Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong and Wang-chun Woo

{xshiab,zgaoag,lelausen,hwangaz,dyyeung}@cse.ust.hk, {wkwong,wcwoo}@hko.gov.hk

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Highlights

1. We propose the first large-scale benchmark for precipitation nowcasting called HKO-7. The benchmark includes:
   - **7-year** precipitation data from Hong Kong Observatory.
   - New performance measures which emphasize more on the heavier rainfall events.
   - A clear evaluation protocol that covers not only the offline setting but also the **online setting** of the nowcasting problem.
2. We propose a new RNN structure called *Trajectory Gated Recurrent Unit* (TrajGRU). In contrast to the *Convolutional Gated Recurrent Unit* (ConvGRU), TrajGRU uses a subnetwork to generate **a *location-variant* recurrence structure**.
3. Our experimental validation shows that:
   - All the deep learning models outperform the optical flow based models.
   - TrajGRU attains the best overall performance among all the deep learning models.
   - After applying online fine-tuning, the models tested in the online setting consistently outperform those in the offline setting.

Precipitation Nowcasting

Precipitation nowcasting refers to the problem of providing very short range (e.g., 0-6 hours) forecast of the rainfall intensity in a local region based on radar echo maps, rain gauge and other observation data as well as the Numerical Weather Prediction (NWP) models. It helps to facilitate drivers by predicting road conditions, enhances flight safety by providing weather guidance for regional aviation, and avoids casualties by issuing citywide rainfall alerts.

HKO-7 Dataset

The radar CAPPI reflectivity images, which have resolution of 480×480 pixels, are taken from an altitude of 2km and cover a $512\text{km} \times 512\text{km}$ area centered in Hong Kong. The data are recorded every 6 minutes and hence there are 240 frames per day. As rainfall events occur sparsely, we select the rainy days based on the rain barrel information to form our final dataset.

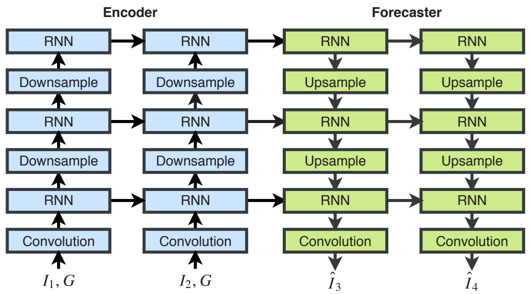
	Train	Validate	Test
Years	2009-2014	2009-2014	2015
#Days	812	50	131
#Frames	192,168	11,736	31,350

Imbalanced Data

The radar reflectivity values are converted to rainfall intensity values (mm/h) using the Z-R relationship: $\text{dBZ} = 10\log a + 10b\log R$ where R is the rain-rate level, $a = 58.53$ and $b = 1.56$. We can observe that **heavier rainfall occurs less often but has a higher real-world impact**.

Rain Rate (mm/h)	Proportion (%)	Rainfall Level
$0 \leq x < 0.5$	90.25	No / Hardly noticeable
$0.5 \leq x < 2$	4.38	Light
$2 \leq x < 5$	2.46	Light to moderate
$5 \leq x < 10$	1.35	Moderate
$10 \leq x < 30$	1.14	Moderate to heavy
$30 \leq x$	0.42	Rainstorm warning

Overall Structure



ConvGRU

$$
\begin{aligned}
\mathcal{Z}_t &= \sigma(\mathcal{W}_{xz} * \mathcal{X}_t + \mathcal{W}_{hz} * \mathcal{H}_{t-1}),\\
\mathcal{R}_t &= \sigma(\mathcal{W}_{xr} * \mathcal{X}_t + \mathcal{W}_{hr} * \mathcal{H}_{t-1}),\\
\mathcal{H}'_t &= f(\mathcal{W}_{xh} * \mathcal{X}_t + \mathcal{R}_t \circ (\mathcal{W}_{hh} * \mathcal{H}_{t-1})),\\
\mathcal{H}_t &= (1 - \mathcal{Z}_t) \circ \mathcal{H}'_t + \mathcal{Z}_t \circ \mathcal{H}_{t-1}.
\end{aligned}
$$

$\mathcal{X}_t, \mathcal{H}_t, \mathcal{R}_t, \mathcal{Z}_t, \mathcal{H}'_t$ are the input, memory state, reset gate, update gate, and new information. The recurrent connection structure is ***location-invariant***:

$$
\begin{aligned}
\mathcal{H}'_{t,:,i,j} &= f(\mathbf{W}_{hh}\text{concat}(\langle \mathcal{H}_{t-1,:,p,q} \mid (p,q) \in \mathcal{N}^h_{i,j} \rangle))\\
&= f(\sum_{l=1}^{|\mathcal{N}^h_{i,j}|} \mathbf{W}^l_{hh} \mathcal{H}_{t-1,:,p_{l,i,j},q_{l,i,j}}).
\end{aligned}
$$

TrajGRU

To capture **spatiotemporal patterns**, a location-variant recurrence structure is more reasonable, e.g,

$$
\mathcal{H}'_{t,:,i,j} = f(\sum_{l=1}^{L} \mathbf{W}^l_{hh} \mathcal{H}_{t-1,:,p_{l,i,j}(\theta),q_{l,i,j}(\theta)}), \tag{1}
$$

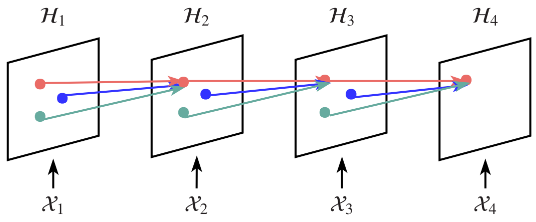
where L is the total number of local links, $(p_{l,i,j}(\theta), q_{l,i,j}(\theta))$ is the lth neighborhood parameterized by θ.
The TrajGRU uses the current input and previous state to generate the local neighborhood set for each location at each timestamp. Since the location indices are discrete and non-differentiable, we use **a set of continuous optical flows to represent these "indices"**. The main formulas of TrajGRU are given as follows:

$$
\begin{aligned}
\mathcal{U}_t, \mathcal{V}_t &= \gamma(\mathcal{X}_t, \mathcal{H}_{t-1}),\\
\mathcal{Z}_t &= \sigma(\mathcal{W}_{xz} * \mathcal{X}_t + \sum_{l=1}^{L} \mathcal{W}^l_{hz} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})),\\
\mathcal{R}_t &= \sigma(\mathcal{W}_{xr} * \mathcal{X}_t + \sum_{l=1}^{L} \mathcal{W}^l_{hr} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})),\\
\mathcal{H}'_t &= f(\mathcal{W}_{xh} * \mathcal{X}_t + \mathcal{R}_t \circ (\sum_{l=1}^{L} \mathcal{W}^l_{hh} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l}))),\\
\mathcal{H}_t &= (1 - \mathcal{Z}_t) \circ \mathcal{H}'_t + \mathcal{Z}_t \circ \mathcal{H}_{t-1}.
\end{aligned}
$$

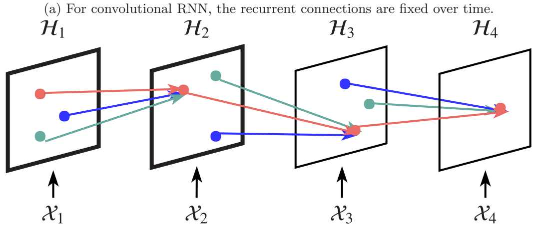
If we denote $\mathcal{M} = \text{warp}(\mathcal{I}, \mathbf{U}, \mathbf{V})$, $\mathcal{M}_{c,i,j}$ is equal to

$$
\sum_{m=1}^{H} \sum_{n=1}^{W} \mathcal{I}_{c,m,n} \max(0, 1 - |i + \mathbf{V}_{i,j} - m|) \max(0, 1 - |j + \mathbf{U}_{i,j} - n|).
$$

Comparison of ConvGRU and TrajGRU



(a) For convolutional RNN, the recurrent connections are fixed over time.



(b) For trajectory RNN, the recurrent connections are **dynamically determined**.

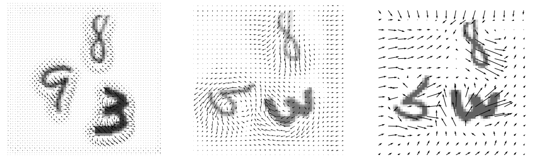
Evaluation Protocols

In the real-world scenario, the nowcasting algorithms can adopt online learning to **adapt to the newly emerging patterns dynamically**. To take into account this setting, we use two testing protocols in our benchmark: the offline setting in which the algorithm can only use 5 previous frames as input and predict 20 frames ahead and the online setting in which the algorithm is free to use all the historical data and any online learning algorithm.

Results on MovingMNIST++

MovingMNIST++ is an extension of MovingMNIST that allows **random rotations, scale changes, and illumination changes**.

	Test MSE $\times 10^{-2}$	$\sigma \times 10^{-2}$	#Param.
Conv-K3-D2	1.495	0.003	2.84M
Conv-K5-D1	1.310	0.004	4.77M
Conv-K7-D1	1.254	0.006	8.01M
Traj-L5	1.351	0.020	2.60M
Traj-L9	1.247	0.015	3.42M
Traj-L13	1.170	0.022	4.00M
Traj-L17	**1.138**	0.019	4.77M
DFN	1.461	0.002	4.83M
Conv2D	1.681	0.001	29.06M
Conv3D	1.637	0.002	32.52M



(a) Link of the first layer of the encoder.

(b) Link of the second layer of the encoder.

(c) Link of the third layer of the encoder.

Balanced Loss Measures

To deal with the data imbalance problem, we propose the *Balanced Mean Squared Error* (B-MSE) and *Balanced Mean Absolute Error* (B-MAE) measures for training and evaluation, which **assign more weights to heavier rainfalls** in the calculation of MSE and MAE. Specifically, we assign a weight $w(x)$ to each pixel according to its rainfall intensity x and then compute the weighted MSE and MAE.

HKO-7 Benchmark

We have investigated the performance of 7 models. The results are given in the following. Here r means the rainrate threshold. The CSI score is calculated as $\frac{\text{TP}}{\text{TP+FN+FP}}$ and the HSS score is calculated as $\frac{\text{TP}\times\text{TN}-\text{FN}\times\text{FP}}{(\text{TP+FN})(\text{FN+TN})+(\text{TP+FP})(\text{FP+TN})}$. In the online setting, we use AdaGrad to fine-tune the deep models.

Algorithms	CSI ↑ $r \geq 0.5$	CSI ↑ $r \geq 2$	CSI ↑ $r \geq 5$	CSI ↑ $r \geq 10$	CSI ↑ $r \geq 30$	HSS ↑ $r \geq 0.5$	HSS ↑ $r \geq 2$	HSS ↑ $r \geq 5$	HSS ↑ $r \geq 10$	HSS ↑ $r \geq 30$	B-MSE ↓	B-MAE ↓
Offline Setting												
Last Frame	0.4022	0.3266	0.2401	0.1574	0.0692	0.5207	0.4531	0.3582	0.2512	0.1193	15274	28042
ROVER + Linear	0.4762	0.4089	0.3151	0.2146	0.1067	0.6038	0.5473	0.4516	0.3301	0.1762	11651	23437
ROVER + Non-linear	0.4655	0.4074	0.3226	0.2164	0.0951	0.5896	0.5436	0.4590	0.3318	0.1576	10945	22857
2D CNN	0.5095	0.4396	0.3406	0.2392	0.1093	0.6366	0.5809	0.4851	0.3690	0.1885	7332	18091
3D CNN	0.5109	0.4411	0.3415	0.2424	0.1185	0.6334	0.5825	0.4862	0.3734	0.2034	7202	17593
ConvGRU-nobal	0.5476	0.4661	0.3526	0.2138	0.0712	0.6756	0.6094	0.4981	0.3286	0.1160	9087	19642
ConvGRU	0.5489	0.4731	0.3720	0.2789	0.1776	0.6701	0.6104	0.5163	0.4159	0.2893	5951	15000
TrajGRU	**0.5528**	**0.4759**	**0.3751**	**0.2835**	**0.1856**	**0.6731**	**0.6126**	**0.5192**	**0.4207**	**0.2996**	**5816**	**14675**
Online Setting												
2D CNN	0.5112	0.4363	0.3364	0.2435	0.1263	0.6365	0.5756	0.4790	0.3744	0.2162	6654	17071
3D CNN	0.5106	0.4344	0.3345	0.2427	0.1299	0.6355	0.5736	0.4766	0.3733	0.2220	6690	16903
ConvGRU	0.5511	0.4737	0.3742	0.2843	0.1837	0.6712	0.6105	0.5183	0.4226	0.2981	5724	14772
TrajGRU	**0.5563**	**0.4798**	**0.3808**	**0.2914**	**0.1933**	**0.6760**	**0.6164**	**0.5253**	**0.4308**	**0.3111**	**5589**	**14465**

Correlations between skill scores

Skill Scores	CSI $r \geq 0.5$	CSI $r \geq 2$	CSI $r \geq 5$	CSI $r \geq 10$	CSI $r \geq 30$	HSS $r \geq 0.5$	HSS $r \geq 2$	HSS $r \geq 5$	HSS $r \geq 10$	HSS $r \geq 30$
MSE	-0.24	-0.39	-0.39	-0.07	-0.01	-0.33	-0.42	-0.39	-0.06	0.01
MAE	-0.41	-0.57	-0.55	-0.25	-0.27	-0.50	**-0.60**	-0.55	-0.24	-0.26
B-MSE	-0.70	-0.57	**-0.61**	**-0.86**	-0.84	-0.62	-0.55	**-0.61**	**-0.86**	-0.84
B-MAE	**-0.74**	**-0.59**	-0.58	-0.82	**-0.92**	**-0.67**	-0.57	-0.59	-0.83	**-0.92**