ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 40, MAY 2023, 887–899

• **Original Paper** •

Convective Storm VIL and Lightning Nowcasting Using Satellite and Weather Radar Measurements Based on Multi-Task Learning Models※

Yang LI[1,2], Yubao LIU*[1,2], Rongfu SUN[3], Fengxia GUO[1], Xiaofeng XU[4], and Haixiang XU[3]

[1]*Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration*, *Nanjing University of Information Science and Technology*, *Nanjing* 210044, *China*

[2]*Precision Regional Earth Modeling and Information Center*, *Nanjing University of Information Science and Technology*, *Nanjing* 210044, *China*

[3]*Jibei Electric Power Company*, *Beijing* 100054, *China*

[4]*China Meteorological Administration*, *Beijing* 100081, *China*

(Received 22 March 2022; revised 8 August 2022; accepted 31 August 2022)

ABSTRACT

Convective storms and lightning are among the most important weather phenomena that are challenging to forecast. In this study, a novel multi-task learning (MTL) encoder-decoder U-net neural network was developed to forecast convective storms and lightning with lead times for up to 90 min, using GOES-16 geostationary satellite infrared brightness temperatures (IRBTs), lightning flashes from Geostationary Lightning Mapper (GLM), and vertically integrated liquid (VIL) from Next Generation Weather Radar (NEXRAD). To cope with the heavily skewed distribution of lightning data, a spatiotemporal exponent-weighted loss function and log-transformed lightning normalization approach were developed. The effects of MTL, single-task learning (STL), and IRBTs as auxiliary input features on convection and lightning nowcasting were investigated.

The results showed that normalizing the heavily skew-distributed lightning data along with a log-transformation dramatically outperforms the min-max normalization method for nowcasting an intense lightning event. The MTL model significantly outperformed the STL model for both lightning nowcasting and VIL nowcasting, particularly for intense lightning events. The MTL also helped delay the lightning forecast performance decay with the lead times. Furthermore, incorporating satellite IRBTs as auxiliary input features substantially improved lightning nowcasting, but produced little difference in VIL forecasting. Finally, the MTL model performed better for forecasting both lightning and the VIL of organized convective storms than for isolated cells.

Key words: convection/lightning nowcasting, multi-task learning, geostationary satellite, weather radar, U-net model

Citation: Li, Y., Y. B. Liu, R. F. Sun, F. X. Guo, X. F. Xu, and H. X. Xu, 2023: Convective storm VIL and lightning nowcasting using satellite and weather radar measurements based on multi-task learning models. *Adv. Atmos. Sci.*, **40**(5), 887–899, https://doi.org/10.1007/s00376-022-2082-6.

Article Highlights:

- A novel multitask learning U-net model was developed for forecasting VIL and lightning, and it outperformed the single-task learning model.
- Spatiotemporal exponent-weighted loss function and a logarithmic normalization were formulated to improve VIL and lightning nowcasting.
- Incorporating satellite infrared brightness temperatures as auxiliary input significantly improves lightning nowcasting.

1. Introduction

Warm-season convective storms frequently produce severe weather conditions, including heavy rainfall, lightning, hail, damaging winds, and tornadoes (Houze et al., 2007; Yao et al., 2015). Accurate and timely convection nowcasting is crucial for mitigating this damage. The intrinsic complexity of convection makes it challenging to forecast its location and intensity. Convection nowcasting approaches can be divided into two classes: radar-echo extrapolation or satellite images (Dixon and Wiener, 1993; Ayzel et al., 2019) and numerical weather models (NWP) (Sun et al.,

※ This paper is a contribution to the special issue on the 14th International Conference on Mesoscale Convective Systems and High-Impact Weather.

* Corresponding author: Yubao LIU
Email: ybliu@nuist.edu.cn

2014). The latter typically requires vast computational resources to integrate mathematical–physical forecasting equations at high temporal and spatial resolutions, making it challenging for the desired frequent update cycles. Radar extrapolation convection nowcasting approaches have been used operationally in many national and regional weather centers.

Lightning is a concomitant phenomenon of convective storms. Lightning forecasting remains a significant challenge in operational meteorology (Sun et al., 2014; Zhou et al., 2020). There are primarily two methods for lightning nowcasting. One is a lightning diagnostic scheme based on NWP outputs that contain electrification parameterizations (Fierro et al., 2013) or cloud microphysical parameters as a proxy for lightning (Tippett and Koshak, 2018). The other is based on an extrapolation scheme using measurements of satellite, weather radar, NWP model output, and lightning location system measurements, alone or combined (Meng et al., 2019). With the rapid increase in modern observation sources, many efforts have been made to explore novel methods to improve convective storms and lightning nowcasting.

Recent advances in machine learning and deep learning have greatly improved data inference ability in the diverse domains of business and science (LeCun et al., 2015). For weather applications, several machine learning models have been developed to represent unresolved physical processes in coarse-scale climate models (Gentine et al., 2018; Rasp et al., 2018; Pal et al., 2019; Seifert and Rasp, 2020; Beucler et al., 2021), improve medium-range weather forecasting (Rasp et al., 2020; Rasp and Thuerey, 2021), and post-process numerical weather model outputs (Krasnopolsky and Lin, 2012; Rasp and Lerch, 2018; Li et al., 2019; Han et al., 2021). Furthermore, deep learning technologies have the potential to obtain the best possible estimate of the current atmospheric state for initializing numerical weather models (Abarbanel et al., 2018; Arcucci et al., 2021; Geer, 2021).

Deep learning methods have been successfully applied to predict the spatiotemporal evolution of convective precipitation based on radar-image extrapolation (Shi et al., 2015; Wang et al., 2017; Sønderby et al., 2020; Wu et al., 2021). Shi et al. (2015) introduced a convolutional long short-term memory neural network model (ConvLSTM) to forecast precipitation using weather radar measurements. Several other deep learning models have been developed for convection precipitation forecasting using weather radar, satellites, or combined observations (Franch et al., 2016; Han et al., 2020, 2021; Ravuri et al., 2021). Mostajabi et al. (2019) applied a machine learning method for lightning nowcasting for up to 30 min using four meteorological parameters: air pressure, air temperature, relative humidity, and wind speed. Zhou et al. (2020) proposed a U-net model for 0–1 h ahead cloud-to-ground (CG) lightning occurrence probabilities using observations from the Himawari-8 geostationary satellite, Doppler weather radar networks, and CG lightning location systems in central-eastern and southern China.

The above deep learning models for convective storms and lightning nowcasting are based on convolutional neural networks (CNN) with single-task learning (STL). STL is a standard method for training a machine learning model for a given task (Shi et al., 2015; Zhou et al., 2020; Han et al., 2021; Liu et al., 2021). Recently, multi-task learning (MTL), which aims to simultaneously optimize multiple objectives using shared weight parameters of neural network models to learn the common patterns of related tasks, has been proposed (Crawshaw, 2020). The MTL exhibits good potential for improving the accuracy and generalization of neural network models (Zhang and Yang, 2018; Zhuo and Tan, 2021). Herein, we explore the MTL for convective storms and lightning nowcasting and compare it to STL models. To the best of our knowledge, this is the first application of the MTL paradigm to convection and lightning nowcasting.

In this study, GOES-16 geostationary satellite infrared brightness temperatures (IRBTs) and lightning flashes (including both intra-cloud and cloud-to-ground lightning events) from the Geostationary Lightning Mapper (GLM), and mosaics of vertically integrated liquid (VIL) from Next Generation Weather Radar (NEXRAD) were used to train a U-net neural network model for 0–90 min lightning and VIL nowcasting, and algorithms including MTL, batch normalization, and cost function strategies were introduced to improve the U-net model for lightning and VIL forecasting. Moreover, the effects of GOES-16 IRBTs on lightning and VIL nowcasting were investigated using the MTL U-net model. The remainder of this paper is organized as follows. In section 2, we introduce the data used for training the neural network models and data preprocessing approaches. Section 3 describes the architecture of the neural network models developed in this study and the specification of the temporal-spatial weighted loss functions. Section 4 presents and analyzes the results of experiments with varying training parameters and input data. Finally, the findings and conclusions of this study are summarized in section 5.

2. Satellite, radar, and lightning data

The GOES-16 6.9 and 10.7 μm infrared (IR) brightness temperature, total lightning flashes of GLM (including intra-cloud and CG lightning), and mosaics of VIL from NEXRAD (Table 1) were collected from the SEVIR (Storm Event ImagRy) dataset (Veillette et al., 2020). A total of 6960 convective events were selected to train the neural network models. Each event in SEVIR is on a 384 km × 384 km patch, selected based on the National Centers for Environmental Information (NCEI) Storm Events Database[a]. Details regarding the SEVIR datasets can be found in the Veillette et al. (2020) study. In this study, each convective event consisted of a 155 min time sequence of images at 5-min time steps in this study. It contained a set of observed variables with fixed dimensions and different geographic location information, as shown in Table 1. The data were split into three subsets: training and validation datasets (4324 and 764 samples, respectively) from February 2018 to May 2019 and a

[a] https://www.ncdc.noaa.gov/stormevents/

test dataset (647 samples) from June to November 2019. Training and validation datasets were used to train the neural network models, and the test dataset was used to evaluate them.

The GOES-16 measurements used in this study are from two IR channels, 6.9 and 10.7 μm, with 2 km grid resolution, the gridded lightning density measured by GLM is with a resolution of 8 km, and the VIL data is at 1−km intervals (Fig. 1). The VIL data were resampled at 2 km grid intervals to align with the IR channel data. To improve the neural network model training, VIL with a range of 0−255 in the SEVIR dataset (including training, validation, and testing datasets) was normalized to 0−1 using the min-max normalization method. The 6.9 and 10.7 μm brightness temperatures were normalized by subtracting their average and dividing by the IRBT standard deviation from the training and validation dataset.

As the lightning flash density data were severely skewed, they were normalized as follows:

$$\log_{lght} = \log_2(lght + 1)\ , \tag{1}$$

where lght is the number of GLM lightning flashes measured in an 8×8 km^2 grid box. Figures 2a and b compare the distribution of lightning flashes normalized by the min-max and Log2Norm normalization methods, respectively. Log2Norm greatly reduces the skewness of min-max normalization.

3. Deep-learning models and experiment design

Convective and lightning nowcasting can be viewed as an extended image-to-image (or video) projection issue. Single-task learning (STL), in which the cost function is defined as a function of the inference task itself, is a standard method adopted for training machine learning models for convection nowcasting by previous researchers. Multi-task learning (MTL) is a subfield of machine learning in which multiple projecting targets are inferred simultaneously with

Table 1. Description of variables was used to train the neural network models.

Image Type	Description	Spatial Resolution (km)	Patch Size
IRBTs 6.9 μm	Mid-level water vapor	2	192 × 192
IRBTs 10.7 μm	Clean longwave window	2	192 × 192
VIL	NEXRAD radar mosaic of VIL	2	192 × 192
Lightning	Intercloud and cloud to ground lightning	8	48 × 48

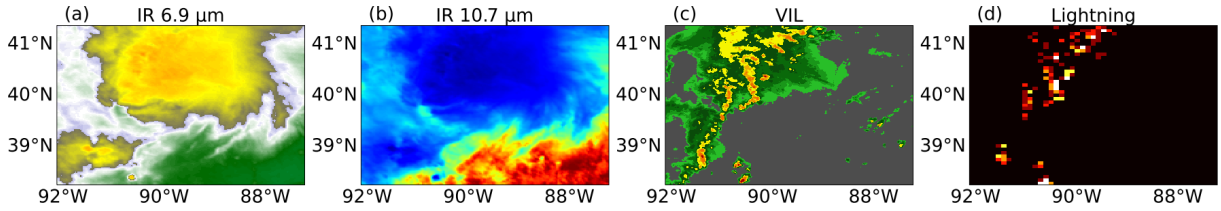

Fig. 1. Measurements of a sample convection storm: (a) 6.9 μm and (b) 10.7 μm infrared brightness temperatures with 2-km grid resolution and 5-min intervals from GOES-16 geostationary satellite measurements, (c) a mosaic of VIL from NEXRAD with 1 km grid resolution, and (d) lightning density at 8 km grid resolution from GLM.

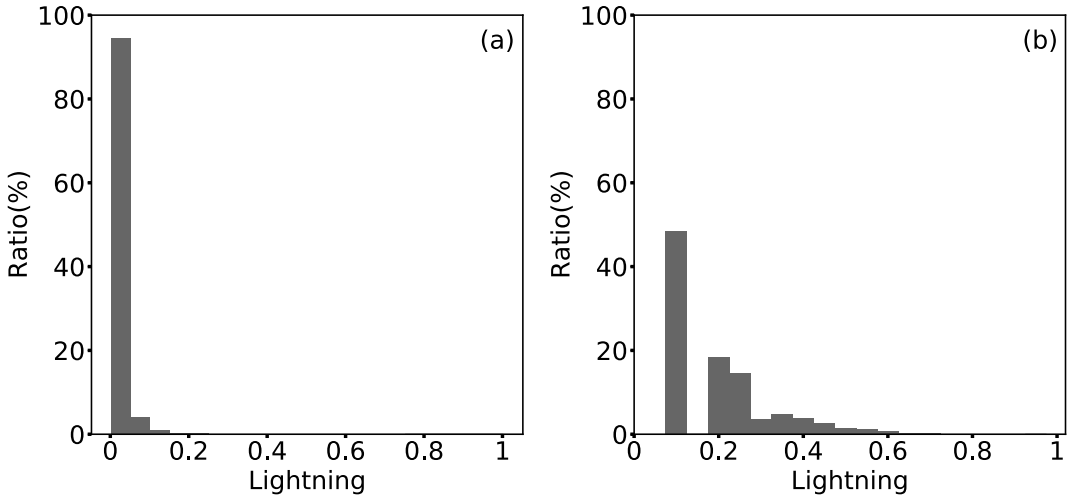

Fig. 2. Distribution of normalized lightning flashes in training dataset by (a) min-max (b) Log2Norm normalization method, respectively.

minimization of the shared weight parameters of a model (Crawshaw, 2020). Optimizing multiple objectives at once, the MTL reduces the cost of running multiple models to complete multiple tasks and has the potential to improve the accuracy of the results. The MTL is particularly effective when the training data are sparse (Zhang and Yang, 2018). In this study, the MTL was constructed along with an encoder-decoder U-Net deep learning model to jointly forecast lightning flash and VIL, and the results were evaluated and compared to those generated with the STL approach that optimized VIL and lightning nowcasting separately.

3.1. *MTL and STL neural network models*

The architecture of the encoder-decoder U-Net model constructed for convection and lightning nowcasting (Fig. 3) is similar to that of Ronneberger et al. (2015). This model consists of convolution (Conv2d), pooling (Maxpool2d), upsampling, batch normalization, and rectified linear unit (i.e., activation function) layers. The convolutional layers are used to extract the key features using a specified number of filters with assigned kernel sizes from the image data and to generate the feature maps. The pooling layers reduce the dimensions of the feature maps and control the possible model overfitting. For example, a Maxpool2d layer of size 2 × 2 takes nonoverlapping 2 × 2 patches of an image and maps each to a single-pixel containing the maximum value of the 2 × 2 patch.

In contrast, the upsampling layers invert operations in the pooling layers. Conv2dTrans layers with a size of 2 × 2 windows were used to expand the image resolution. The upsampling layers are not intended to invert the pooling layers to restore the information lost during the pooling operations, but to simply expand the resolution of an image, that is, restore the data dimension. As small-scale features were lost during the pooling operations, skip connection operations were used to concatenate the large- and small-scale features to maintain multi-scale features during upsampling. The batch normalization layer was used to accelerate neural network model training by transforming the intermediate results of the neural network model to reduce the internal covariate shift (Ioffe and Szegedy, 2015). An activation function (e.g., LeakyReLu) was applied to the feature maps to introduce nonlinearity.

Unlike the original U-net model (Ronneberger et al., 2015), the U-net model configured herein contains three feature extractor modules for VIL, GLM lightning flashes, and both: 1) the VIL feature extractor module (light gray rectangle in Fig. 3) extracts the VIL and IRBTs features in the encoder and extracts the features of the VIL, lightning, and/or IRBTs in the decoder; 2) the common feature extractor modules (the deep gray rectangle in Fig. 3) share the weight parameters for extracting the features from VIL, GLM lightning flashes, and/or IRBTs in both the encoder and decoder; 3) the feature extractors of lightning in the decoder (the blue rectangle in Fig. 3) are used to extract features from VIL, lightning, and/or IRBTs for predicting lightning flashes.

Each convolution layer in Fig. 3 is followed by a batch normalization layer and LeakyReLu activation layer. In the input layer, VIL and/or IRBTs were fed into the U-net model. Furthermore, 13 additional lightning frames corresponding to VIL and IRBTs as auxiliary input features were fed into the model after the second Maxpool2d layer for lightning and/or VIL nowcasting (shown in Fig. 3). The encoder section consisted of six sets of two convolution layers and five maxpooling layers. Two output layers were used for lightning and VIL forecasting, respectively, in the decoder sec-

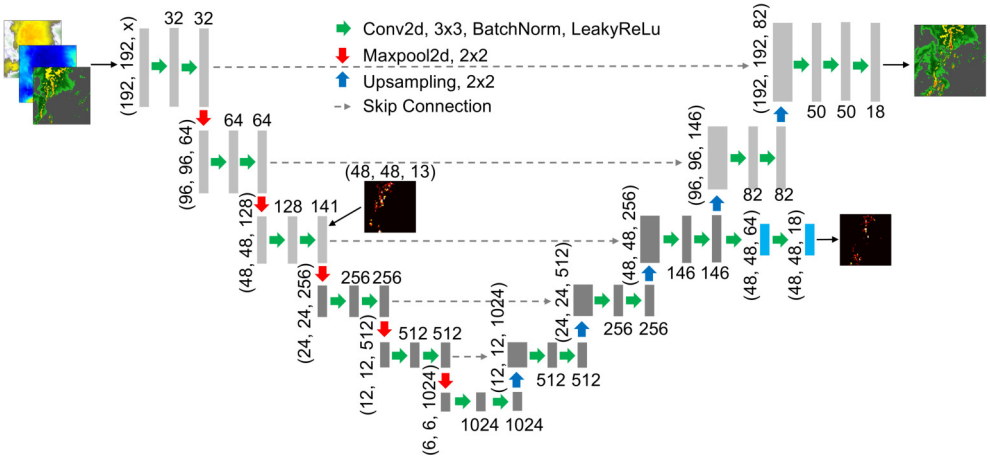

Fig. 3. Description of the encoder-decoder U-net model with multi-task learning for convection and lightning nowcasting. The light gray rectangle, navy blue rectangle, and deep gray rectangle blocks denote the features extractor of vertically integrated liquid (VIL), lightning flashes, and both, respectively. 3 × 3 represents the dimensions of the filter applied in the convolutional layers. 2 × 2 is the window size of maxpooling layers and upsampling layers. The large-scale and small-scale features are concatenated together by skip connections (dashed lines with arrows) to keep multi-scale features. *x* in the input layer indicates that input channels 13 (39) of VIL (VIL and IRBTs) were fed into the model.

tion. In the decoder section, the common feature extractor module consists of three sets of two convolution layers and three upsampling layers; a lightning forecasting module consists of two convolution layers without a LeakyReLu activation layer; and a VIL forecasting module consists of two sets of two convolution layers, two upsampling layers, and the last output layer (one convolution layer without the LeakyReLu activation layer).

The encoder-decoder U-net models were built using the open-source Python library PyTorch Lightning (https://www.pytorchlightning.ai, version 1.3.8) with PyTorch (https://pytorch.org, version 1.7.1). The U-net models were run on a graphics processing unit (GPU: RTX 8000). The batch size was empirically set to 8. The Adam optimization scheme with default values was used as the learning strategy. The initial learning rate was set to 3×10^{-4}, and a learning rate optimization method called ReduceLROnPlateau[b] was used, which reduces the learning rate by a factor of 0.9 when a metric shows no improvement for three subsequent epochs. The above optimal hyperparameters are based on the results of a group of experiments. The U-net models were set with the same random seed and weight initialization methods to reproduce comparable results.

3.2. *Design of loss function*

To cope with the heavily uneven distribution of VIL and lightning flashes, the following temporal-spatial exponent weighted loss function was used to train the neural network model:

$$\text{Loss} = \sum_{t=0}^{T} W_t \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{N} \exp^{\alpha y_m} \left(y_{m,i,j} - \hat{y}_{m,i,j} \right), \quad (2)$$

where W_t is a weight parameter at the input time t[c], T is the total number of input times (i.e., 18), M is the total number of grid points, N is the dimension of the image, and α is the weight factor with a power function. y_m and $\hat{y}_m$ represent the observed and predicted values, respectively. A time-exponent-weighted MSE loss function was used to optimize the VIL and lightning flash nowcasting tasks. The weight factor α for VIL and lightning nowcasting was set at 3.5 and 2 based on the results of a group of experiments, respectively. With MTL, the total of loss is defined as follows:

$$\text{TLoss} = \beta \times \text{VILLoss} + \gamma \times \text{LghtLoss}, \quad (3)$$

where VILLoss and LghtLoss are the losses of VIL and lightning nowcasting, respectively; β and γ are the weight parameters of VILLoss and LghtLoss, respectively, to scale the VILLoss and LghtLoss to the similar magnitude ranges.

[b] https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html

[c] W_t is set to (1, 1.2, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5, 5, 5, 5, 5, 5, 5, 5) and (1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3, 3, 3, 3) from first frame to 18-th frame for VIL and lightning, respectively.

3.3. *Description of evaluating metrics and experiments*

To evaluate the forecasting skill of the neural network models, the equitable threat score (ETS), critical success index (CSI), probability of detection (POD), false alarm ratio (FAR), and bias score (BIAS) were calculated for the training, validation, and test datasets. CSI, POD, FAR and BIAS can be calculated from using the following equations:

$$\text{ETS} = \frac{\text{TP} + \text{TP}_{\text{random}}}{\text{TP} + \text{FN} + \text{FP} - \text{TP}_{\text{random}}}, \quad (4)$$

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (5)$$

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (6)$$

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}}, \quad (7)$$

$$\text{BIAS} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FN}}, \quad (8)$$

where $\text{TP}_{\text{random}} = (\text{TP} + \text{FN}) \times (\text{TP} + \text{FP}) / (\text{TP} + \text{FN} + \text{FP} + \text{CN})$, TP (True Positive) is the number of grid points where both prediction and observation satisfy an assigned threshold condition, FP (False Positive) is the number of grid points where the prediction satisfies the assigned threshold condition, but the observation does not; FN (False Negative) is the number of grid points where the observation satisfies the assigned threshold condition, but not the prediction; and CN (Correct Negative) indicates correct negative. ETS, CSI, POD, and FAR have a range between 0 and 1, but BIAS ranges from 0 to infinity (a frequency bias closer to1 means a perfect score).

The U-net model was formulated to run for convection and lightning nowcasting using MUNet (MTL U-net) and SUNet (STL U-net) models, respectively. The experimental design is summarized in Table 2. The MUNet and SUNet models were configured to forecast the VIL and lightning over the next 90 min (18 frames) based on the immediate last 65-minute observations (13 frames). VIL and lightning were optimized for MUNet (including M2DIR and M2DNR) but separately for SUNet (e.g., S2DIR). Two common baseline models, the persistence and optical flow models based on the rainymotion package (Ayzel et al., 2019) were used as baseline references for assessing the performance of the neural network models developed in this study. The persistence model assumes the last input (value) not changing in future time steps. The optical-flow model from the "rainymotion package" estimates a dense flow at each pixel using the last two frames in the inputs and creates a nowcast by advecting each pixel using a semi-Lagrangian scheme, same as Veillette et al. (2020). The persistence and optical-flow model do not easily grasp the generation, disappearance and

changes of convection storms from weather radar and/or satellite cloud images.

During model training, we first determined the number of epochs to obtain the optimal results for MUNet/SUNet and IRBTs on convection and lightning nowcasting. For STL, the epoch that allowed the models to achieve the best CSI of VIL (or lightning) nowcasting was selected. Nevertheless, for the MTL, the epoch number that resulted in the best CSI for both VIL and lightning nowcasting was chosen.

4. Experiments and analysis

4.1. *Performance Evaluation*

The skill scores of 0–90 min VIL ($\geqslant$3.53 kg m^{-2} corresponding to 133 in Veillette et al., 2020) and lightning [$\geqslant$1 flash (64 km^{-2})] nowcasting from the persistence, M2DIR, M2DNR and S2DIR are shown in Table 3. The neural network models significantly outperformed the persistence model for both VIL and lightning nowcasting. All neural network models underpredicted weak lightning and overestimated intense lightning (data not shown). However, the neural network models slightly overpredicted the VIL for lower VIL thresholds and underpredicted it for higher VIL thresholds.

Figure 4 shows CSI of all models for VIL and lightning nowcasting with 95% confidence intervals (obtained by a bootstrapping method with 500 iterations). The range of confidence interval is so small that it makes it difficult to distinguish it in the figure. It is possible that each neural network model has similar performance for the convection storms in the test data set, leading to the confidence interval differences in the skill scores being not significant. All the neural network models showed much better skills than the persistence model with a lead time beyond 5–10 min (Fig. 4b). The MUNet model evidently improved the lightning forecast for all forecast lead times. For the larger lightning density thresholds, the neural network models (MUNet and SUNet) did not exceed the persistence model until the 10-min lead time (Fig. 4c). Lightning density tends to have smaller scale structures, exhibits more randomness than VIL, making it more challenging to forecast than VIL. This is particularly true for intense lightning cores, where the CSI of the neural network model forecast remains low within the first 5–10 min. Therefore the persistence model, whose forecast skill starts at 100% accuracy, performs better at this very short lead time.

4.2. *Sensitivity to the normalization method*

To investigate the effect of the normalization method on lightning nowcasting, we trained two SUNet models for lightning nowcasting with the same mean squared error (MSE) loss function but using the Min-max and Log2Norm methods, respectively, to normalize lightning data. The skill metrics of lightning flash nowcasting for the Min-max and Log2Norm normalization methods are listed in Table 4. The skill scores (CSI and ETS) of lightning nowcasting suggest that the Min-max normalization method is better than Log2Norm at the lower lightning density threshold for lightning nowcasting. However, the Log2Norm normalization method is superior for larger lightning density thresholds, and its performance is dramatically improved for extreme lightning cases. The SUNet model, using either the Min-max or Log2Norm for lightning data normalization, underpredicts the lightning flash density, especially for large lightning density thresholds.

As shown in Fig. 5, the Log2Norm normalization

Table 2. Description of experiments for VIL and lightning nowcasting.

Abbreviation	Description	Input Features	Targets
M2DIR	Unet-2D with MTL, using IRBTs	IRBTs, VIL and Lightning	VIL and Lightning
M2DNR	Unet-2D with MTL, w/o IRBTs	VIL and Lightning	VIL and Lightning
S2DIR	Unet-2D with STL using IRBTs	IRBTs, VIL and Lightning	VIL or Lightning

Table 3. Evaluation metrics of the VIL ($\geqslant$3.53 kg m^{-2}) and lightning [($\geqslant$1 flash (64 km^{-2})] nowcasting for lead times of 0–90 min using the persistence model and the MUNet (multi-task learning U-net) model with/without including IRBTs as auxiliary input features, and the SUNet (single-task learning U-net) model including IRBTs as auxiliary input features. The bold black means the best skill scores.

Target	Model	POD	FAR	BIAS	CSI	ETS
VIL	Persistence	0.233	0.754	1.027	0.138	0.116
	Optical flow	0.349	0.654	1.078	0.212	0.190
	S2DIR	0.461	0.593	1.170	0.267	0.242
	M2DIR	0.488	0.600	1.247	0.277	0.252
	M2DNR	0.496	0.599	1.283	**0.278**	**0.253**
lightning	Persistence	0.199	0.800	1.156	0.110	0.096
	Optical flow	0.210	0.721	0.836	0.136	0.124
	S2DIR	0.334	0.716	1.301	0.169	0.154
	M2DIR	0.355	0.651	1.064	**0.208**	**0.195**
	M2DNR	0.324	0.643	0.950	0.198	0.185

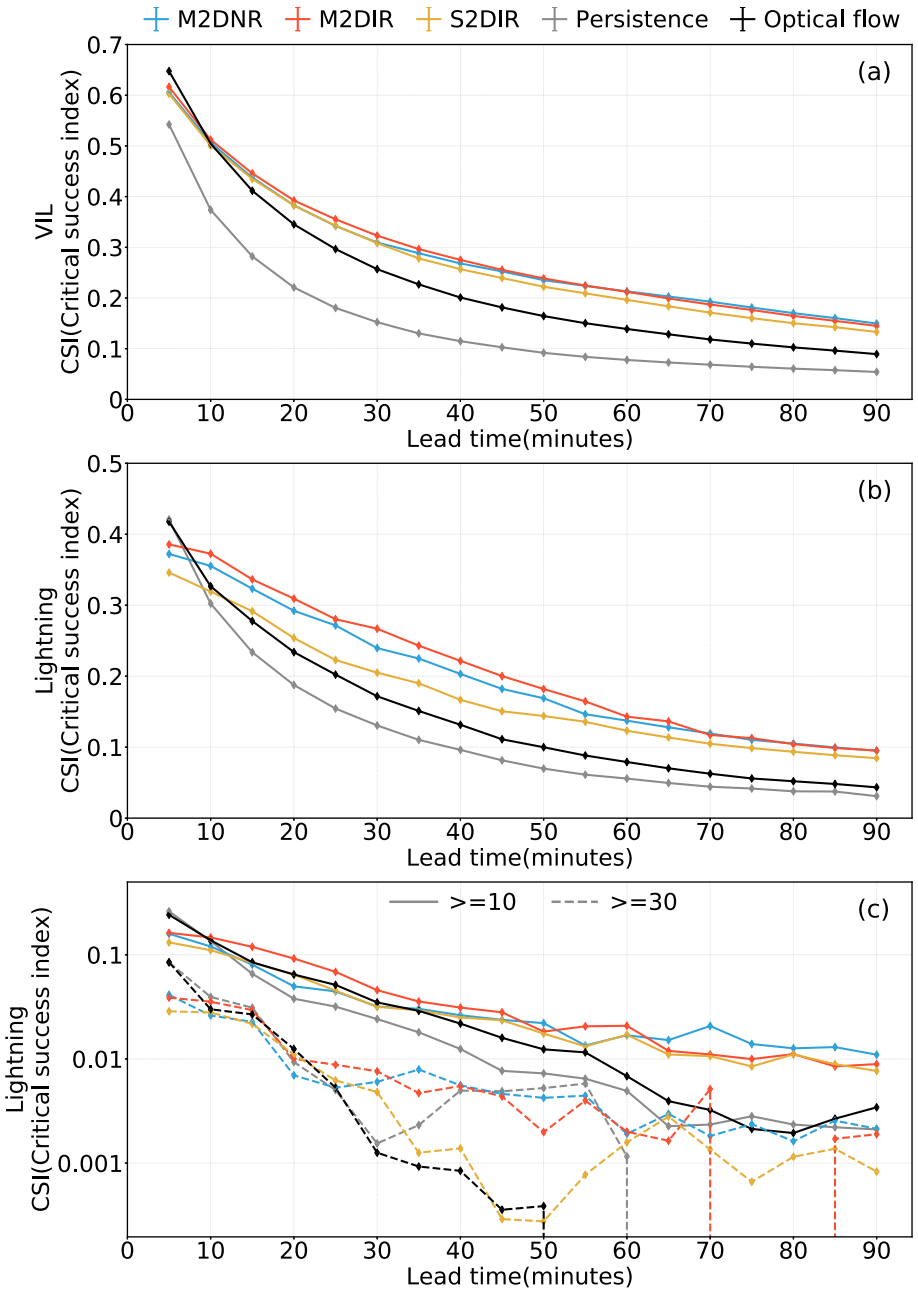

Fig. 4. CSI of time series of (a) VIL (⩾3.53 kg m^{-2}) and (b) lightning flash nowcasting [⩾1 flash (64 km^{-2})] all models, respectively. (c) CSI of time series for different lightning thresholds [⩾10 and ⩾30 flash (64 km^{-2})] of lightning flash nowcasting all models. Persistence, M2DIR, M2DNR and S2DIR are the persistence model, the MUNet (multi-task learning U-net) models with/without the inclusion of IRBTs as auxiliary input features, and the SUNet (single-task learning U-net) model with the inclusion of IRBTs as auxiliary input features, respectively. The skill scores shown exceed the 95% confidence interval (obtained by a bootstrapping method with 500 iterations).

method achieves significantly better scores, especially for longer lead times and large lightning densities. The Log2Norm normalization method handles the skewed lightning density distribution better than the Min-max normalization methods (Fig. 2), which helps the neural network models to learn the features with respect to intense lightning activities in the lightning data.

4.3. *Effects of MTL and STL*

Table 3 shows that for VIL nowcasting, M2DIR yields a CSI of 0.277 and ETS of 0.252, which outperforms S2DIR (with a CSI of 0.267 and ETS of 0.242). Both M2DIR and S2DIR slightly overpredict the weak and middle VIL values. However, M2DIR and S2DIR both significantly underpredict at large VIL thresholds, e.g., M2DIR yields a

Table 4. Evaluation of lightning nowcasting with the Min-max and Log2Norm normalization method for different lightning density thresholds at lead time up to 90 min.

Method	Thresholds	POD	FAR	BIAS	CSI	ETS
Min-max	1	0.302	0.644	0.989	0.164	0.150
	5	0.079	0.426	0.250	0.055	0.054
	10	0.030	0.261	0.093	0.025	0.024
	15	0.018	0.175	0.056	0.015	0.015
	30	0.001	0.041	0.003	0.001	0.001
Log2Norm	1	0.142	0.496	0.316	0.114	0.107
	5	0.054	0.459	0.207	0.041	0.039
	10	0.027	0.334	0.134	0.022	0.021
	15	0.018	0.294	0.135	0.014	0.014
	30	0.006	0.207	0.223	0.005	0.005

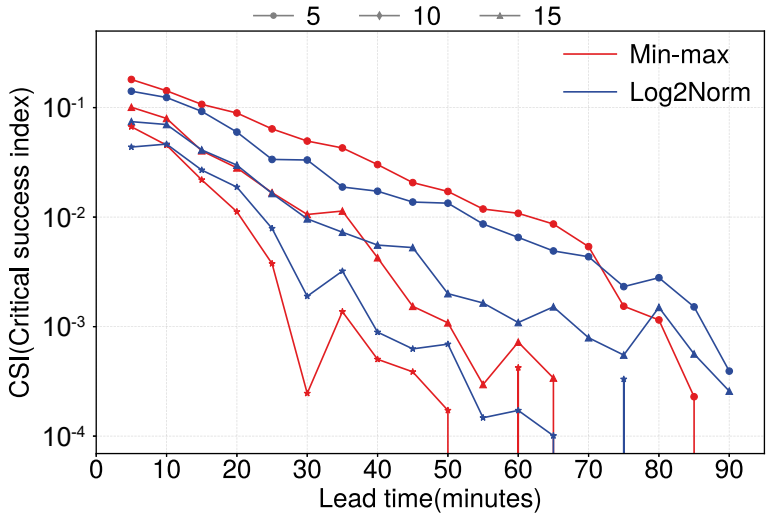

Fig. 5. CSI time series for lightning nowcasting at different lightning density thresholds [5, 10, and 15 flashes (64 km^{-2})] using the SUNet model, and the lightning data normalized by the Min-max (red line) and Log2Norm (blue line) normalization method, separately.

CSI of 0.119 and BIAS of 0.544, and S2DIR yields a CSI of 0.112 and BIAS of 0.333 when VIL > 12.14 kg m^{-2}. The M2DIR CSI scores are slightly better than S2DIR for all forecast lead times (Fig. 4a).

The MUNet model (M2DIR) significantly outperformed the lightning nowcasting of the SUNet model (S2DIR), with the CSI increasing by 0.039 and ETS by 0.041 (Table 3). Moreover, the frequency bias of M2DIR was closer to 1, which is better than that of S2DIR. Figures 4b and 4c show that the skill metrics (CSI) of M2DIR were evidently superior to those of S2DIR for all lead times.

In summary, the deep model experimental results clearly demonstrate that MTL significantly improves lightning nowcasting in comparison with STL, and it also improves the VIL forecast to a certain extent. However, the superiority of MTL tends to diminish with lead times over 60 min. In the following section, we explore the impact of adding satellite IRBTs data to the deep learning models.

4.4. *Effects of IR brightness temperature*

Satellite IRBTs contain information on the height and thermodynamic and microphysical processes of deep moist convection (Mecikalski et al., 2008; Rosenfeld et al., 2008). Herein, the effect of IRBTs on VIL and lightning nowcasting using MUNet models is analyzed. Table 3 shows that the GOES-16 IRBTs have different effects on VIL and lightning nowcasting. IRBTs have little effect on VIL nowcasting (Table 3), and for intense VIL, the forecast is slightly degraded. This is likely because IRBTs may be dominated by the cold cloud top temperature of the anvils for deep convective cloud anvils (Müller et al., 2018), which contain very little extra information about VIL, but a considerable amount of less relevant information about the ice crystal anvils that mask the VIL signals.

In contrast, IRBTs have a significant influence on lightning forecasting (Table 3 and Figs. 4b and c). Lightning occurrence increases as the brightness temperature decreases, and a brightness temperature lower than 217 K

can be used as a proxy for the presence of lightning (Katsanos et al., 2007). Thus, IRBTs contain rich information regarding the occurrence of lightning. When IRBTs were included as auxiliary input features (Fig. 4b), the MUNet model with IRBTs achieved the best performance for lightning nowcasting, although it slightly overpredicted lightning more than the MUNet model without IRBTs.

The overprediction of the density of lightning flashes with the MUNet model that includes IRBTs is likely because the model slightly overestimates the relationship between lightning flash occurrence and lower brightness temperature. A brightness temperature lower than 217 K can be used as a proxy for lightning flash occurrence; however, it is not a sufficient condition for lightning occurrence. A lightning discharge is a combination of complex dynamic and microphysical processes (Mostajabi et al., 2019). In-cloud mixed-phase microphysics driven by the presence of a strong updraft (e.g., > 6 m s^{-1} in the mean, Zipser and Lutz, 1994) in the lower- to mid- level is necessary to facilitate significant cloud electrification and the strong electric fields required to initiate lightning (Carey and Rutledge, 1996; Petersen et al., 1996; Deierling and Petersen, 2008). Furthermore, low IRBTs may occur after vigorous convection has dissipated.

4.5. *Organized versus isolated convective storms*

4.5.1. *Overall results*

Organized convection evolves more continuously than scattered isolated convective storms. The latter involves more cells that initiate or disperse, and thus, are more difficult to forecast. To obtain a visual perception of the impact of the different neural network models on the VIL and lightning nowcasting of convection events of different types, the performance of nowcasting of two categories of convection and lightning events— organized convective storms and isolated convective cells—each composed of 50 cases selected from the testing datasets, are summarized in Table 5 and Fig. 6. The statistical verification of the experiments was consistent with the earlier discussions. The MUNet model outperformed the SUNet model in terms of both VIL and lightning nowcasting. Both the MUNet and SUNet models achieve better VIL and lightning nowcasting for the organized convective storms than isolated convective cells, especially for the long nowcasting lead times (Fig. 6).

For VIL nowcasting, M2DNR appears to be superior for forecasting organized convective storms; however, the skill scores (CSI and ETS) for M2DIR are better than those of M2DNR for isolated convective cells (Table 5), which demonstrates that IRBTs provide useful information for isolated convective cell nowcasting. Figure 6a shows that the neural network models clearly overpredict for organized convective storms with all lead times for VIL nowcasting, but its frequency bias shows almost unbiased scores for isolated convective cells. As illustrated in Fig. 6b, the skill score of M2DIR was better than that of M2DNR and S2DIR. M2DNR and S2DIR slightly underpredicted and overpredicted the lightnings of both organized convective storms and isolated convective cells, respectively.

4.5.2. *A case of convective storms*

Figure 7 shows the distribution of VIL observations and forecasts of the three models at different lead times for a typical case taken from the testing dataset with multiple small isolated convective cells and an organized convective storm. The MUNet model is more capable of capturing convective storm development and merging within the white rectangles B in Fig. 7 than the SUNet model. Compared to M2DNR and S2DIR, M2DIR produces less overestimation of intense convective cores. Nevertheless, the MUNet model still significantly underestimates the multiple small and isolated convection cells and the convection intensity and misses some cells completely (e.g., those within the white rectangle area A in Fig. 7). Finally, our models failed to capture the convection initiation (white rectangles A in Fig. 7). This may be due to the current input data not contain-

Table 5. The verification metrics for the two categories of convection (organized convective storms and isolated convective cells referred to as C1 and C2, respectively) VIL (≥3.53 kg m^{-2}) and lightning (≥1 flash 64 km^{-2}) nowcasting using the MUNet model with/without including IRBTs (i.e., M2DIR and M2DNR, respectively) and the SUNet model including IRBTs (i.e., S2DIR). The bold black notifies the best skill scores.

Target	Types	Experiments	POD	FAR	BIAS	CSI	ETS
VIL	C1	S2DIR	0.753	0.530	1.639	0.405	0.332
		M2DIR	0.751	0.518	1.590	0.413	0.342
		M2DNR	0.802	0.535	1.768	**0.418**	**0.343**
	C2	S2DIR	0.369	0.585	0.893	0.238	0.220
		M2DIR	0.406	0.599	1.022	**0.249**	**0.231**
		M2DNR	0.390	0.581	0.978	0.241	0.223
Lightning	C1	S2DIR	0.495	0.594	1.244	0.291	0.253
		M2DIR	0.477	0.520	0.999	**0.316**	**0.283**
		M2DNR	0.454	0.520	0.936	0.307	0.274
	C2	S2DIR	0.362	0.732	1.327	0.182	0.161
		M2DIR	0.387	0.650	1.104	**0.221**	**0.202**
		M2DNR	0.369	0.645	1.044	0.218	0.199

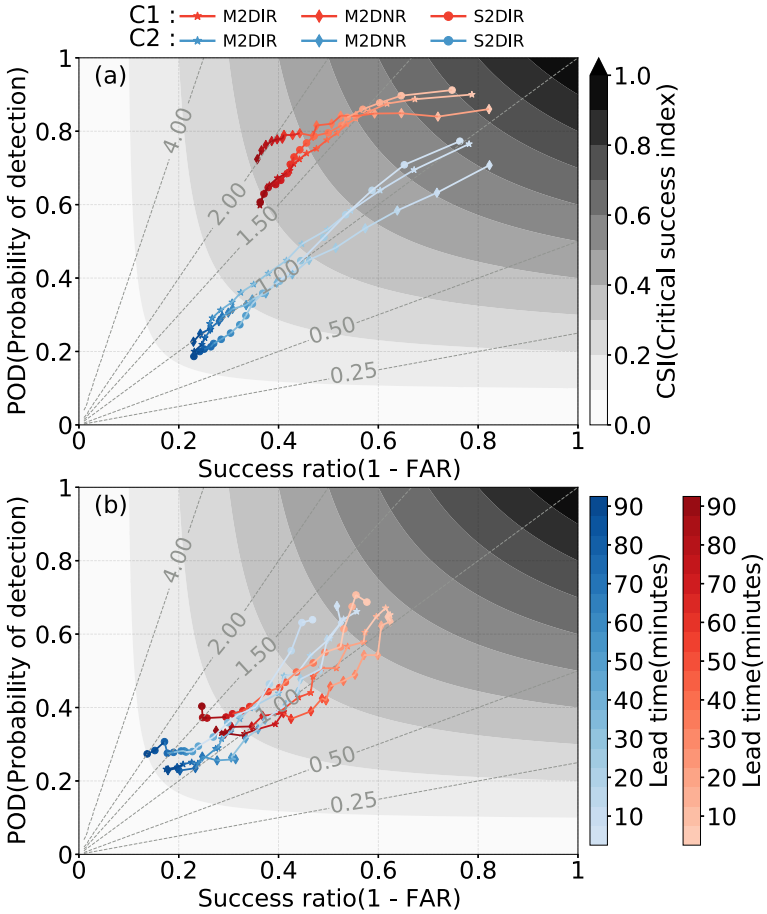

Fig. 6. The performance diagram of (a) VIL (⩾3.53 km m^{-2}) and (b) lightning (⩾1 flash 64 km^{-2}) nowcasting for different lead times for organized convective storms (C1, red line) and isolated convective cells (C2, blue line) using the MUNet model with/without IRBTs (i.e., M2DIR and M2DNR) and the SUNet model with IRBTs (i.e., S2DIR).

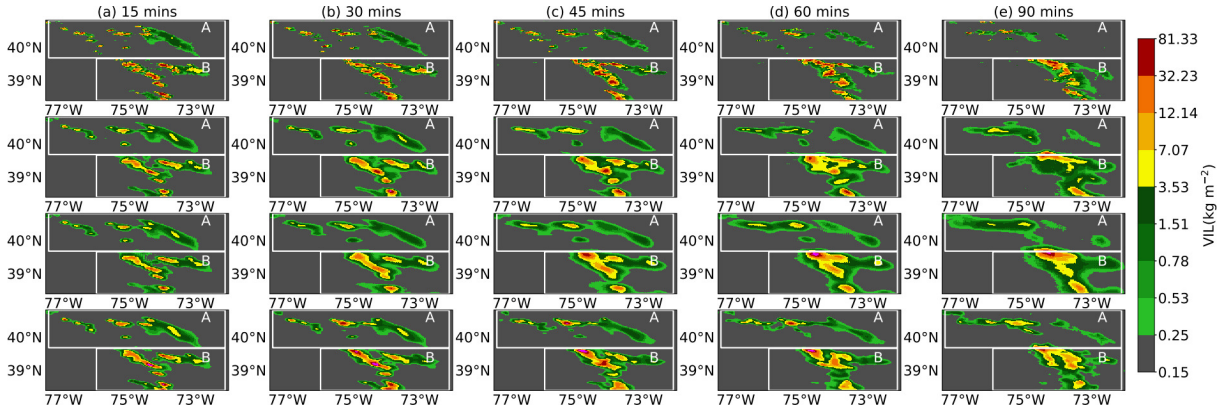

Fig. 7. VIL (kg m^{-2}) nowcasting and observation (384 km × 384 km) from 19 August 2019 at (a) 15-min, (b) 30-min, (c) 45-min, (d) 60-min, and (e) 90-min using the MUNet model with/without including IRBTs as auxiliary input features, which refer to M2DIR and M2DNR, respectively, and the SUNet model including IRBTs, which refer to S2DIR. White rectangles A indicate multiple isolated and small convective cells. White rectangles B indicate organized convective storms development and merging.

ing sufficient information for convection initiation, such as boundary layer convergence and atmospheric instability.

For this case, although the neural network models tended to underestimate the lightning activities of the isolated convective cells within region A, the MUNet model in particular accurately captures the overall evolution of lightning activ-

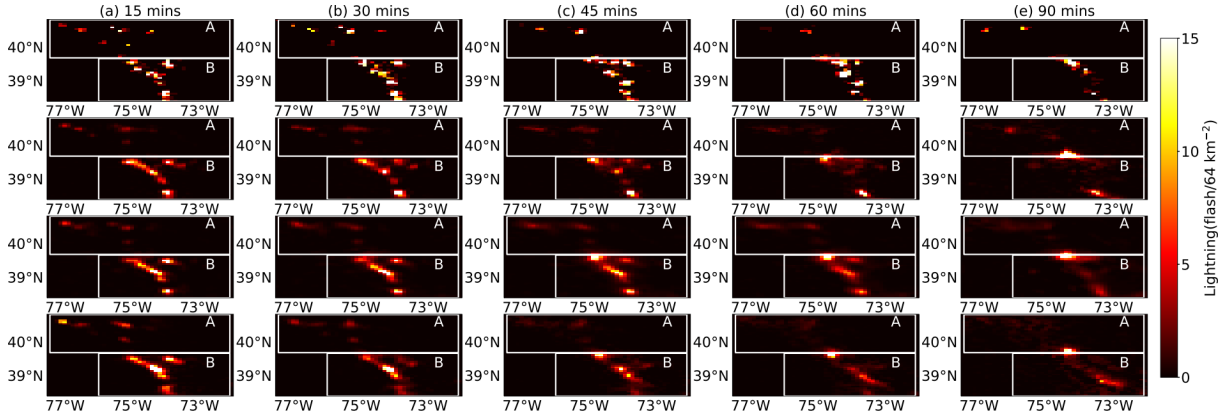

Fig. 8. Same as Fig. 7, but for lightning (flash 64 km^{-2}) observation and nowcasting (384 km × 384 km).

ities of the organized convective storms within region B for all lead times (Fig. 8).

5. Summary and discussions

In this study, a multi-task learning (MTL) encoder-decoder U-net with models was constructed to forecast VIL and lightning (0–90 min) and compared to the traditional single-task learning (STL) model. The input data included GOES-16 geostationary satellite IR brightness temperatures (6.9 μm and 10.7 μm channels) with high temporal and spatial resolutions (2 km grid resolution and 5-min intervals), VIL derived from NEXRAD data, and gridded lightning flashes obtained from satellite GLM. Given the extremely skewed nature of the lightning distribution, a log-transform normalization method was used to normalize the lightning data for lightning nowcasting. To alleviate the decay of lightning and VIL nowcasting skill with increasing forecast lead time, a spatiotemporal exponent-weighted loss function was introduced. Furthermore, the effects of GOES-16 IRBTs on lightning and VIL nowcasting were also investigated using the MTL model. By performing the experiments with various training models and training data, we arrived at the following conclusions:

1) The MUNet and SUNet models significantly outperformed the persistence and optical flow models for both lightning and VIL nowcasting beyond ~5–10 min lead time. As the lightning density tends to have smaller scale structures and behaves more randomly than VIL, neither the MUNet nor the SUNet models can beat the persistence model at very short lead times (0–15 min, depending on the intensity of lightning).

2) Applying a log-transform normalization for lightning density data dramatically improves the nowcasting of intense lightning at all lead times in comparison to the min-max normalization method, although the latter shows a slight advantage for weak lightning.

3) M2DIR evidently outperforms S2DIR for lightning and VIL nowcasting, particularly for those with intense lightning thresholds. However, the superiority of M2DIR gradually decays with lead time.

4) Including IRBTs as auxiliary input features significantly improves lightning nowcasting, although it has only a small influence on VIL nowcasting. This is because IRBTs contain important proxy (ice particle) information of lightning occurrences but little more about convection features than the VIL itself. Furthermore, IRBTs may be dominated by lumpy ice particle anvils, which may mask important signals relevant to VIL.

5) Neural network models, particularly the MUNet model, are more capable of nowcasting the VIL and lightning of organized convective storms than those with isolated convective cells.

This study demonstrates that the lightning and VIL MTL approach for joint lightning and VIL optimization can improve the accuracy and generalization of the neural network model for lightning and VIL nowcasting. The MTL significantly improved lightning nowcasting. This research complements previous work by Mostajabi et al. (2019), Zhou et al. (2020), and Shrestha et al. (2021) for 30- and 60- min lightning nowcasting. The models presented herein extended the lightning nowcasting to 90 min with appreciable skill.

In addition to lightning, heavy rainfall, hail, and damaging winds are also produced by severe convective storms. We are currently expanding our research on the nowcasting of these weather phenomena by considering longer forecast lead times (up to 6 h) and using more data sources, including rapid update NWP model outputs and dual polarization radar observations, which are advantageous for improving convection initiation (Pan et al., 2021) and lightning initiation (Katuzienski, 2019) forecasting. Data pre-processing and manipulation of the learning parameters are crucial to account for those data with extremely skewed distributions, as was done with lightning in this study. Learning multiple objectives at once is a difficult optimization problem (Parisotto et al., 2016), and more effective optimization methods, such as gradient surgery (Yu et al., 2020) and weighting schemes of the multi-task loss function (Liu et al., 2019), should be explored for the model training.

Acknowledgements. This work is supported by the Science and Technology Grant No. 520120210003, Jibei Electric Power

Company of the State Grid Corporation of China. The authors would like to thank Mark VEILLETTE to provide the SEVIR dataset. SEVIR dataset can be downloaded from AWS (https://registry.opendata.aws/sevir).

REFERENCES

Abarbanel, H. D. I., P. J. Rozdeba, and S. Shirman, 2018: Machine learning: Deepest learning as statistical data assimilation problems. *Neural Computation*, **30**(8), 2025–2055, https://doi.org/10.1162/neco_a_01094.

Arcucci, R., J. C. Zhu, S. Hu, and Y.-K. Guo, 2021: Deep data assimilation: Integrating deep learning with data assimilation. *Applied Sciences*, **11**(3), 1114, https://doi.org/10.3390/app11031114.

Ayzel, G., M. Heistermann, and T. Winterrath, 2019: Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1). *Geoscientific Model Development*, **12**(4), 1387–1402, https://doi.org/10.5194/gmd-12-1387-2019.

Beucler, T., I. Ebert-Uphoff, S. Rasp, M. Pritchard, and P. Gentine, 2021: Machine learning for clouds and climate (invited chapter for the AGU geophysical monograph series "clouds and climate"). *Earth and Space Science Open Archive*, https://doi.org/10.1002/essoar.10506925.1.

Carey, L. D., and S. A. Rutledge, 1996: A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm. *Meteorol. Atmos. Phys.*, **59**, 33–64, https://doi.org/10.1007/BF01032000.

Crawshaw, M., 2020: Multi-task learning with deep neural networks: A survey. arXiv preprint arXiv:2009.09796.

Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *J. Geophys. Res.*, **113**(D16), D16210, https://doi.org/10.1029/2007JD009598.

Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.*, **10**(6), 785–797, https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

Fierro, A. O., E. R. Mansell, D. R. MacGorman, and C. L. Ziegler, 2013: The implementation of an explicit charging and discharge lightning scheme within the WRF-ARW model: Benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. *Mon. Wea. Rev.*, **141**(7), 2390–2415, https://doi.org/10.1175/MWR-D-12-00278.1.

Franch, G., A. Nardelli, C. Zarbo, V. Maggio, G. Jurman, and C. Furlanello, 2016: Deep learning for rain and lightning nowcasting. *Proc. NIPS 2016 Workshop on ML for Spatiotemporal Forecasting*, Barcelona, Zenodo, https://doi.org/10.5281/zenodo.3594325.

Geer, A. J., 2021: Learning earth system models from observations: Machine learning or data assimilation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **379**(2194), 20200089, https://doi.org/10.1098/rsta.2020.0089.

Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**(11), 5742–5751, https://doi.org/10.1029/2018GL078202.

Han, L., J. Z. Sun, and W. Zhang, 2020: Convolutional neural network for convective storm nowcasting using 3-D Doppler weather radar data. *IEEE Trans. Geosci. Remote Sens.*, **58**(2), 1487–1495, https://doi.org/10.1109/TGRS.2019.2948070.

Han, L., M. X. Chen, K. K. Chen, H. N. Chen, Y. B. Zhang, B. Lu, L. Y. Song, and R. Qin, 2021: A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Adv. Atmos. Sci.*, **38**(9), 1444–1459, https://doi.org/10.1007/s00376-021-0215-y.

Houze, R. A. Jr., D. C. Wilton, and B. F. Smull, 2007: Monsoon convection in the Himalayan region as seen by the TRMM precipitation radar. *Quart. J. Roy. Meteor. Soc.*, **133**(627), 1389–1411, https://doi.org/10.1002/qj.106.

Ioffe, S., and C. Szegedy, 2015: Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proc. 32nd International Conf. on Machine Learning*, Lille, France, JMLR.org, 448–456.

Katsanos, D. K., K. Lagouvardos, V. Kotroni, and A. A. Argiriou, 2007: The relationship of lightning activity with microwave brightness temperatures and spaceborne radar reflectivity profiles in the central and eastern mediterranean. *J. Appl. Meteor. Climatol.*, **46**(11), 1901–1912, https://doi.org/10.1175/2007JAMC1454.1.

Katuzienski, D. O., 2019: Comparing dual-polarization radar lightning forecast methods across southwest Utah. M.S. thesis, Dept. of Engineering Physics, Air Force Institute of Technology.

Krasnopolsky, V. M., and Y. Lin, 2012: A neural network nonlinear multimodel ensemble to improve precipitation forecasts over continental US. *Advances in Meteorology*, **2012**, 649450, https://doi.org/10.1155/2012/649450.

LeCun, Y., Y., Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**(7553), 436–444, https://doi.org/10.1038/nature14539.

Li, H. C., C. Yu, J. J. Xia, Y. C. Wang, J. Zhu, and P. W. Zhang, 2019: A model output machine learning method for grid temperature forecasts in the Beijing area. *Adv. Atmos. Sci.*, **36**(10), 1156–1170, https://doi.org/10.1007/s00376-019-9023-z.

Liu, C., S. Yang, D. Di, Y. J. Yang, C. Zhou, X. Q. Hu, and B.-J. Sohn, 2021: A machine learning-based cloud detection algorithm for the Himawari-8 spectral image. *Adv. Atmos. Sci.*, in press, https://doi.org/10.1007/s00376-021-0366-x.

Liu, S. K., E. Johns, and A. J. Davison, 2019: End-to-end multi-task learning with attention. *Proc. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, IEEE, 1871–1880, https://doi.org/10.1109/CVPR.2019.00197.

Mecikalski, J. R., K. M. Bedka, S. J. Paech, and L. A. Litten, 2008: A statistical evaluation of GOES cloud-top properties for nowcasting convective initiation. *Mon. Wea. Rev.*, **136**(12), 4899–4914, https://doi.org/10.1175/2008MWR2352.1.

Meng, Q., W. Yao, and L. T. Xu, 2019: Development of lightning nowcasting and warning technique and its application. *Advanceds in Meteorology*, **2019**, 2405936, https://doi.org/10.1155/2019/2405936.

Mostajabi, A., D. L. Finney, M. Rubinstein, and F. Rachidi, 2019: Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *npj Climate and Atmospheric Science*, **2**(1), 41, https://doi.org/10.1038/s41612-019-0098-0.

Müller, R., S. Haussler, and M. Jerg, 2018: The role of NWP filter for the satellite based detection of cumulonimbus clouds. *Remote Sensing*, **10**(3), 386, https://doi.org/10.3390/rs10030386.

Pal, A., S. Mahajan, and M. R. Norman, 2019: Using deep neural networks as cost-effective surrogate models for super-parameterized E3SM radiative transfer. *Geophys. Res. Lett.*, **46**(11), 6069–6079, https://doi.org/10.1029/2018GL081646.

Pan, X., Y. H. Lu, K. Zhao, H. Huang, M. J. Wang, and H. N. Chen, 2021: Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model. *Geophys. Res. Lett.*, **48**, e2021GL095302, https://doi.org/10.1029/2021GL095302.

Parisotto, E., L. J. Ba, and R. Salakhutdinov, 2016: Actor-mimic: Deep multitask and transfer reinforcement learning. *Proc. 4th International Conf. on Learning Representations*, San Juan, Puerto Rico.

Petersen, W. A., S. A. Rutledge, and R. E. Orville, 1996: Cloud-to-ground lightning observations from TOGA COARE: Selected results and lightning location algorithms. *Mon. Wea. Rev.*, **124**(4), 602–620, https://doi.org/10.1175/1520-0493(1996)124<0602:CTGLOF>2.0.CO;2.

Rasp, S., and S. Lerch, 2018: Neural networks for postprocessing ensemble weather forecasts. *Mon. Wea. Rev.*, **146**(11), 3885–3900, https://doi.org/10.1175/MWR-D-18-0187.1.

Rasp, S., and N. Thuerey, 2021: Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for WeatherBench. *Journal of Advanceds in Modeling Earth Systems*, **13**(2), e2020MS002405, https://doi.org/10.1029/2020MS002405.

Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(39), 9684–9689, https://doi.org/10.1073/pnas.1810286115.

Rasp, S., P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, 2020: WeatherBench: A benchmark data set for data-driven weather forecasting. *Journal of Advanceds in Modeling Earth Systems*, **12**(11), e2020MS002203, https://doi.org/10.1029/2020MS002203.

Ravuri, S., and Coauthors, 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672–677, https://doi.org/10.1038/s41586-021-03854-z.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Proc. 18th International Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, Springer, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Rosenfeld, D., W. L. Woodley, A. Lerner, G. Kelman, and D. T. Lindsey, 2008: Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase. *J. Geophys. Res.*, **113**, D04208, https://doi.org/10.1029/2007JD008600.

Seifert, A., and S. Rasp, 2020: Potential and limitations of machine learning for modeling warm - rain cloud microphysical processes. *Journal of Advanceds in Modeling Earth Systems*, **12**(12), e2020MS002301, https://doi.org/10.1029/2020MS002301.

Shi, X. J., Z. R. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Proc. 28th International Conf. on Neural Information Processing Systems*, Montreal, Canada, MIT Press, 802–810.

Shrestha, Y., Y. Zhang, R. Doviak, and P. W. Chan, 2021: Lightning flash rate nowcasting based on polarimetric radar data and machine learning. *Int. J. Remote Sens.*, **42**(17), 6762–6780, https://doi.org/10.1080/01431161.2021.1933243.

Sønderby, C. K., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. arXiv preprint arXiv: 2003.12140.

Sun, J. Z., and Coauthors, 2014: Use of NWP for nowcasting convective precipitation: recent progress and challenges. *Bull. Amer. Meteor. Soc.*, **95**(3), 409–426, https://doi.org/10.1175/BAMS-D-11-00263.1.

Tippett, M. K., and W. J. Koshak, 2018: A baseline for the predictability of U.S. cloud-to-ground lightning. *Geophys. Res. Lett.*, **45**(9), 10 719–10 728, https://doi.org/10.1029/2018GL079750.

Veillette, M., S. Samsi, and C. J. Mattioli, 2020: SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Proc. 34th Conf. on Neural Information Processing Systems*, Vancouver, Canada.

Wang, Y. B., M. S. Long, J. M. Wang, Z. F. Gao, and P. S. Yu, 2017: PredRNN: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Proc. 31st International Conf. on Neural Information Processing Systems*, Long Beach, California, Curran Associates Inc., 879–888.

Wu, H. X., Z. Y. Yao, J. M. Wang, and M. S. Long, 2021: MotionRNN: A flexible model for video prediction with spacetime-varying motions. *Proc. 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Nashville, TN, USA, IEEE, 15 435–15 444, https://doi.org/10.1109/CVPR46437.2021.01518.

Yao, Y. Q., X. D. Yu, Y. J. Zhang, Z. J. Zhou, W. S. Xie, Y. Y. Lu, J. L. Yu, and L. X. Wei, 2015: Climate analysis of tornadoes in China. *Journal of Meteorological Research*, **29**(3), 359–369, https://doi.org/10.1007/s13351-015-4983-0.

Yu, T. H., S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, 2020: Gradient surgery for multi-task learning. arXiv preprint arXiv:2001.06782.

Zhang, Y., and Q. Yang, 2018: An overview of multi-task learning. *National Science Review*, **5**(1), 30–43, https://doi.org/10.1093/nsr/nwx105.

Zhou, K. H., Y. G. Zheng, W. S. Dong, and T. B. Wang, 2020: A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *J. Atmos. Oceanic Technol.*, **37**(5), 927–942, https://doi.org/10.1175/JTECH-D-19-0146.1.

Zhuo, J.-Y., and Z.-M. Tan, 2021: Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Mon. Wea. Rev.*, **149**(7), 2097–2113, https://doi.org/10.1175/MWR-D-20-0333.1.

Zipser, E. J., and K. R. Lutz, 1994: The vertical profile of radar reflectivity of convective cells: A strong indicator of storm intensity and lightning probability. *Mon. Wea. Rev.*, **122**(8), 1751–1759, https://doi.org/10.1175/1520-0493(1994)122<1751:TVPORR>2.0.CO;2.