2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Context-Aware Attention LSTM Network for Flood Prediction

Yirui Wu[†‡*], Zhaoyang Liu[‡*], Weigang Xu[†], Jun Feng[†], Shivakumara Palaiahnakote[§], and Tong Lu[‡]
[†]College of Computer and Information, Hohai University, Nanjing, China
[‡]National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China
[§]Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
{wuyirui@hhu.edu.cn;zyliumy@gmail.com;wei_gangxuu@163.com;fengjun@hhu.edu.cn;
shiva@um.edu.my;lutong@nju.edu.cn}

Abstract—**To minimize the negative impacts brought by floods, researchers from pattern recognition community utilize artificial intelligence based methods to solve the problem of flood prediction. Inspired by the significant power of Long Short-Term Memory (LSTM) networks in modeling the dynamics and dependencies of sequential data, we intend to utilize LSTM networks to predict sequential flow rate values based on a set of collected flood factors. Since not all factors are informative for flood prediction and the irrelevant factors often bring a lot of noise, we need to pay more attention to the informative ones. However, original LSTM doesn't have strong attention capability. Hence we propose an context-aware attention LSTM (CA-LSTM) network for flood prediction, which is capable to selectively focus on informative factors. During training, the local context-aware attention model is constructed by learning probability distributions between flow rate and hidden output of each LSTM cell. During testing, the learned local attention model assign weights to adjust relations between input factors and predictions at all steps of LSTM network. We conduct experiments on a flood dataset with several comparative methods to demonstrate high accuracy of the proposed method and the effectiveness of the proposed context-aware attention model.**

I. Introduction

As one of the most common and largely distributed natural diasters, flood happens and brings damage. If we could accurately forecast flood by predicting its sequential flow rate values in advance, hundreds of lives and quantity of property could be saved. In the past decade, researchers from both pattern recognition and hydrology community have proposed a variety of methods to construct accurate, and robust flood prediction models. We generally category them into two types, namely hydrology model [1], [2], [3] and data-driven model [4], [5], [6]. The methods in the first group solve highly non-linear systems, which describe the complex hydrology processes from clues to results by functions. However, such methods are extremely sensitive to parameters [7]. Meanwhile, adjusting theses parameters requires special research effort on quantity of historical flood sequences, which prevents the usage of hydrology model for flood prediction in rivers without special research interest, *e.g.* small rivers. The methods in the second group usually estimate the river flow rate based on historical collected flood factors, *e.g* former rainfall, river runoff and so on, without considering the detailed physical processes. Due to the complex mechanism of flood producing, it's efficient and costless to directly learn the relation between flooding cues and flow rates, especially for rivers with few research efforts.

Inspired by the significant performance [8], [9], [10] of Convolutional Neural Networks (CNNs) and LSTMs, we intend to utilize deep learning methods to discover the inherent relations between flood factors and flow rates. However, floods don't happen frequently, which leads to the small size of flood dataset when utilizing a flood as a training sample. Since deep learning methods generally require large set of samples to train for an highly effective feature representation, it's difficult to directly deploy deep learning methods on flood prediction. Moreover, the collected factors are not all representative and informative for flood prediction. For example, the water retained in soil has great effect on floods in humid area, while it's not important for flood prediction in dry places [3]. This is due to different capacities to contain water corresponding to various types of soil, *i.e.* high and low capacity for soil in humid and dry areas respectively. The informativeness degree of a flood factor may vary over different floods and even over different time points in the same flood. Therefore, it is beneficial to selectively focus on the informative factors at important time points and try to ignore the irrelevant ones, since the latter contribute very little for flood prediction, and even bring in noise that can decrease the accuracy of flood prediction. This selectively focusing mechanism is known as attention model, which has been demonstrated to be very effective in various applications, such as speech recognition [11], action recognition [12] and so on.

In this paper, we propose an context-aware attention LSTM (CA-LSTM) network for accurate and robust flood prediction, which has strong attention ability for flood prediction. We first transform the original problem to predicting flow rates at time points, which enables to use flow rates and flood factors at various time points as training samples, thus increases the size of dataset to fit the requirement of deep learning methods. In the proposed CA-LSTM network, we describe the context-aware attention by constructing several weight schemes to assign weights based on both flood factors and time points. During training, the local attention information is extracted by learning probability distributions between flow

* indicates equal contribution.

rate and hidden output (representing contextual information [13]) of each LSTM cell. During testing, the local attention information, i.e. parameters of the learned distributions at all steps, is fed to the network, thus the network can use it to measure the informativeness weights of the new input factors at all steps.

The main contribution of the paper is to propose a context-aware attention LSTM network for flood prediction, which retains the sequential modeling ability of the original LSTM, meanwhile promoting its selective attention ability. To the best of our knowledge, this is the first LSTM architecture with explicit local context-aware attention as its fundamental capability for flood prediction. The proposed context-aware attention model is simple but effective, which fits the problem of flow rate prediction by constructing probability distribution to represent the local context. The learned distribution is further utilized as a weight scheme to clearly show the informative factors at each time points of flood sequence, which is reasonable to help discover and analyze inherent patterns between factors and flow rate, especially for regions whose flood mechanism is too complex to build a physical hydrology model, such as small rivers.

II. Related Work

Considering the relevance to the proposed method, we detailly describe the data-driven model and context model in this section.

Data-driven Model. With the development of artificial intelligence technologies, researchers from the machine learning community have proposed a quantity of methods to predict flood, including Bayesian-based methods [4], SVM [14], Neural Network [15], deep learning methods [5], [6], [16] and so on.

Early, Reggiani *et al.* [17] construct a modified Bayesian predicting system by involving numerical weather information to address the spatial-temporal variabilities of precipitation during prediction. Yu *et al.* [14] utilize the support vector machine to establish a real-time forecasting model by applying a two-step grid search method to find the optimal parameters for SVM. Later, cheng *et al.* [18] perform accurate daily runoff forecasting by proposing an artificial neural network based on quantum-behaved particle swarm optimization, which trains the ANN parameters in an alternative way and achieves much better forecast accuracy than the basic ANN model.

Due to high potentials of discovering effective features from data, researchers utilize deep learning architectures for flood prediction. For example, Bai *et al.* [16] propose a multi-scale deep feature learning structure with hybrid models to handle the daily reservoir inflow forecasting. In their hybrid model, ensemble empirical mode decomposition and Fourier spectrum are first employed to extract multi-scale features, which are then represented by three deep belief networks (DBNs) respectively. Zhuang *et al.* [5] design a novel Spatio-Temporal Convolutional Neural Network (ST-CNN) to fully utilize the spatial and temporal information and automatically learn underlying patterns from data for extreme flood cluster prediction. Liu *et al.* [6] proposes a deep learning approach by integrating stacked auto-encoders (SAE) and back propagation neural networks (BPNN) for the prediction of stream flow, which simultaneously takes advantages of the powerful feature representation capability of SAE and superior predicting capacity of BPNN. Unlike the deep learning methods mentioned above, the proposed method performs context-aware attention over all steps of the LSTM network to emphasize the importance of informative factors for different kinds of flood sequences.

Attention Model. Human perception focuses selectively on parts of the scene to acquire information at specific places and times. This kind of processes is named as attention model, which has drawn increasing attentions to deal with languages, images and other data. Early, Itti *et al.* [19] incorporate the attention model for object detection by modeling it as saliency maps, i.e. pixelwise weighting of image parts that locally stand out, without learning process.

For deep neural networks, attention models is constructed as a dimension of interpretability into their internal representations by selectively focusing on specific information, when performing a particular task. Recently, attention models are gradually categorized into two classes, *i.e.* hard attention [20] and soft attention [21]. Hard attention makes hard decisions on choosing parts of the input data as focuses, which results in improper algorithms to be learned through gradient descent and back-propagation. Mnit *et al.* [22] propose a recurrent neural network model that is capable to extract information from an image by adaptively selecting a sequence of regions and only processing the selected regions at high resolution.

On the contrary, soft attention takes the entire input into account by weighting each part or step of the observations dynamically. The objective function is usually differentiable, making gradient-based optimization possible. Sharma *et al.* [12] proposed a recurrent mechanism for action recognition from RGB data, which integrates convolutional features from different parts of a space-time volume. Yeung *et al.* [23] report a temporal recurrent attention model for dense labelling of videos [43], which integrates multiple input frames and soft predictions generated for multiple frames at each time step. Recently, Liu *et al.* [13] propose Global Context-Aware Attention LSTM for 3D action recognition, which recurrently optimize global contextual information and further utilize it as informative functions to assist accurate action recognition. The proposed context-aware attention model could be classified into soft attention using a recurrent scheme to optimize. It describes local context information as a form of weight scheme, which is reasonable to understand and analyze the regression problem of flow rate prediction.

III. The Proposed Method

In this subsection, we firstly describe the construction of the proposed context-aware attention LSTM. Then, we describe the context-aware attention Model in detail, which is capable of selectively focusing on the informative factors and time points in the flood sequence.

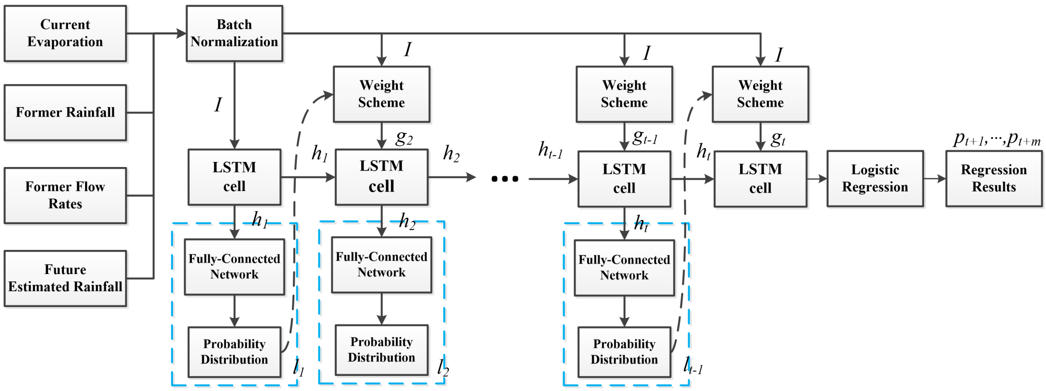

Fig. 1. Illustration of the proposed context-aware attention LSTM network for flood prediction, where blue rectangles indicates the local context-aware attention model.

Fig. 2. The map of various types of stations in Changhua river. Note that Station CH is not only a rainfall station, but also a river gauging station whose river flow needs to be predicted. The station SS is a rainfall station and evaporation station.

A. *Construction of CA-LSTM*

Take Changhua river as an example, we describe the steps to predict and analyze the hourly flood using the proposed method. The general information about Changhua watershed is shown in Fig. 2, in which we can see 7 rainfall stations, 1 evaporation station and 1 river gauging station. In general, we aim to predict the flow rate at the river gauging station CH for next m hours (represented by $\{p_{t+1}, ..., p_{t+m}\}$) by utilizing various flood factors, including rainfalls observed at the rainfall stations from former k hours, current evaporation observed at the evaporation station SS, flow rates observed at CH from former k hours and the overall estimated ranifall for next m hours achieved directly from weather report. With the objective of flow rate prediction, we design the proposed context-aware attention LSTM network as shown in Fig. 1. The input of the proposed LSTM network is a feature achieved by concatenating all former mentioned flood factors. Then, we adopt the layer of batch normalization to process the input, in order to accelerate training with higher learning rates and less careful parameter initialization [24]. We then follow the conventional structure of LSTM cell to build each step of CA-LSTM network. A typical LSTM unit consists of an input gate n_i, a forget gate n_f, an input modulation gate n_g, an output gate o_t, an output state h_t and an internal memory cell state c_t. By utilizing the gating mechanism, the unit can learn and memorize a complex representation for long-term dependencies at memory cell c_t among the input sequence data. More detailed, the representation in c_t is constructed as a combination of former memory information after forgetting and new information generated from input, i.e. $c_t = n_f \bigodot c_{t-1} + n_i \bigodot n_g$, where $\bigodot$ denotes element-wise multiplication.

Inspired by [22] which considers the attention problem as the sequential decision process of how an agent interact with a visual environment, we design the proposed context-aware attention model as defining ”interaction level” with the input feature, where the ”interaction level” is essentially described by weights assigned to variables inside the input feature. Therefore, the normalized input I is fed to all steps of CA-LSTM as the original description for flood factors, which is kindly similar to the visual environment defined in [22]. Meanwhile, the context-aware attention model recurrently defines the corresponding weights for the normalized input I, leads to attentions on informative factors. Such weight scheme for one step thus could be represented as:

$$g_t = I \cdot l_{t-1} \tag{1}$$

where $\cdot$ represents the element-wise multiplication, g_t is the input for next step representing a feature with extracted informative flood factors, l_{t-1} is the learned weight computed by the proposed context-aware attention model. In a global sense, the proposed CA-LSTM network involves combinations of informative flood factors in all steps to achieve a more accuracy and robust flow rate prediction.

Finally, we utilize the logistic regression classifier connected to the last step of CA-LSTM to makes regression about predicted flow rates for next m hours. The smooth L1 loss function [25] to measure the difference between the true flow rate values $\tilde{p}$ and the predicted result p, is defined as

$$loss(\tilde{p}, p) = \frac{1}{n}\sum_{i=0}^{n} s(\tilde{p}_i, p_i) \tag{2}$$

TABLE I
COMPARISON OF RMSE ON CHANGHUA DATASET WITH CA-LSTM AND SEVERAL COMPARATIVE METHODS.

Method	T+1	T+2	T+3	T+4	T+5	T+6	Average
CA-LSTM	37.24	46.44	**52.37**	**61.10**	**74.16**	**88.36**	**59.95**
LSTM	41.71	54.20	66.11	75.50	85.52	96.49	69.97
FCN	**27.30**	**41.88**	56.87	84.32	110.8	125.3	74.41
SVM	180.0	179.8	179.8	179.9	180.1	180.4	180.0

where p_i and $\tilde{p}_i$ are prediction and ground truth vectors at time i containing flow rate values for next m hours respectively, n refer to the size of dataset which utilizes flow rates and flood factors at various time points as training samples and function $s()$ is defined as:

$$s(\tilde{p}_i, p_i) = \begin{cases} 0.5(\tilde{p}_i - p_i)^2 & if|p_i - p_i| < 1 \\ |\tilde{p}_i - p_i| - 0.5 & otherwise \end{cases} \tag{3}$$

We use the back-propagation through time (BPTT) algorithm to minimize the loss function. Note that we adopt smooth L1 loss function, since it make the loss value convergent in a faster and more stable way comparing with using MSE as loss function.

B. Context-aware Attention Model

Previous hydrology work [3] has already shown that there is often a subset of informative factors which are important as they contribute much more to generating floods, while the other ones can be irrelevant (or even noisy) to the flow rate. Consequently, to achieve a high accuracy for flood prediction, we need to identify the informative factors and concentrate more on their features, meanwhile trying to ignore the features of the irrelevant ones, i.e., selectively focusing (attention) on the informative flood factors is beneficial for reliable flood prediction.

Hence we propose to introduce a context-aware attention model to the LSTM network, which holds the local contextual information for the flood prediction and can be fed to each step of LSTM to assist the attention procedure. The context-aware attention model is built around a recurrent neural network, shown as blue rectangles in Fig. 1. At each step, it processes the input data, integrates information over time, i.e. the information contained in parameters, and decides how to weight the flood factors at next time step. The key idea of the proposed attention model stems from the the supposition that we could encode the attention information extracted from training sets by probability distribution and decode such information by sampling and regarding as weights.

Specifically, the hidden output h_t (representing contextual information) is firstly fed into a fully-connected network, transforming h_t into a probability distribution based on the learned parameters of the network. After generating distribution with fully-connected work, we adopt sigmoid activation function to normalize the result distribution from 0 to 1, leading the corresponding sample value to be proper as a weight. Above all, the learned weight l_t, utilized to construct the weight scheme in the last subsection, is subject to the learned distribution, which could be written as follows:

$$l_t \sim d(sig(f_p(h_t; \theta_p))) \tag{4}$$

where d refers to a distribution, function $f_p()$ and $sig()$ represents the fully-connected network and sigmoid activation respectively and θ_p represents the learned parameters during training.

In fact, the proposed context-aware attention model is fit for dynamic environments due to the design of probability description and simple one-layer structure. Moreover, unlike former methods [22] requiring additional training process to learn weights for attention model, the proposed attention model could be optimized as part of CA-LSTM network using BPTT method. For flood prediction, the result weight generated by the proposed model is rational to discover and analyze inherent relations between flood factors and flow rate.

IV. EXPERIMENTS

We collect hourly data of 40 floods happened from 1998 to 2010 in Changhua river as our original dataset. Recall that we design to use flow rates and flood factors at various time points as training samples to increase the size of dataset. After such transformation, the number of samples is increased to 8555. We utilize 8-fold cross validation to evaluate our CA-LSTM network and comparative works. We use standard quality measures, i.e. Relative Mean Square Error (RMSE) for measuring the quality of flood predicting, which could be represented as

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - q_j)^2} \tag{5}$$

where n refers to the number of testing samples, y_j and q_j represent the predicted and observed flow rates respectively. Note that smaller values of MSE imply better performance the predicting achieves.

We train the CA-LSTM network by defining its sequence length as 32 and the dimension of hidden output as 128. The learning rate, weight decay and batch size are settled as 0.00225, 10^{-6} and 100, respectively. We utilize 500 epoches to

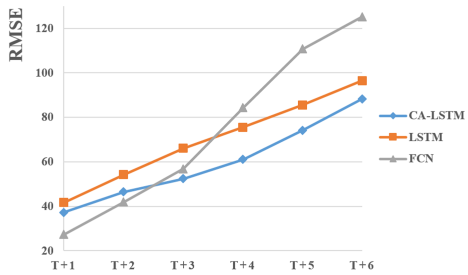

Fig. 3. Comparison of RMSE among CA-LSTM, LSTM and FCN when predicting at different time points.

convergent for the final prediction results. The proposed CA-LSTM network runs on a workstation (2.4GHz 6-core Xeon CPU, 60G RAM, Nvidia GeForce GTX 1080Ti and Ubuntu 64-bit OS) for all the experiments.

Table. I gives the detailed statistics of the proposed CA-LSTM network and several comparative methods for Changhua dataset. Note that we implement the conventional LSTM network, FCN (Fully-Connected Network) and SVM for comparison. The structure and training parameters of LSTM network is exactly the same as CA-LSTM except for the context-aware attention model. The FCN is designed with 10 fully-connected layers and SVM is implemented with the kernel of radial basis function. As shown in Table. I, CA-LSTM network achieves the lowest RMSE values except for prediction at T+1 and T+2. In fact, LSTM is designed to solve the problem of long-term dependencies with the structure of cell memory. In other words, LSTM based network is good at handling and predicting important events with relatively long intervals or delays. We thus observe that FCN performs better than both LSTM and CA-LSTM for prediction at T+1 and T+2, meanwhile performs much worse than LSTM and CA-LSTM for prediction from T+3 to T+6. The difference of RMSE values between CA-LSTM and FCN is as large as 36.94 when predicting time is set as T+6. The RMSE values corresponding to SVM is much larger than these achieved by other three methods, which proves that deep models performs better at flow rate regression problem than SVM. With the context-aware attention model, CA-LSTM network achieves much smaller MSE than that of LSTM network, which proves the efficiency of the proposed context-aware attention model. Essentially, it's rational to focus only on the informative flood factors, since the mechanism of flood in small rivers is too complex to analyze and only few factors contribute to the flood in small rivers.

We show the comparison of RMSE between among CA-LSTM, LSTM and FCN in Fig. 3. We could find that CA-LSTM and LSTM gets similar and stable performance, *i.e.* slight raise, in RMSE when the predicting hour increases from T+1 to T+6, while the RMSE of FCN changes greatly when predicting time is increased. The stable performance of CA-LSTM and LSTM proves that LSTM based network is suitable for long-time prediction by learning and memorizing a complex representation for long-term dependencies at memory cell among the input sequence data.

In Fig. 4, we compare the flow rates prediction results of CA-LSTM, LSTM and FCN with the ground-truth values during a sample period of time. We could see the CA-LSTM achieves nearly the same flow rates as the observed flow rates for prediction at T+2, T+4 and T+6. For LSTM, we could find it get obvious wrong predictions labeled by blue rectangles, when predicting at T+4 and T+6. The areas to view wrong predictions enlarge when predicting with FCN at T+4 and T+6. In fact, we find wrong predictions are easy to occur during flood peak period for both methods. Considering its terrible damage, decision makers mainly concern the predicted flow rates during flood peak. We could conclude LSTM and FCN are not suitable and operable to predict floods of small rivers.

The average time to process a prediction for CA-LSTM is 0.302 ms, which is much smaller than 6.953 ms per prediction achieved by the original LSTM network. The main reason for much shorter processing time lies in the construction of the context-aware attention model, which ignore some input factors for faster computation since these factor are useless for predicting. Moreover, original LSTM need to resize the input for each LSTM layer increasing the number of parameters, while the input of CA-LSTM is always the same, i.e. a feature constructed by concatenating flood factors. The involve of context-aware attention model also results in faster convergency and lower probability to overfit, since additional attention information helps the network to be more task-specified.

V. Conclusions

In this paper, we extend the original LSTM network to achieve a context-aware attention LSTM network for flood prediction, which is capable to selectively focus on informative flood factors. In the proposed CA-LSTM network, we describe the context-aware attention by constructing several weight schemes to assign weights based on both flood factors and time points. During training, the local context- aware attention model is constructed by learning probability distributions between flow rate and hidden output of each LSTM cell. During testing, the learned local attention model assign weights to adjust relations between input factors and predictions at all steps of LSTM network. Experiment results on the Changhua dataset show the proposed method outperforms several comparative methods and the effectiveness of the proposed context-aware attention model. Our future work includes the exploration on other hydrology purposes with the proposed method, such as mid-term flood predicting and flood frequency analysis.

Acknowledgment

This work was supported by the Natural Science Foundation of China under Grant 61702160, Grant 61370091, Grant 61672273, the Fundamental Research Funds for the Central Universities under Grant 2016B14114, the Science Foundation of JiangSu under Grant BK20170892, the Science Foundation for Distinguished Young Scholars of Jiangsu under Grant BK20160021 and the open Project of the National

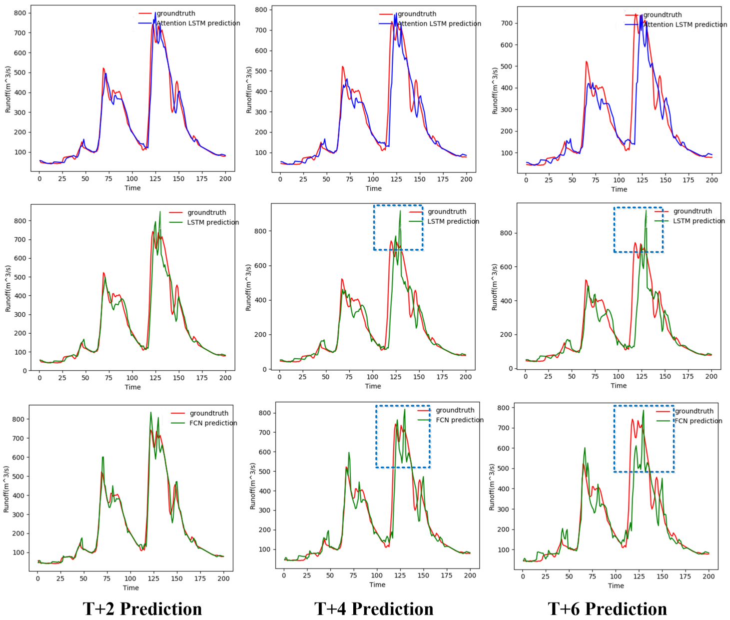

Fig. 4. Comparison with the ground truth flow rates and predicted flow rates computed by CA-LSTM, LSTM and FCN, where top, middle and bottom row represents prediction results of CA-LSTM, LSTM and FCN respectively. Note that the blue rectangles indicates several obvious wrong predictions.

Key Lab for Novel Software Technology in NJU under Grant KFKT2017B05.

REFERENCES

[1] E. Paquet, F. Garavaglia, R. Garçon, and J. Gailhard, "The schadex method: A semi-continuous rainfall–runoff simulation for extreme flood estimation," *Journal of Hydrology*, vol. 495, pp. 23–37, 2013.

[2] M. Rogger, A. Viglione, J. Derx, and G. Blöschl, "Quantifying effects of catchments storage thresholds on step changes in the flood frequency curve," *Water Resources Research*, vol. 49, no. 10, pp. 6946–6958, 2013.

[3] Z. Ren-Jun, "The xinanjiang model applied in china," *Journal of hydrology*, vol. 135, no. 1-4, pp. 371–381, 1992.

[4] S. Han and P. Coulibaly, "Bayesian flood forecasting methods: A review," *Journal of Hydrology*, 2017.

[5] W. Y. Zhuang and W. Ding, "Long-lead prediction of extreme precipitation cluster via a spatiotemporal convolutional neural network," in *Proceedings of the 6th International Workshop on Climate Informatics: CI*, 2016.

[6] F. Liu, F. Xu, and S. Yang, "A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network," in *Proceesdings of IEEE International Conference on Multimedia Big Data*, 2017, pp. 58–61.

[7] C. Yao, K. Zhang, Z. Yu, Z. Li, and Q. Li, "Improving the flood prediction capability of the xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph," *Journal of hydrology*, vol. 517, pp. 1035–1048, 2014.

[8] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceesdings of IEEE CVPR*, 2016, pp. 779–788.

[9] Y. Wu, W. Wang, S. Palaiahnakote, and T. Lu, "A robust symmetry-based method for scene/video text detection through neural network," in *Proceesdings of International Conference on Document Analysis and Recognition*, 2017.

[10] L. Wei, Y. Wu, W. Wang, and T. Lu, "A novel 3d human action recognition framework for video content analysis," in *Proceesdings of International Conference on Multimedia Modeling*, 2018.

[11] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Proceedings of NIPS*, 2015, pp. 577–585.

[12] S. Sharma, R. Kiros, and R. Salakhutdinov, "Action recognition using visual attention," *CoRR*, vol. abs/1511.04119, 2015.

[13] J. Liu, G. Wang, P. Hu, L. Duan, and A. C. Kot, "Global context-aware attention LSTM networks for 3d action recognition," in *Proceedings of IEEE CVPR*, 2017, pp. 3671–3680.

[14] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3, pp. 704–716, 2006.

[15] G. Corani and G. Guariso, "Coupling fuzzy modeling and neural networks for river flood prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 382–390, 2005.

[16] Y. Bai, Z. Chen, J. Xie, and C. Li, "Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models," *Journal of Hydrology*, vol. 532, pp. 193–206, 2016.

[17] P. Reggiani and A. Weerts, "Probabilistic quantitative precipitation forecast for flood prediction: An application," *Journal of Hydrometeorology*, vol. 9, no. 1, pp. 76–95, 2008.

[18] C.-t. Cheng, W.-j. Niu, Z.-k. Feng, J.-j. Shen, and K.-w. Chau, "Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization," *Water*, vol. 7, no. 8, pp. 4232–4246, 2015.

[19] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[20] S. Yeung, O. Russakovsky, G. Mori, and L. Fei-Fei, "End-to-end learning of action detection from frame glimpses in videos," in *Proceedings of IEEE CVPR*, 2016, pp. 2678–2687.

[21] F. Baradel, C. Wolf, and J. Mille, "Pose-conditioned spatio-temporal attention for human action recognition," *CoRR*, vol. abs/1703.10106, 2017.

[22] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proceedings of NIPS*, 2014, pp. 2204–2212.

[23] S. Yeung, O. Russakovsky, N. Jin, M. Andriluka, G. Mori, and F. Li, "Every moment counts: Dense detailed labeling of actions in complex videos," *CoRR*, vol. abs/1507.05738, 2015.

[24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of ICML*, 2015, pp. 448–456.

[25] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.