Precipitation Retrieval From Fengyun-3D MWHTS and MWRI Data Using Deep Learning

Kangwen Liu, Jieying He, Member, IEEE, and Haonan Chen, Senior Member, IEEE

Abstract—In this article, two multitask deep learning models, multilayer perceptron (MLP) and convolutional neural networks (CNNs) are constructed to detect precipitation flags and retrieve precipitation rates simultaneously over the Northwest Pacific area. The retrieval results are verified by integrated multisatellite retrievals for GPM. Compared with using a single payload, the results show the retrieval advantages of incorporating two microwave spaceborne payloads (microwave humidity and temperature sounder and microwave radiation imager). In addition, each separated task in multitasking deep learning models can outperform the single-task models. The MAE values of MLP and CNN are decreased by 0.05 and 0.04 mm/h, respectively. Nevertheless, it is still a challenge to extend the application of these models, which demands extracting the features learned from a certain area to other areas characterized by different precipitation features. A flexible network framework is developed with two transfer learning (TL) methods (freeze and fine-tuning) to tackle this problem. The pretrained models (MLP and CNN) are trained from the Northwest Pacific area, whereas the transferred models are applied to the Northeast Pacific area. The encouraging results show that the performance of the TL model using the fine-tune method has exceeded that of the comparison model without TL, which is trained by four times more data. The main advantage of using TL is the time reduction and efficiency improvements.

Index Terms—Deep learning (DL), microwave humidity and temperature sounder (MWHTS), microwave radiation imager (MWRI), precipitation retrieval, transfer learning (TL).

I. INTRODUCTION

FOR a long time, quantitative precipitation measurement has been a problem of great concern to meteorological researchers [1]. Traditionally, ground-based rain gauges and weather radars can well measure the precipitation at high temporal resolution [2], [3], [4]. Nevertheless, these ground-based instruments are extremely lacking in many parts of the world, including oceans, mountainous regions, and sparsely populated remote areas. In addition, precipitation might not show at all due to the issues, such as under catch and blind zone or may show with a reduced intensity. Meteorological satellites have irreplaceable advantages in terms of time continuity and observation scale, etc. [5], [6], [7]. Visible light and infrared (IR) waves have poor penetration of clouds and precipitation. In contrast, microwaves can penetrate clouds of a certain thickness and even reach the surface, which can directly reflect the microphysical features of precipitation clouds. The meteorological satellite series Fengyun-3 (FY-3) is the second generation of polar-orbiting meteorological satellites in China, which provides rich meteorological satellite data [8], [9], [10]. The microwave humidity and temperature sounder (MWHTS), loaded on FY-3D, can estimate the vertical distribution of global temperature and humidity, inverse precipitation, and improve the monitoring and warning ability of rainstorms, typhoons, and other weather [11], [12], [13]. Similarly, as another payload of FY-3D, microwave radiation imager (MWRI) can monitor typhoons and other severe convective weather by providing thermal radiance emitted from the surface and thus quantitative rainfall information [14], [15].

Precipitation retrieval methods can be roughly divided into three categories: statistical methods (e.g., [16] and [17]), physical methods (e.g., [18], [19], and [20]), and combined physical–statistical methods (e.g., [21]). A direct relation between the satellite observations and precipitation can be made using a statistical algorithm, reducing the need for any physical assumption and offering researchers unprecedented opportunities to process large-volume datasets in near-real-time systems, such as Earth-observing satellite data [22], [23], [24]. Min et al. [1] estimated summer precipitation from Himawari-8 and global forecast system based on random forest, an advanced machine learning (ML) method. Chen [25] proposed an ML system for precipitation estimation using satellite and ground radar network observations. Using FY-3 C MWHTS, He et al. developed a precipitation inversion algorithm with many neural network (NN) estimators. These estimators are trained and evaluated using the validated global reference physical model NCEP/WRF/ARTS [26]. On this basis, Li [27] constructed a linear regression model and neural network model to carry out global precipitation detection, precipitation inversion, and typhoon simulation.

The above studies showed that emerging ML techniques had been successfully and extensively applied in precipitation

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
rate retrieval, increasingly becoming an important direction of precipitation estimation. Nevertheless, few studies used modern deep learning (DL) techniques [28], where multiple processing layers representing multiple levels of abstraction exist between the input and the output of a DL NN [29], [30]. In addition, there is a key problem in the current DL algorithm for precipitation retrieval: Although the trained model may have high accuracy and beat previous benchmarks, they are only applied to some specific datasets. When used for new tasks or areas, the performance of this model will decline significantly. On the other hand, traditional DL needs plenty of training data to retrain the specific model for a new area or task, which will consume unnecessary time. In this case, transfer learning (TL) can solve this problem in theory when the new task is related to or similar to the original task [31], [32], resulting in that the performance of the TL model reaches or exceeds the level of the new model as far as possible. The new model is trained directly based on new target or task data. The fundamental idea of TL is to extract the feature from a previous or source task and apply the extracted knowledge to a new or target task. A conceptual metaphor is that a person who has learned to recognize cats does not need to completely relearn the skills of recognizing animals to learn how to recognize dogs. Nevertheless, TL is rarely applied to precipitation inversion with TB data, and its feasibility needs to be further explored. Also, the previously constructed models are single-task models and rarely focused on combining separated tasks for improved precipitation estimation.

This study aims to advance DL-based TL for precipitation retrieval. The structure of the article is arranged as follows.

1) Based on the matched brightness temperature (TB) data over the Northwest Pacific area, this article first constructs two multitask DL models [multilayer perceptron (MLP) and convolutional neural network (CNN)] and shows the retrieval advantages by incorporating two microwave spaceborne payloads (MWHTS and MWRI).

2) Then this article illustrates whether multitask models which detect precipitation flags and retrieve precipitation rates simultaneously can improve learning efficiency and retrieval performance when compared to training a separate model for each task. This lies in contrast to single-task learning typically employed in previous ML architectures.

3) After training two multitask DL models, the features learned from these models are extracted and applied to the precipitation retrieval area in the Northeast Pacific, a new target region with only a small amount of data. This is the basic idea of TL. Because the model layer’s features are gradually evolving from general to specific, and the TL model needs to be transferred to the general feature level, this leads to many problems such as whether a layer’s features can be quantified as general or specific. To this end, this article develops a flexible network framework with two TL methods (freeze and fine-tuning), which can avoid the underlying implementation of the model, and only needs to focus on the logical structure of the model and change the transfer layers $N$ of the network to explore these issues and evaluate the feasibility and rationality of TL for precipitation retrieval.

| TABLE I | DETAILED PARAMETERS OF MWHTS |
| --- | --- |
| Channel | Center frequency (GHz) | Polarization model | Weight peak height (hPa) |
| 1 | 89.0 | V | Window |
| 2 | 118.75±0.08 | H | 30 |
| 3 | 118.75±0.2 | H | 50 |
| 4 | 118.75±0.3 | H | 100 |
| 5 | 118.75±0.8 | H | 250 |
| 6 | 118.75±1.1 | H | 350 |
| 7 | 118.75±1.5 | H | Surface |
| 8 | 118.75±3.0 | H | Surface |
| 9 | 118.75±5.0 | H | Surface |
| 10 | 150.0 | V | Window |
| 11 | 183.31±1.0 | H | 300 |
| 12 | 183.31±1.8 | H | 400 |
| 13 | 183.31±3.0 | H | 500 |
| 14 | 183.31±4.5 | H | 700 |
| 15 | 183.31±7.0 | H | 800 |

| TABLE II | DETAILED PARAMETERS OF MWRI |
| --- | --- |
| Channel | Center frequency (GHz) | Polarization model |
| 1 | 10.65 | V |
| 2 | 10.65 | H |
| 3 | 18.7 | V |
| 4 | 18.7 | H |
| 5 | 23.8 | V |
| 6 | 23.8 | H |
| 7 | 36.5 | V |
| 8 | 36.5 | H |
| 9 | 89.0 | V |
| 10 | 89.0 | H |

II. Datasets and Input Selection

A. MWHTS Data and MWRI Data

The research data used in this article include MWHTS data (Level-1 products), MWRI data (Level-1 products), and IMERG observations (Level-3 products). When FY-3D operates in polar orbit, the MWHTS and MWRI data can be obtained every 102 min [33]. The Level-0 voltage data is first unpacked, the instrument status and data are inspected, and then through the calibration process, Level-1 TB data can be obtained. The specific parameters of MWHTS and MWRI are shown in Tables I and II.

B. IMERG Data

With the granule size of 30 min, the IMERG Late Level-3 product (Version 6) used in this article provides precipitation estimates at 0.1° by the integrated multisatellite retrievals for GPM (IMERG) [34]. IMERG is intended to intercalibrate, merge, and interpolate satellite microwave precipitation estimates, together with microwave-calibrated IR satellite estimates, and precipitation gauge analyses. After obtaining the satellite observations, IMERG-Late Run is computed about 14 h after observation time using both forward and backward morphing. A number of these
evaluations have found the overall improved performance of GPM IMERG products over the TMPA 3B42 predecessor [35].

C. Temporal and Spatial Matching, and Preprocessing

Fig. 1 shows the source (orange rectangle) and target study domains (red rectangle), which represent Northwest Pacific area and Northeast Pacific area for precipitation retrieval in this article. MWHTS data (Level-1 products), MWRI data (Level-1 products), and IMERG observations (Level-3 products) are selected from July 1, 2019 to August 15, 2020, for precipitation retrieval. Before precipitation retrieval, earth-observing satellite data are matched and preprocessed. First, according to the channel data integrity quality and earth observation TB quality score fields, MWHTS and MWRI data are read, and the TB, geographical location, and time information of each channel are extracted; meanwhile, the precipitation, geographic location, and time information are extracted from IMERG observations using precipitationQualityIndex, PrecipitationCal data field, etc. Finally, data obtained from the previous two steps are matched according to the following rules:

1) Select the TB data between 50 and 340 K to remove the irrational data.
2) Select the precipitation data <100 mm/h.
3) Select the data over the source (0°N–40°N, 110°E–155°E) and target study domains (5°S–40°N, 180°W–120°W) according to LandSeaMask flag of MWHTS/MWRI (orange and red rectangles in Fig. 1).
4) Given the nadir point size of 150.0 GHz channel (MWHTS) with 16 km × 16 km, 89.0 GHz channel (MWRI) with 9 km × 15 km, and the interval of half an hour for IMERG observations, a global grid data with a resolution of 0.1° × 0.1°, FY data and GPM product are matched for each 15-min window with a spatial resolution of 0.15° × 0.15°.
5) TB data of each channel was normalized before training the model.

In total, after matching the FY-3D data and the IMERG data, 3 667 211 pairs of matched datasets were obtained.

D. TB Input

A crucial step in precipitation retrieval is to determine the optimal input for the model. To this end, different brightness temperature-derived variables have been considered. In addition to using TB in the different MWHTS and MWRI channels, the TB differences between each two 183-GHz channels of MWHTS, 183.31 ± 1.0 to 183.31 ± 3.0 GHz, 183.31 ± 1.0 to 183.31 ± 7.0 GHz, and 183.31 ± 3.0 to 183.31 ± 7.0 GHz, are also selected as the combination factors for input. Three channels (183.31 ± 1.0 GHz, 183.31 ± 3.0 GHz, and 183.31 ± 7.0 GHz) have different weight heights, and channels away from 183.31 GHz have lower weight peak heights (Table I). With the increase in precipitation rate, the channel away from 183.31 GHz is more vulnerable to the scattering of ice particles in the cloud. Therefore, the TB observed by the 183.31 ± 7.0 GHz channel is lower than that of the other two channels. TB differences between each two 183-GHz channels are conducive to characterizing precipitation of different intensities and further improve the retrieval accuracy, which plays a promoting role in precipitation retrieval.

III. METHODOLOGIES

A. Retrieval Framework

Fig. 2 shows the overall framework of precipitation retrievals in this article. First, the MWHTS data, MWRI data, and IMERG observations are matched and preprocessed to control the data quality (Block 1 in Fig. 2). Then, combined with the IMERG
observations, in addition to using TB in different MWHTS and MWRI channels, TB differences are proposed as inputs (Block 2). TB data matched with precipitation product is the input dataset (Block 3). The input vector size is $28 \times 1$; this is because TB data at 25 different channels from MWHTS and MWRI and 3-TB difference are taken as inputs. For an accurate performance evaluation, a proper data splitting into a training set (accounts for 80\%), a test set (10\%), and a verification set (10\%) are required. The training set is used for data samples of model fitting, and the test set is a separate sample set in the process of model training, which can be used to preliminarily evaluate the ability of the model and monitor whether the model is overfitted. The validation set is used to evaluate the retrieval ability of the trained model. The outputs of the multitask model are precipitation probability and precipitation rate (Block 5). The precipitation flag (labeled 0 or 1) is determined with rain probability (threshold is 50\%). Finally, if the precipitation flag is labeled as 1, the retrieved precipitation rate is taken as the final precipitation rate. Otherwise, set the precipitation rate to 0 mm/h (no precipitation).

**B. Deep Learning Models**

This study aims to advance DL-based TL for precipitation retrieval. DL models (Fig. 3) play a key role in TL.

The MLP model [36] established in this article uses three dense layers (from layer 2 to layer 4; Block 1 in Fig. 3) with 64 neurons. In addition, the dropout regularization layer (layer 5; Block 1 in Fig. 3) is used to improve the robustness of the NN. Two subnetworks (layers 8 and 9) at the end of the dense layers 6 and 7 output, respectively, the precipitation probability and the precipitation rate. Indeed, instead of training a different model for each task, this study trains one multitask model capable of outputting both the classification and regression results simultaneously. This approach is justified because a network able to retrieve precipitation rate should also be able to estimate the precipitation possibility to a certain extent [37]. The DL model needs to map the output to a value between 0\% and 100\% for detecting precipitation probability. For layer 9, the activation function is $Sigmoid$. Precipitation rate retrieval is the regression problem; so there is no activation function for layer 8. After constructing the above network layers, the $Adam$ optimizer is adopted to continuously reduce losses.

Neurons in the MLP model are used to describe the features of a dense layer. Similarly, the kernel of the CNN model (in this article, kernel size is set to $5 \times 1$) will determine the features of the CNN layer [38]. As a window, the kernel slides the TB and TB difference (input’s vector size is set to $28 \times 1$, detailed information in Section III-A) from top to bottom. This operation is called convolution ($conv1D$), and its function is to transform the TB data into a feature map, which can be considered as a representation of what is learned from the inputs. The feature map is transformed into another map in the subsequent layer, and the process continues. The number of feature maps created by each $conv1D$ operation is controlled by the filter numbers. The $maxpooling1D$ operation after the convolution layer is to compress each feature map, and the function of the flatten layer (layer 7; Block 2 in Fig. 3) is to transform the stack of feature maps into a vector format. The CNN model established in this article uses three convolution and two pooling layers (layers 2–6; Block 2 in Fig. 3) with 64 feature maps in each layer and uses the $Adam$ optimizer. CNN model also adds a dropout layer (layer 8), and the dense layers (layers 9–12) for precipitation flag and precipitation rate are consistent with the MLP model.
C. Transfer Learning

For the traditional framework of ML or DL, a model is trained on the basis of plenty of data. Then the model is applied to solve specific tasks. Once the feature space distribution changes or is faced with new tasks or areas, the model needs to be retrained, which will consume an inordinate amount of time and waste the valuable features learned by the existing model. In this case, TL can solve this problem in theory when the new task is related or similar to the original task. The model that extracts the valuable features from the source area and is applied to the target area is called the TL model, and the essence of TL is to train/fine-tune the model based on a pretrained model A (Fig. 4). For example, a person who has learned to recognize cats does not need to completely relearn the skills of recognizing animals to learn how to recognize dogs. These things are inherently related. When a person masters a skill, he can naturally learn new related skills faster.

In the training process, the model learns the general features first and then features become more and more specific [31]. For example, pretrained A can identify cats, while control model B can identify dogs. The purpose of the TL model AtoB is also to identify dogs. Model A is trained on the source data (such as pictures of cats), and then the features of the first $N$ layers are readjusted or transferred to the model AtoB, which is trained by target data (such as pictures of dogs). If these features are general, TL often works, which means that the features are
suitable for both source task and target task. On the other hand, if model AtoB cannot recognize dogs or the recognition accuracy is far lower than model B, it can be inferred that the first $N$ layer features of model A learn the specific features to identify cats; so it is not feasible to transfer these features to model AtoB.

Therefore, finding out the general features between the source and target dataset is very important. According to the change of retrieval accuracy, whether a layer learns general or specific features and whether the learning of specific features happens suddenly or gradually can be measured.

In this article, features refer to the weights of each layer. Two TL models (MLP-TL and CNN-TL) are constructed as follows: According to the retrieval framework shown in Fig. 2, model A is trained from Northwest Pacific (source area), then the weights of model A are transferred to TL model AtoB trained on Northeast Pacific Ocean (target area). When training the TL model, this article adopts the following two methods.

1) Freeze: Take the weight parameters of the first $N$ layers of model A as the initial parameters of the TL model AtoB, and then freeze the weights of the first $N$ layers. When training the AtoB, only the weights of the remaining layers will be adjusted.

2) Fine-tuning: Instead of freezing the first $N$ layers, take the weight parameters of the first $N$ layers of model A as the initial parameters of TL model AtoB, and continuously adjust the weight of all layers of AtoB during the training process.

The MLP model has three dense layers and a dropout layer followed by two dense layers (Block1; Fig. 3). Parameters of layers 6, 7, 8, or 9 are very few and will learn specific features in the process of training. TL is not feasible when these layers are set as the remaining layers. To this end, for MLP, from layer 2 to layer 3, and then to layer 4 is frozen or fine-tuned (layer 1 is the input layer; detailed information is shown in Fig. 3). The remaining layers were randomly initialized. It is similar to the CNN model (from layer 2 to layer 3, and then to layer 4, until layer 7 is frozen or fine-tuned).

D. Training Strategy and Model Evaluation Criteria

For each data sample, the sparse categorical cross-entropy loss ($L_{scc}$) and the mean squared error loss ($L_{mse}$) are computed respectively for detecting precipitation flag and retrieving precipitation rate. To train multitask model, this article combines the different task losses into a single one ($L$), which is done by choosing a linear combination between them as follows:

$$L = \sum \alpha_{scc} L_{scc} + \alpha_{mse} L_{mse}.$$  \hfill (1)

In the above formula, the parameter $\alpha$ is tuned to scale properly the different losses and to maximize each task’s performance. The epoch of each model is set to 40. Fig. 5 shows the retrieval accuracy of the MLP model with different weight ratios. It can be seen that when the ratio of classification loss weight and regression loss weight is 0.6:0.4 ($\alpha_{scc} : \alpha_{mse} = 0.6:0.4$), the retrieval scores are the highest (equations of accuracy and corr are shown in Table III). The experiment on the CNN model is also carried out, and the weight ratio is the same as the MLP model.
This study evaluates the model abilities by computing various scores (Table III) already used in previous studies [39]. For detecting precipitation, accuracy (Acc) is selected as the evaluation criteria. Here, tp is the number of true positives, fp is the number of false positives, tn is the number of true negatives, and fn is the number of false negatives. For precipitation retrieval, mean absolute error (MAE) and correlation coefficients (Corr) are selected as the evaluation criteria. Here, $\hat{y}_i$ is the predicted value of the $i$th sample; $y_i$ is the corresponding truth value; $n$ is the total number of samples; Cov($y$, $\hat{y}$) represents the covariance of $y$ and $\hat{y}$; $D(y)$ and $D(\hat{y})$ are the variances of $y$ and $\hat{y}$, respectively. All these classification and regression scores are computed from the contingency table made from the retrieval results of DL models.

IV. RESULTS AND ANALYSIS

A. Performances of Deep Learning Models

For two multitask DL models constructed in this article, retrieval advantages by incorporating two microwave spaceborne payloads (MWHTS and MWRI) are first explored. For the same model, we have the following:

1) Only TB and TB difference at different MWHTS channels (the size of the input vector is $18 \times 1$, top two modules in Block 2, Fig. 2).

2) Only TB at different MWRI channels (the size of the input vector is $10 \times 1$, bottom module in Block 2, Fig. 2).

3) Both MWHTS and MWRI TB data and TB difference (the size of the input vector is $28 \times 1$, input (1) plus input of (2), Block 2 in Fig. 2) are selected as inputs.

This will allow the separate quantification of the benefits of MWHTS and MWRI. The results of precipitation retrieval using the verification set are shown in Table IV and Fig. 6 (MWHTS and MWRI are taken as inputs). Multitask models are proposed to detect precipitation flags and retrieve precipitation rates simultaneously. This leads to whether multitask model is better than training a separate model for each task. To this end, the retrieval results of multitask learning models are compared with the single task model trained for each task (Table IV).
TABLE VI
MLP-TL PERFORMANCE EVALUATED BY VERIFICATION SET FROM NORTHEAST PACIFIC

| Method           | Evaluation criteria | MLP (Control Model B) Trained by 1654732 data | Network layer N of MLP-TL Trained by 413683 data |
|------------------|---------------------|-----------------------------------------------|--------------------------------------------------|
|                  |                     | 2:Dense 3:Dense 4:Dense                       |                                                  |
| Freeze           | Acc                 | 0.95 0.95 0.94                                |                                                  |
|                  | MAE                 | 0.35 0.36 0.38 0.40                          |                                                  |
|                  | Corr                | 0.81 0.75 0.68 0.65                          |                                                  |
| Fine tune        | Acc                 | 0.95 0.95 0.95 0.95                          |                                                  |
|                  | MAE                 | 0.35 0.31 0.32 0.34                          |                                                  |
|                  | Corr                | 0.81 0.83 0.79 0.76                          |                                                  |

TABLE VII
CNN-TL PERFORMANCE EVALUATED BY VERIFICATION SET FROM NORTHEAST PACIFIC

| Method           | Evaluation criteria | CNN (Control Model B) Trained by 1654732 data | Network layer N of CNN-TL Trained by 413683 data |
|------------------|---------------------|-----------------------------------------------|--------------------------------------------------|
|                  |                     | 2:Conv1D 3:MaxPool 4:Conv1D 5:MaxPool 6:Conv1D 7:Flatten |                                                  |
| Freeze           | Acc                 | 0.95 0.95 0.95 0.95 0.95 0.95 0.94            |                                                  |
|                  | MAE                 | 0.33 0.35 0.37 0.38 0.39 0.41 0.42            |                                                  |
|                  | Corr                | 0.80 0.78 0.74 0.69 0.66 0.63 0.62            |                                                  |
| Fine tune        | Acc                 | 0.95 0.95 0.95 0.95 0.95 0.95 0.94            |                                                  |
|                  | MAE                 | 0.33 0.32 0.30 0.30 0.31 0.32 0.33            |                                                  |
|                  | Corr                | 0.80 0.81 0.82 0.79 0.77 0.75 0.73            |                                                  |

For MLP single-task model, network layers 6, 8, or 7, 9 are removed from the model structure (see Block 1 in Fig. 3) to detect precipitation flags or retrieve precipitation. CNN model also adopts a similar processing method.

Compared with using the single payload MWHTS/MWRI observations as the independent source for inputs, the classification results (detect whether there is precipitation or not) indicate the retrieval advantages by incorporating two microwave spaceborne payloads. It can be seen that no matter which DL model, the accuracy of precipitation identification is higher than 94%. Good scores imply that DL models can accurately detect precipitation areas and particularly suited for precipitation detection.

No matter which DL model is applied, regression results (precipitation rate retrieval) also show joint inversion has the best performance. MLP model has the highest retrieval accuracy, with MAE and Corr of 0.31 mm/h and 0.80, respectively, and the performance of the CNN model is almost as good as MLP, with MAE and Corr of 0.32 mm/h and 0.79, respectively. This indicates that the CNN model commonly used in the 2-D or multidimensional image fields can also be applied to the 1-D input (such as TB at multichannels in this article) model to retrieve precipitation.

In addition, after optimizing the loss function weight of each task, each separated task in multitasking DL models can outperform the single-task models, especially for regression task. This multitasks approach is justified because a network able to retrieve precipitation rate should also be able to estimate the precipitation possibility to a certain extent [37]. It can be considered as an approach to inductive knowledge transfer that improves generalization by sharing the domain information between complementary tasks (two tasks are both related to precipitation, and precipitation identification can be considered as a preliminary step of precipitation inversion and is considered the key to obtaining good inversion performance [40]). It does this by using shared weights or other representations to learn multiple tasks—what is learned from one task can help learn other tasks during model training [41].
For a qualitative evaluation, visual inspections (Fig. 7) of the precipitation retrieval are performed and compared to the IMERG products, MWHTS, and MWRI observations (MWHTS and MWRI are taken as inputs simultaneously). The location, distribution, and structure can be clearly seen in the precipitation retrieval maps obtained by DL models [panels (d) and (e) in Fig. 7], which show a consistent spatial pattern and high correlation with IMERG product. Overall, the DL models proposed in this article are scientific and effective.

B. Outcomes of Transfer Learning

In this article, it is not operational and practical to directly apply the model based on matched data from the Northwest Pacific area to retrieve the precipitation from the Northeast Pacific area. This inference is further verified in Table V. To this end, this section verifies the effect of TL. According to the method in the TL section, different $N$ values are selected to train MLP-TL and CNN-TL models (Fig. 8) and compare these models with model B (trained by 1,654,732 data). Each TL model adopts two methods, freeze and fine-tuning. Tables VI and VII show the regression and classification scores of MLP-TL and CNN-TL models after TL under different $N$ values, which are trained by a small amount of data (413,683 in total) from Northeast Pacific and evaluated by verification set from Northeast Pacific.

As can be seen from Tables VI and VII, TL models show varying degrees of negative migration adopting freeze method. That is, the source task has a negative effect on the target domain task. When $N = 2$, MLP-TL shows a slight degree of negative migration. When $N = 2$, CNN-TL also shows a slight degree of negative migration. Compared with model B, MAE and Corr values of the MLP-TL model only increased by 0.01 mm/h and 0.06, MAE value of the CNN-TL model increased from 0.33 to 0.35 mm/h, and Corr value only decreased by 0.02, indicating that MLP-TL and CNN-TL with these $N$ values basically achieved the same performance as model B. With the increase of the $N$ value, the remaining adjustable parameters of the model become less and the phenomenon of negative migration becomes increasingly serious. TL is not feasible with the freeze method.

The encouraging results show that the performance of the TL model using the fine-tune method has reached or even exceeded that of the comparison model, which is trained by four times more data (Tables VI and VII). With the increase of $N$, the retrieval accuracy decreases slightly. The improvement in performance is due to the similarity between the two datasets and the readjustment of model weights. The experiment is equivalent to training on the basis of the weight of model A rather than from scratch, and the effect of source data played a positive role. TL model can not only utilize features learned from the original model but also learn the precipitation features in the Northeast Pacific during the training process. In order to further highlight the feasibility and rationality of TL, two precipitation cases are shown in Fig. 9. Choosing $N = 2$ (dense layer, Table VI), MLP-TL model with fine-tuning is applied to inverse precipitation [Fig. 9(a)], and compared with the control model B [Fig. 9(b)]. Similarly, the CNN-TL model with fine-tuning...
is applied to retrieve precipitation [Fig. 9(d)] in the case of $N = 3$ (MaxPool layer; Table VII). Results show a consistent spatial pattern and high correlation between GPM IMERG product, corresponding model B observations, and precipitation retrieved by two TL models. Fig. 10 shows the retrieval performance using MLP-TL/CNN-TL and MLP/CNN models, respectively. It can be known that MLP-TL performs better than MLP, and $\text{Corr}$ is increased by 0.02. Similar phenomena are also shown for CNN-TL and CNN models.

Overall, the main advantage of using TL is the time reduction and efficiency improvements. The MLP-TL and CNN-TL models perform well in precipitation retrieval with the fine-tuning method, indicating the feasibility and rationality of the TL for precipitation retrieval. Compared with the freeze method, the fine-tune method can continuously optimize the inversion accuracy by adjusting the migration layer parameters during the training process. Simultaneously, the precipitation identification accuracy of TL model is over 94\% as seen in Tables VI and VII. This result shows the good identification accuracy is strongly robust with the transfer layer restrictions.

V. SUMMARY AND FUTURE WORK

In this article, MLP and CNN are constructed to detect precipitation flags and retrieve precipitation rates simultaneously. Compared with using single payload observations as an independent source, the results show the retrieval advantages of incorporating two microwave spaceborne payloads (MWHTS and MWRI). Simultaneously, after optimizing the loss function, each separated task in multitask DL models can outperform the single-task models, especially for regression task. Additional visual inspection shows a consistent spatial pattern and high correlation of precipitation between IMERG product and retrieved precipitation rate by models. Overall, multitask DL models are scientific and effective.

However, a DL model trained using TB data in one region may not be suitable for other regions with different precipitation characteristics and environments. On this basis, this article explores the feasibility of TL for precipitation retrieval, which aims to extract the features learned from a model trained from the source area and apply these features to the precipitation retrieval over the target area. To this end, this article develops a flexible network framework with two TL methods (freeze and fine-tuning), which can avoid the underlying implementation of the model and change the transfer layers $N$ of the network. Pretrained models (MLP and CNN) trained using TB data in the Northwest Pacific area are used as the benchmark, based on which two TL models are developed, namely, MLP-TL and CNN-TL. The encouraging results show that the performance of the TL model using the fine-tune method has reached or even exceeded that of the comparison model (the DL model without TL). Besides, the training data used for DL is over four times the volume of that used for TL. The advantage of TL is the time reduction and efficiency improvements. Simultaneously, the classification accuracy of TL models can accurately detect the precipitation area, showing the strong robustness and generalization ability of multitasking TL models. Overall, TL for precipitation retrieval is feasible and rational.

Nevertheless, the specific problems such as precipitation retrieval, how to better quantify the common features between
source and target fields, and how to design the most appropriate algorithm to extract and transfer valuable features have not yet been thoroughly explored. In addition, how to avoid negative transfer has become a difficult issue. Reducing the feature structure transferred between domains can alleviate this problem, such as sharing the prior probability of the model among domains instead of sharing the model parameters. These efforts avoid negative transfer by reducing knowledge transfer, but this usually falls into the dilemma of under-transfer. Because DL models in this article rely on the passive microwave radiometer to retrieve precipitation, future work can focus on the flexible application of this TL framework that saves training time in other meteorological fields and the introduction of radar data and other multisource data for joint retrieval.

REFERENCES

[1] M. Min et al., “Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 5, pp. 2557–2570, May 2019.

[2] H. Chen and V. Chandrasekar, “The quantitative precipitation estimation system for Dallas–Fort Worth (DFW) urban remote sensing network,” J. Hydro. Appl. Weather Radar, vol. 531, pp. 259–271, 2015.

[3] H. Chen, V. Chandrasekar, and R. Bechini, “An improved dual-polarization radar rainfall algorithm (dop2:0): Application in NASA ILSF field campaign,” J. Hydrometeorol., vol. 18, no. 4, pp. 917–937, 2017. [Online]. Available: https://journals.ametsoc.org/view/journals/hymet/18/4/jhm-d-16-0124_1.xml

[4] R. Cifelli, V. Chandrasekar, H. Chen, and L. E. Johnson, “High resolution radar quantitative precipitation estimation in the San Francisco bay area: Rainfall monitoring for the urban environment,” J. Meteorol. Soc. Japan. Ser. II, vol. 96A, pp. 141–158, 2018.

[5] F. Sun, M. Min, D. Qin, F. Wang, and J. Hu, “Refined typhoon geometric center derived from a high spatiotemporal resolution geostationary satellite imaging system,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 4, pp. 499–503, Apr. 2018.

[6] H. Chen, V. Chandrasekar, T. Tan, and R. Cifelli, “Rainfall estimation from ground radar and trmm precipitation radar using hybrid deep neural networks,” Geophys. Res. Lett., vol. 46, no. 17–18, pp. 10669–10678, 2019.

[7] K. Liu, J. He, and H. Chen, “Precipitation retrieval from Fengyun-3D microwave humidity and temperature sounder data using machine learning,” Remote Sens., vol. 14, no. 2022, Art. no. 848.

[8] R. You, S. Gu, Y. Guo, X. Wu, H. Yang, and W. Chen, “Long-term calibration and accuracy assessment of the FengYun-3 microwave temperature sounder radiance measurements,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 12, pp. 4854–4859, Dec. 2012.

[9] X. Song et al., “Estimation of land surface temperature using FengYun-3E (FY-3E) data: A case study of the source area of the yellow river,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 8, pp. 3744–3751, Aug. 2017.

[10] X. Wang and X. Zou, “Quality assessments of chinese FengYun-3A microwave temperature sounder (MWTs) measurements,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 12, pp. 4875–4884, Dec. 2012.

[11] X. Zou, Y. Ma, and Z. Qin, “Fengyun-3B Microwave humidity sounder (MWHS) data noise characterization and filtering using principle component analysis,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 12, pp. 4892–4902, Dec. 2012.

[12] Z. Wang et al., “Performance analysis of microwave humidity and temperature sounder onboard the FY-3D satellite from prelaunch multangle calibration data in thermal/vacuum test,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 3, pp. 1664–1683, Mar. 2019.

[13] Z. Wang et al., “Performance analysis of microwave humidity and temperature sounder onboard the FY-3D satellite,” IEEE Trans. Instrum. Meas., vol. 69, no. 9, pp. 6582–6594, Sep. 2020.

[14] H. Yang et al., “The FengYun-3 microwave radiation imager on-orbit verification,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 11, pp. 4552–4560, Nov. 2011.
[37] A. Kendall, Y. Gal, and R. Cipolla, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,” in Proc. IEEE Conf. Comput. Vis. Pattern Recogn., 2018, pp. 7482–7491.

[38] X. Lei, H. Pan, and X. Huang, “A dilated CNN model for image classification,” IEEE Access, vol. 7, pp. 124087–124095, 2019.

[39] L. Beusch, L. Foresti, M. Gabella, and U. Hamann, “Satellite-based rainfall retrieval: From generalized linear models to artificial neural networks,” Remote Sens., vol. 10, no. 6, 2018, Art. no. 939.

[40] P.-E. Kirstetter, N. Viltard, and M. Gosset, “An error model for instantaneous satellite rainfall estimates: Evaluation of BRAIN-TMI over west Africa,” Quart. J. Roy. Meteorological Soc., vol. 139, no. 673, pp. 894–911, 2013.

[41] R. A. Caruana, Multitask Learning. Dordrecht, Netherlands: Kluwer Academic Publishers, 1998.

Xi Kangwen Liu was born in ShanDong, China. He received the B.S. degree in computer science and technology from Qufu Normal University, Qufu, China, in 2020. He is currently working toward the master’s degree in computer application technology with the University of Chinese Academy of Sciences, Beijing, China.

His research interests include remote sensing of precipitation and deep learning.

Jieying He (Member, IEEE) received the Ph.D. degree specializing in computer science and applied research from National Space Science Center, Chinese Academy of Sciences, Beijing, China, in 2012. She is currently a Researcher and Supervisor of Ph.D. She focuses on design and development of satellite-based microwave radiometer, calibration, and quantification retrievals for earth and planetary exploration. She is currently the Deputy Chief Designer of microwave humidity sounder for Fengyun-3 satellite and salinity radiometer for ocean satellites, and the leader of the conceptual research project of the Space Science (Phase II) Mission. She is also one of the 9th/10th China–US Young Leaders in Space Science, and excellent members of Youth Innovation Promotion Association of Chinese Academy of Sciences.

Haonan Chen (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Colorado State University (CSU), Fort Collins, CO, USA, in 2017. He has been an Assistant Professor with the Department of Electrical and Computer Engineering, CSU since 2020. He is also an Affiliate Faculty with the Data Science Research Institute (DSRI), CSU. Before joining the CSU faculty, he worked with the Cooperative Institute for Research in the Atmosphere (CIRA) and the NOAA Physical Sciences Laboratory, Boulder, CO, USA, from 2012 to 2020, first as a Research Student and then a National Research Council Research Associate and a Radar, Satellite, and Precipitation Research Scientist. His research interests span a broad range of remote sensing and multidisciplinary data science, including radar and satellite observations of natural disasters, polarimetric radar systems and networking, clouds and precipitation processes, big data analytics, and deep learning.

Dr. Chen serves as an Associate Editor for the Journal of Atmospheric and Oceanic Technology, URSI Radio Science Bulletin, and the IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing.