \mathbf{x}_{tp}, \mathbf{x}_{ti}, \mathbf{x}_{ah}]$  be the observation data at time  $t$ , where  $\mathbf{x}_{re} = [\mathbf{x}_{re_t}, \mathbf{x}_{re_{t-1}}, \dots, \mathbf{x}_{re_{t-m}}]$  are the radar echo maps of the target area.  $\mathbf{x}_{si} = [\mathbf{x}_{si_t}, \mathbf{x}_{si_{t-1}}, \dots, \mathbf{x}_{si_{t-m}}]$  are the satellite images of the same area.  $\mathbf{x}_{tp}$  is the topographic map

of the same area.  $\mathbf{x}_{ti} = [\mathbf{x}_{ti_t}, \mathbf{x}_{ti_{t-1}}, \dots, \mathbf{x}_{ti_{t-m}}]$  and  $\mathbf{x}_{ah} = [\mathbf{x}_{ah_t}, \mathbf{x}_{ah_{t-1}}, \dots, \mathbf{x}_{ah_{t-m}}]$  are the temperature and air humidity image sequences from a remote sensing satellite, respectively.  $m$  is the sample index before time  $t$ . If the target area is divided into  $N \times N$  grid cells, the precipitation of all cells is denoted as  $\mathbf{P} = [\mathbf{P}_t, \mathbf{P}_{t-1}, \dots, \mathbf{P}_{t-m}]$ , where  $\mathbf{p}_u = [p_{u_1}, p_{u_2}, \dots, p_{u_{NN}}]$  ( $0 \leq u \leq m$ ). It should be noted that  $p_{u_k}$  will be unknown if no precipitation stations are placed in the corresponding cell. If a cell has several precipitation stations, the cell precipitation values will be the average observation values from the stations. In addition, the precipitation of  $n$  time points in the future is denoted as  $\mathbf{Y} = [\mathbf{y}_{t+1}, \mathbf{y}_{t+2}, \dots, \mathbf{y}_{t+n}]$ , where  $\mathbf{y}_v = [y_{v_1}, y_{v_2}, \dots, y_{v_{NN}}]$  ( $1 \leq v \leq n$ ) are the predicted precipitation in all local areas.

Therefore, the problem can be formulated as follows:

$$\mathbf{Y} = g(\mathbf{X}_t, \mathbf{P}). \quad (1)$$

To address it, we propose a multimodal semisupervised deep graph learning framework herein.

#### 3.2. Multimodal Semisupervised Deep Graph Learning

**3.2.1. Graph Representation.** Multimodal and heterogeneous data are represented by a graph structure. The target region of precipitation nowcasting is divided into many grid cells, and each cell represents a local area (as shown in Figure 2, the rectangular area represents the target region of the precipitation nowcasting and the blue points represent the precipitation stations). Let  $G(\mathbf{V}, \mathbf{E})$  be the graph structure of the observation data of the target region, where nodes  $\mathbf{V}$  represent a certain local area and edges  $\mathbf{E}$  represent the relationship of two local areas. Here, red points mean the precipitation is available while the green means the precipitation is unknown.

Therefore, the first problem is vector representation of  $G(\mathbf{V}, \mathbf{E})$ . To obtain the node and edge representations of  $G(\mathbf{V}, \mathbf{E})$ , we utilize a convolutional autoencoder to learn the feature representation of a radar echo map, an air humidity image, a satellite image, a temperature image, and a topographic map. The convolutional autoencoder (as shown in Figure 3) architecture in our method consists of five parts: an input image, an encoder, a feature representation layer, a decoder, and a reconstructed image. It should be noted that the five convolutional autoencoder architectures have the same structure but are trained separately. Moreover, to ensure that the input images are as similar as possible to the reconstructed images, we utilize mean structural similarity (MSSIM) as our objective function in the convolutional autoencoder training step. If the input and reconstructed images are marked as  $\mathbf{I}$  and  $\mathbf{I}'$ , respectively, then the MSSIM of  $\mathbf{I}$  and  $\mathbf{I}'$  is formulated as follows:

$$\text{MSSIM}(\mathbf{I}, \mathbf{I}') = \frac{1}{K} \sum_{j=1}^K \text{SSIM}(\mathbf{I}_j, \mathbf{I}'_j), \quad (2)$$

where  $K$  represents the image similarity computed in  $K$  patches and  $\mathbf{I}_j$  and  $\mathbf{I}'_j$  represent the  $j$ th image patch. SSMI is

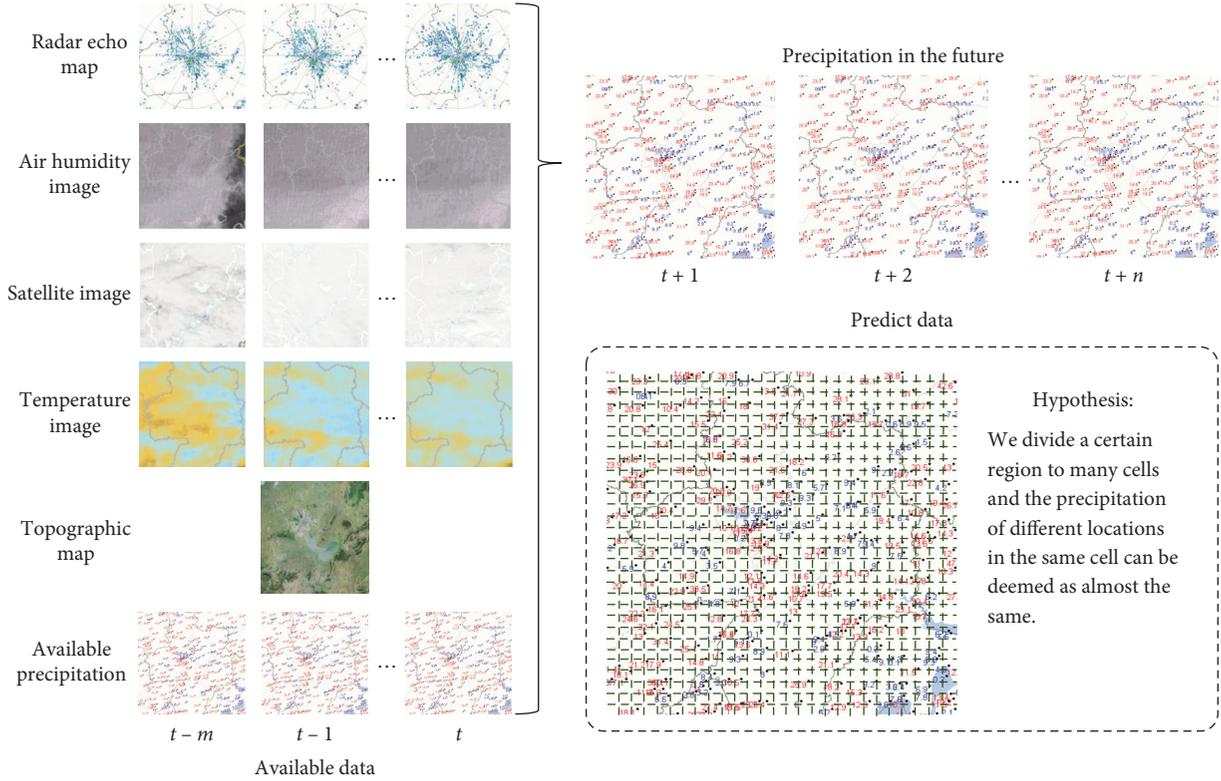


FIGURE 1: Motivation and the problem definition of precipitation nowcasting in this paper. In this paper, we assume that the future precipitation can be predicted via modeling the following factors: radar echo maps, air humidity images, satellite images, temperature images, and available precipitation data over the past several hours with the corresponding topographic map. In addition, we assume that if we can divide a certain region to many cells, then precipitation at different locations in the same cell can be deemed as almost the same.

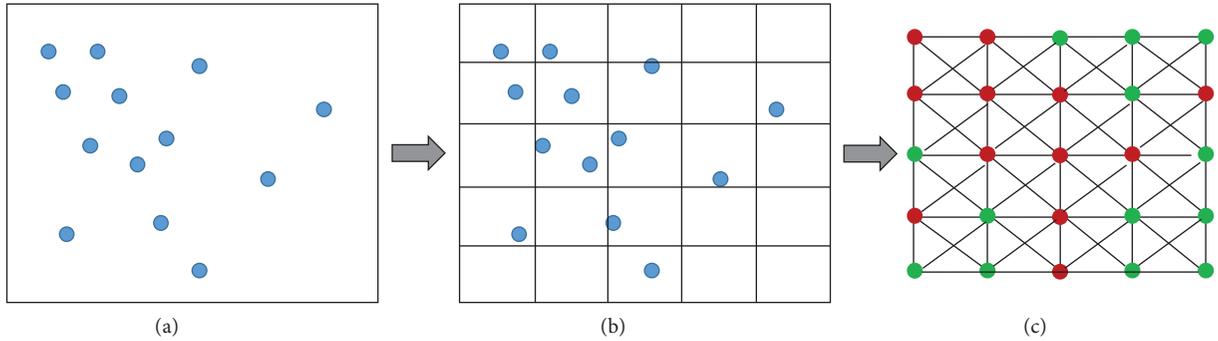


FIGURE 2: Illustration of the pipeline of graph construction. A certain area can be represented by a graph, where nodes represent a local area and the edges represent the relationship between two local areas. The red nodes mean the precipitation of the corresponding local area is available while the green means the precipitation is unknown.

the structural similarity measurement of two image patches formulated as

$$\text{SSIM}(I_j, I'_j) = \frac{2\mu_{I_j} \mu_{I'_j} + 2\sigma_{I_j, I'_j}}{(\mu_{I_j}^2 + \mu_{I'_j}^2)(\sigma_{I_j}^2 + \sigma_{I'_j}^2) + C}, \quad (3)$$

where  $\mu$  and  $\sigma$  are the mean and variance of the image patch and  $C$  is a small constant which avoids zero errors.

Once all convolutional autoencoders are trained, the node representations in graph  $G(\mathbf{V}, \mathbf{E})$  can be denoted as

$$\mathbf{f}_{v_i} = [\mathbf{f}_{\text{re}_{v_i}}, \mathbf{f}_{\text{si}_{v_i}}, \mathbf{f}_{\text{tp}_{v_i}}, \mathbf{f}_{\text{ti}_{v_i}}, \mathbf{f}_{\text{ah}_{v_i}}], \quad (4)$$

where  $\mathbf{f}_{v_i}$  is the representation of node  $v_i$ , and  $\mathbf{f}_{\text{re}_{v_i}}$ ,  $\mathbf{f}_{\text{si}_{v_i}}$ ,  $\mathbf{f}_{\text{tp}_{v_i}}$ ,  $\mathbf{f}_{\text{ti}_{v_i}}$ , and  $\mathbf{f}_{\text{ah}_{v_i}}$  are the representations obtained from the outputs of the decoders in Figure 3.

Then, the edges in graph  $G(\mathbf{V}, \mathbf{E})$  are denoted as

$$s_{i,j} = \frac{1}{\|\mathbf{f}_{v_i} - \mathbf{f}_{v_j}\|_2}, \quad (5)$$

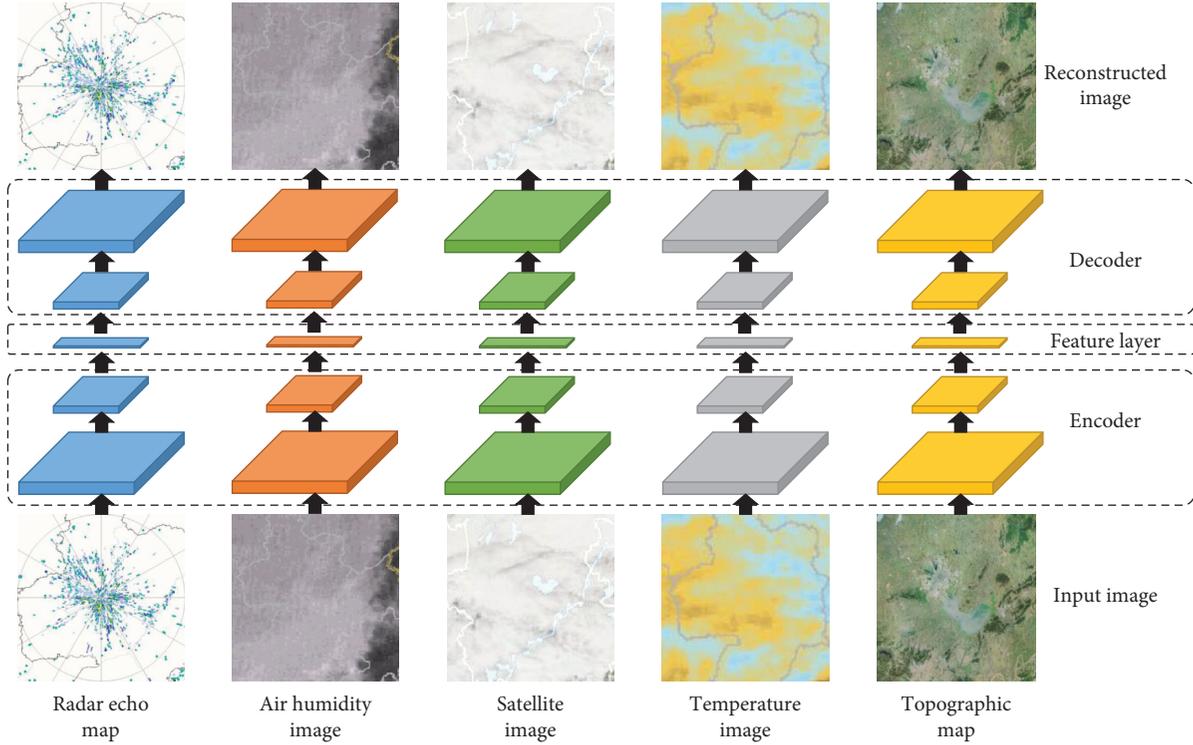


FIGURE 3: Architecture of the convolutional autoencoder used for feature representation learning in our framework.

where  $s_{i,j}$  is the similarity of two nodes  $v_i$  and  $v_j$  and  $\| * \|_2$  is the Euclidean distance of the two node representations.

Therefore, let  $G(\mathbf{H}, \mathbf{A})$  be the graph representation of graph  $G(\mathbf{V}, \mathbf{E})$ , which can be formulated as

$$\mathbf{H} = \begin{bmatrix} \mathbf{f}_{re_{v_1}} & \mathbf{f}_{si_{v_1}} & \mathbf{f}_{tp_{v_1}} & \mathbf{f}_{ti_{v_1}} & \mathbf{f}_{ah_{v_1}} \\ \mathbf{f}_{re_{v_2}} & \mathbf{f}_{si_{v_2}} & \mathbf{f}_{tp_{v_2}} & \mathbf{f}_{ti_{v_2}} & \mathbf{f}_{ah_{v_2}} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{f}_{re_{v_{N \times N}}} & \mathbf{f}_{si_{v_{N \times N}}} & \mathbf{f}_{tp_{v_{N \times N}}} & \mathbf{f}_{ti_{v_{N \times N}}} & \mathbf{f}_{ah_{v_{N \times N}}} \end{bmatrix}, \quad (6)$$

$$\mathbf{A} = \begin{bmatrix} s_{1,1} & s_{1,2} & \dots & s_{1,N} \\ s_{2,1} & s_{2,2} & \dots & s_{2,N} \\ \dots & \dots & \dots & \dots \\ s_{N,1} & s_{N,2} & \dots & s_{N,N} \end{bmatrix}.$$

**3.2.2. Precipitation Nowcasting Model Learning.** Based on the graph representation mentioned above, the observation data of different sampled times can be deemed as a graph sequence, as shown in Figure 4. Therefore, the precipitation nowcasting task is converted to a sequence graph modeling problem, which aims to obtain the precipitation of each node according to a series of graphs.

To address it, as illustrated in Figure 5, a multimodal semisupervised deep graph learning framework is proposed as a solution. A GCN and LSTM are used to model spatial and temporal information, respectively.

GCN contains one input layer, several propagation (hidden) layers, and one final perceptron layer by [22].

Given an input  $\mathbf{H}$  and an adjacency matrix  $\mathbf{A}$ , the GCN conducts the following layerwise propagation in the hidden layer as

$$\mathbf{H}^{k+1} = \sigma(\mathbf{D}^{-(1/2)} \mathbf{A} \mathbf{D}^{-(1/2)} \mathbf{H}^k \boldsymbol{\theta}^k), \quad (7)$$

where  $k = 0, 1, \dots, K-1$ .  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$  is a diagonal matrix with  $d_i = \sum_{j=1}^n \mathbf{A}_{ij}$ .  $\boldsymbol{\theta}^k \in \mathbb{R}^{d_k \times d_{k+1}}$ ,  $d_0 = p$  is a layer-specific weight matrix needing to be trained.  $\sigma$  denotes a nonlinear activation function, and  $\mathbf{H}^{k+1} \in \mathbb{R}^{n \times d_{k+1}}$  denotes the activation output in the  $k+1$ th layer.

As each row in  $\mathbf{H}$  represents a local area, the precipitation can be predicted as follows:

$$\mathbf{P}'_t = \text{Sigmoid}(h(\mathbf{H}'_t)), \quad (8)$$

where  $h$  is a multilayer fully connected neural network. It should be noted that equation (8) is a regression formulation, which is different from the commonly used semisupervised node classification, each row in  $\mathbf{H}$  represents a node, and GCN defines the final perceptron layer as  $\text{softmax}(\mathbf{D}^{-(1/2)} \mathbf{A} \mathbf{D}^{-(1/2)} \mathbf{H}^k \boldsymbol{\theta}^k)$  to classify all nodes.

Considering that  $\mathbf{H}^0$  is crucial for computing equation (7), we design a fusion unit to integrate the five features  $(\mathbf{f}_{re_{v_i}}, \mathbf{f}_{si_{v_i}}, \mathbf{f}_{tp_{v_i}}, \mathbf{f}_{ti_{v_i}}, \mathbf{f}_{ah_{v_i}})$  into one single hidden feature representation. Here, we apply a fully connected layer using a hyperbolic tangent nonlinearity activation function. The detailed transformations are as follows:

$$\mathbf{H}^0 = \tanh\left(W_f \left[ \mathbf{f}_{re_{v_i}}, \mathbf{f}_{si_{v_i}}, \mathbf{f}_{tp_{v_i}}, \mathbf{f}_{ti_{v_i}}, \mathbf{f}_{ah_{v_i}} \right]\right), \quad (9)$$

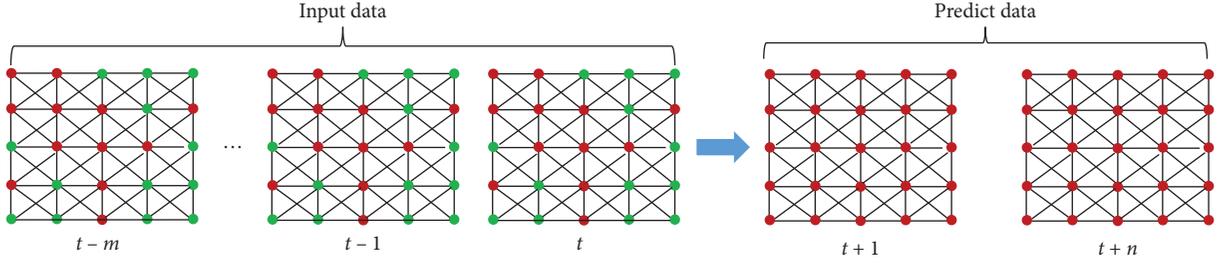


FIGURE 4: Illustration of the precipitation nowcasting which is deemed as a sequence graph modeling problem in this paper.

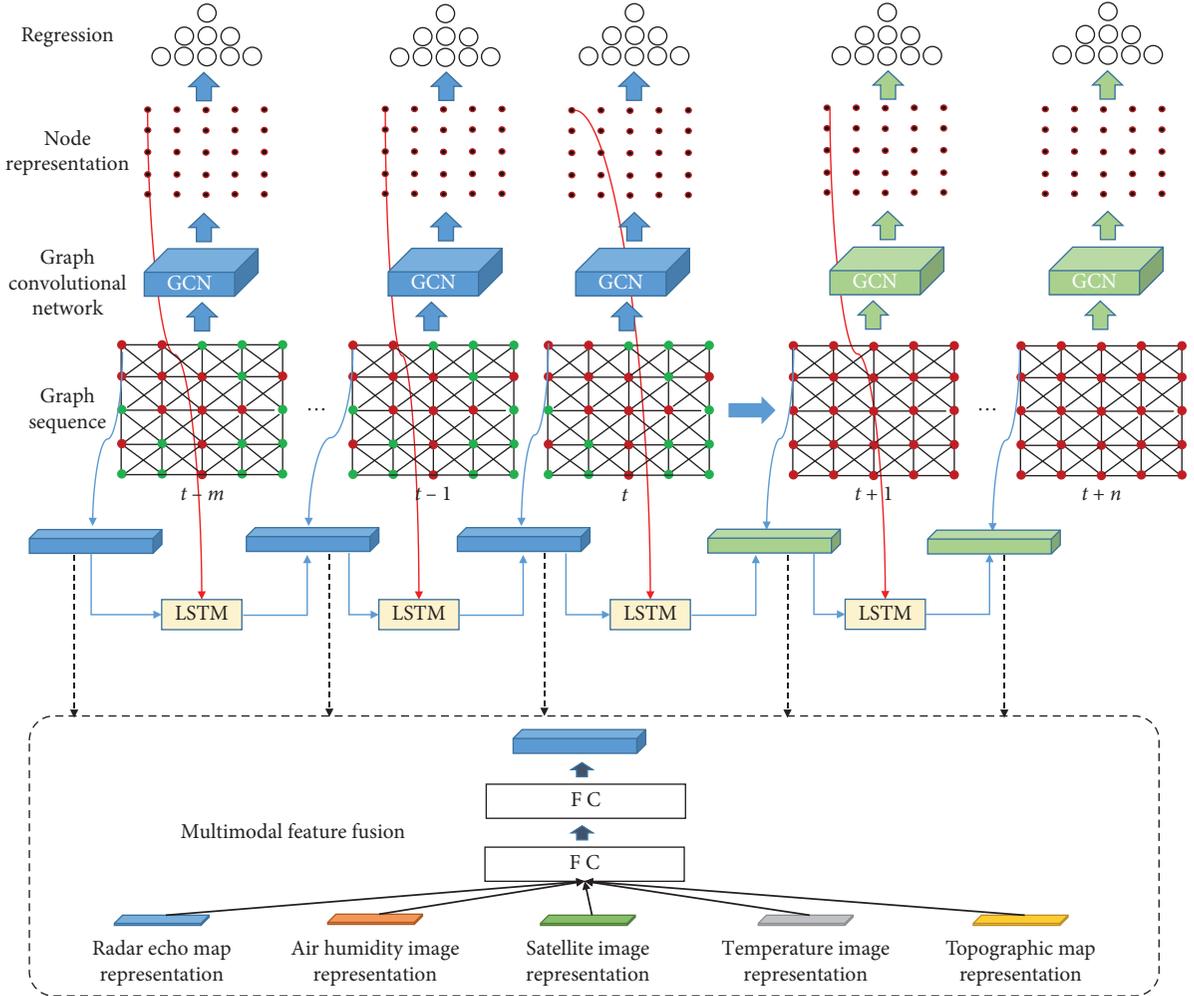


FIGURE 5: The framework of the proposed precipitation nowcasting model. The red nodes mean the precipitation of the corresponding local area is available while the green means the precipitation is unknown.

where  $W_f$  is the transformation parameter, that is, the weights of the fully connected layer, which are learned automatically in the training. This fusion unit is capable of learning the weights for different types of the features.

It is well known that precipitation is related to not only the current weather but also meteorological conditions over the past period. To handle this temporal influence, LSTM is utilized to establish relationships between the node at time  $t$  and the nodes before time  $t$ , which is formulated as

$$\mathbf{H}_t^0 = h(\mathbf{H}_{t-1}^0, \mathbf{H}_{t-1}^K, \mathbf{H}_t^0), \quad (10)$$

where  $h$  is a LSTM-based sequence model.

The optimal weight parameters are trained by minimizing the following loss function over all labeled nodes  $L$ , i.e.,

$$\mathcal{L} = - \sum_{i \in L} \sum_{j=0}^m \left\| \mathbf{P}_{(t-j)_i} - \mathbf{P}'_{(t-j)_i} \right\|, \quad (11)$$

where  $L$  indicates a set of the labeled nodes and  $j$  denotes the index of all observation graphs.

After the model training is completed, the future precipitation is easy to compute because  $\mathbf{y}_{t+n} = \mathbf{P}'_{t+n}$ .

## 4. Experiments

In this section, we first evaluate the proposed method on a large-scale real meteorological dataset. Then, we demonstrate extensive experimental results and analysis.

**4.1. Dataset.** The dataset used in the benchmark contains radar echo images, satellite images from 2016 to 2018, and topographic maps. The radar reflectivity images at 0.5° elevation with a resolution of  $416 \times 416$  pixels were collected from local new-generation weather radar stations (Hefei, Bengbu, and Huangshan) and cover a local area centered in each radar station. The topographic maps were collected from Baidu maps. Infrared and vapor channel images of FengYun-2 meteorological satellite images were chosen to increase the accuracy of precipitation nowcasting. The topographic maps and satellite images have the same spatial resolution and cover the same areas. Thus, a large-scale real meteorological dataset was obtained (partial raw data can be queried from <http://www.amo.org.cn/tcsj/zh/index.jsp>). All data were checked by five volunteers. Any sample data from consecutive 27 h can constitute a whole training sample (this paper aims to forecast precipitation at a place for the next 3 h based on the weather conditions of its neighbors over the past 24 h. Moreover, to match the sampling time, the precipitation herein is accumulated over the past 20 min). For each local radar station, we randomly selected 5000 samples as the training set, 500 samples as the validation set, and 2000 samples as the test set.

**4.2. Evaluation Metrics.** We utilize mean-square error (MSE) to evaluate model's effectiveness. The formulation is as follows:

$$\text{MSE} = \frac{1}{Q} \sum_{n=1}^N \|\mathbf{I}_{\text{pm}} - \mathbf{I}'_{\text{pm}}\|_2, \quad (12)$$

where  $Q$  is the number of test samples and smaller MSE means better performance.

Furthermore, to quantitatively evaluate our model's effectiveness influenced by the time variations, we compute MSE in different sampling time periods in the test phase:

$$\text{MSE}_{i \in \{0,8\}} = \frac{1}{Q} \left( \frac{1}{M \times N} \|\mathbf{I}_{\text{pm}_i} - \mathbf{I}'_{\text{pm}_i}\|_2 \right), \quad (13)$$

where  $i$  is the index of the  $i$ th precipitation map.

**4.3. Implementation Details.** We trained the proposed model from scratch on the training set and considered the model with the smallest mean-square error on the validation set as our final model. All images were normalized to  $[-1, +1]$ . Note that images from different sources have different resolutions. Therefore, we aligned and cropped all images

before the training. We first selected two key points as anchors and resized all images based on the anchor distance of each image. Then, we cropped all images to  $416 \times 416$  pixels based on the radar longitude and latitude. ResNet18 was utilized to extract feature representations of all images, and the feature dimension was set to 256 (that is, the average pooling layer of the ResNet18 was replaced by a 256-d fully connected layer).

The open source deep learning framework PyTorch was employed to implement the proposed semisupervised deep regression model. The parameter initialization for all layers was Xavier, and we used the Adam optimizer with an initial learning rate of 0.001.

## 4.4. Results and Discussion

**4.4.1. Ablation Study.** Table 1 presents the mean-square errors of different time nodes of the three radar stations. We can observe that the proposed framework predicts the precipitation in the next hour well. It also clearly shows that the error increases with time. Moreover, we can see that our model has the worst performance for Huangshan. We believe that the reason is that the causes of precipitation in mountainous areas are more complicated than elsewhere.

To the best of our knowledge, this is the first attempt to model various types of the observed data in a unified network. We describe the performed ablation study below to explore the performance influenced by different factors.

The inputs have five forms:

re: inputs that only include radar echo maps

re + si: inputs that include radar echo maps and satellite images

re + si + ah: inputs that include radar echo maps, satellite images, and air humidity images

re + si + ah + ti: inputs that include radar echo maps, satellite images, air humidity images, and temperature images

re + si + ah + ti + tm: inputs that include radar echo maps, satellite images, topographic map, temperature images, and air humidity images

From Table 2, we can see that all observed data used in this paper contribute to short-term precipitation nowcasting. Table 2 also shows that the topographic map usage improves the accuracy slightly, as compared to the other four factors. We think that the reason is that all other observed data are influenced by the topographic map. Therefore, the features extracted from the other factors represent the topographic map implicitly.

**4.4.2. Comparison with Existing Methods.** We note that some deep learning-based precipitation nowcasting methods have recently been proposed, e.g., [28–30]. However, they are all focused on supervised tasks. As we attempt to address the semisupervised automatic precipitation nowcasting problem, we compare our method with the two widely used semisupervised node representation methods,

TABLE 1: Mean-square errors (MSEs) of different time nodes in three radar stations.

Radar stations	MSE0	MSE1	MSE2	MSE3	MSE4	MSE5	MSE6	MSE7	MSE8	MSE
Hefei	0.06	0.11	0.19	0.28	0.59	1.17	1.93	2.61	3.6	1.17
Bengbu	0.04	0.08	0.15	0.29	0.64	1.12	1.86	2.57	3.16	1.1
Huangshan	0.07	0.12	0.22	0.40	0.92	1.58	2.11	2.86	3.74	1.34

TABLE 2: Mean-square errors (MSEs) of different methods in three radar stations.

Method	MSEs of different radar stations		
	Hefei	Bengbu	Huangshan
ICA by [27]	1.33	1.24	1.59
DeepWalk by [19]	1.67	1.46	2.01
Our (re)	3.62	3.42	3.98
Our (re + si)	2.48	2.51	2.64
Our (re + si + ah)	1.56	1.51	1.72
Our (re + si + ah + ti)	1.19	1.18	1.41
Our (re + si + ah + ti + tm)	<b>1.17</b>	<b>1.10</b>	<b>1.34</b>

e.g., DeepWalk by [19] and ICA by [27]. Specifically, the GCN utilized in the proposed framework is replaced by these two methods. The results are reported in Table 2. It can be observed that the proposed method achieves the best performance. We think that the major reason is that the proposed framework can handle the spatial (local region) relationship more effectively.

## 5. Conclusion and Future Work

In this paper, we propose a multimodal semisupervised deep graph learning framework for precipitation nowcasting. Particularly, multimodal factors, i.e., the radar echo maps, air humidity images, satellite images, temperature images, and available precipitation data over the past several hours with the corresponding topographic map are handled in a unified framework. In addition, we handle areas without precipitation stations, and thus, with unknown precipitation. A GCN with sequence information is presented herein this paper to model temporal and spatial information with semisupervised labels simultaneously. Using the methods above, we successfully implemented efficient precipitation nowcasting via a computer vision technique. To verify our method, we built large-scale real precipitation nowcasting datasets. Extensive experimentation demonstrated that our approach achieves superior performance to the baselines. In the future, we will explore how to embed attention mechanisms in our framework, which may improve the accuracy further.

## Data Availability

The data used to support the findings of this study are available at <http://data.cma.cn/site/index.html>.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Authors' Contributions

Kaichao Miao and Wei Wang contributed equally to this work and should be considered co-first authors.

## Acknowledgments

This work was supported by the Anhui Provincial Natural Science Foundation (no. 2008085QF295), Scientific Research Development Foundation of Hefei University (no. 19ZR15ZDA), and Talent Research Foundation of Hefei University (no. 18-19RC54).

## References

- [1] S. Otsuka, G. Tuerhong, R. Kikuchi et al., "Precipitation nowcasting with three-dimensional space-time extrapolation of dense and frequent phased-array weather radar observations," *Weather and Forecasting*, vol. 31, no. 1, pp. 329–340, 2016.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [3] J. Gao, T. Zhang, and C. Xu, "I know the relationships: zero-shot action recognition via two-stream graph convolutional networks and knowledge graphs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8303–8311, Vancouver, Canada, July 2019.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [5] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, Honolulu, HI, USA, July 2017.
- [6] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5533–5541, Venice, Italy, October 2017.
- [7] C.-C. Chiu, T. N. Sainath, Y. Wu et al., "State-of-the-art speech recognition with sequence-to-sequence models," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Calgary, Canada, pp. 4774–4778, April 2018.
- [8] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 802–810, Montreal, Canada, December 2015.
- [9] X. Shi, Z. Gao, L. Lausen et al., "Deep learning for precipitation nowcasting: a benchmark and a new model," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 5617–5627, Long Beach, CA, USA, December 2017.
- [10] S. Singh, S. Sarkar, and P. Mitra, "Leveraging convolutions in recurrent neural networks for doppler weather radar echo

- prediction,” in *Proceedings of the International Symposium on Neural Networks*, Springer, Hokkaido, Japan, pp. 310–317, June 2017.
- [11] Z. Karevan and J. A. Suykens, “Spatio-temporal stacked LSTM for temperature prediction in weather forecasting,” 2018, <http://arxiv.org/abs/1811.063411>.
- [12] E. Shi, Q. Li, D. Gu, and Z. Zhao, “A method of weather radar echo extrapolation based on convolutional neural networks,” in *Proceedings of the International Conference on Multimedia Modeling*, Springer, Bangkok, Thailand, pp. 16–28, February 2018.
- [13] Pechyonkin, *Deep Learning Vision for Non-Vision Tasks*, 2018, <https://pechyonkin.me/deep-learning-vision-non-vision-tasks>.
- [14] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [15] W.-c. Woo and W.-k. Wong, “Operational application of optical flow techniques to radar-based rainfall nowcasting,” *Atmosphere*, vol. 8, no. 3, p. 48, 2017.
- [16] X. Zhu, Z. Ghahramani, and J. D. Lafferty, “Semi-supervised learning using gaussian fields and harmonic functions,” in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 912–919, Washington, DC, USA, August 2003.
- [17] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: a geometric framework for learning from labeled and unlabeled examples,” *Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.
- [18] J. Weston, F. Ratle, H. Mobahi, and R. Collobert, “Deep learning via semi-supervised embedding,” in *Neural Networks: Tricks of the Trade*, pp. 639–655, Springer, Heidelberg, Germany, 2012.
- [19] B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk: online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, pp. 701–710, August 2014.
- [20] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: large-scale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, Florence, Italy, pp. 1067–1077, May 2015.
- [21] A. Grover and J. Leskovec, “Node2vec: scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, San Francisco, CA, USA, pp. 855–864, August 2016.
- [22] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, <http://arxiv.org/abs/1609.02907>.
- [23] Z.-M. Chen, X.-S. Wei, P. Wang, and Y. Guo, “Multi-label image recognition with graph convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5177–5186, Long Beach, CA, USA, June 2019.
- [24] X. Wang, Y. Ye, and A. Gupta, “Zero-shot recognition via semantic embeddings and knowledge graphs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6857–6866, Salt Lake City, UT, USA, June 2018.
- [25] L. Wu, P. Sun, R. Hong, Y. Fu, X. Wang, and M. Wang, “Socialgcn: an efficient graph convolutional network based model for social recommendation,” 2018, <http://arxiv.org/abs/1811.02815>.
- [26] L. Yao, C. Mao, and Y. Luo, “Graph convolutional networks for text classification,” 2018, <http://arxiv.org/abs/1809.05679>.
- [27] Q. Lu and L. Getoor, “Link-based classification,” in *Proceedings of the 20th International Conference on Machine Learning*, ICML-03, Washington, DC, USA, pp. 496–503, August 2003.
- [28] Q.-K. Tran and S.-k. Song, “Computer vision in precipitation nowcasting: applying image quality assessment metrics for training deep neural networks,” *Atmosphere*, vol. 10, no. 5, p. 244, 2019.
- [29] J. Jing, Q. Li, and X. Peng, “MLC-LSTM: exploiting the spatiotemporal correlation between multi-level weather radar echoes for echo sequence extrapolation,” *Sensors*, vol. 19, no. 18, p. 3988, 2019.
- [30] Q.-K. Tran and S.-k. Song, “Multi-channel weather radar echo extrapolation with convolutional recurrent neural networks,” *Remote Sensing*, vol. 11, no. 19, p. 2303, 2019.