

Machine Learning for Weather Prediction

Zhi Li

June 7, 2019

1 MIM

MIM: A Predictive Neural Network for Learning Higher-Order Non-Stationarity from Spatiotemporal Dynamics.

2 TrajGRU

2.1 Key Points

- Location-variant structure to learn spatial-temporal features
- customized loss function to train the network

2.2 Encoder-Predictor structure

The encoder-predictor can be described by series of equations shown below.

$$H_t^1 H_t^2 H_t^3 \cdots H_t^n = h(I_{t-J+1} I_{t-J+2} \cdots I_t) \quad (1)$$

$$\hat{I}_{t+1} \hat{I}_{t+2} \cdots \hat{I}_{t+K} = g(H_t^1 H_t^2 H_t^3 \cdots H_t^n) \quad (2)$$

$$\hat{I}_{t+1} \hat{I}_{t+2} \cdots \hat{I}_{t+K} = g(H_t^1 H_t^2 H_t^3 \cdots H_t^n) \quad (3)$$

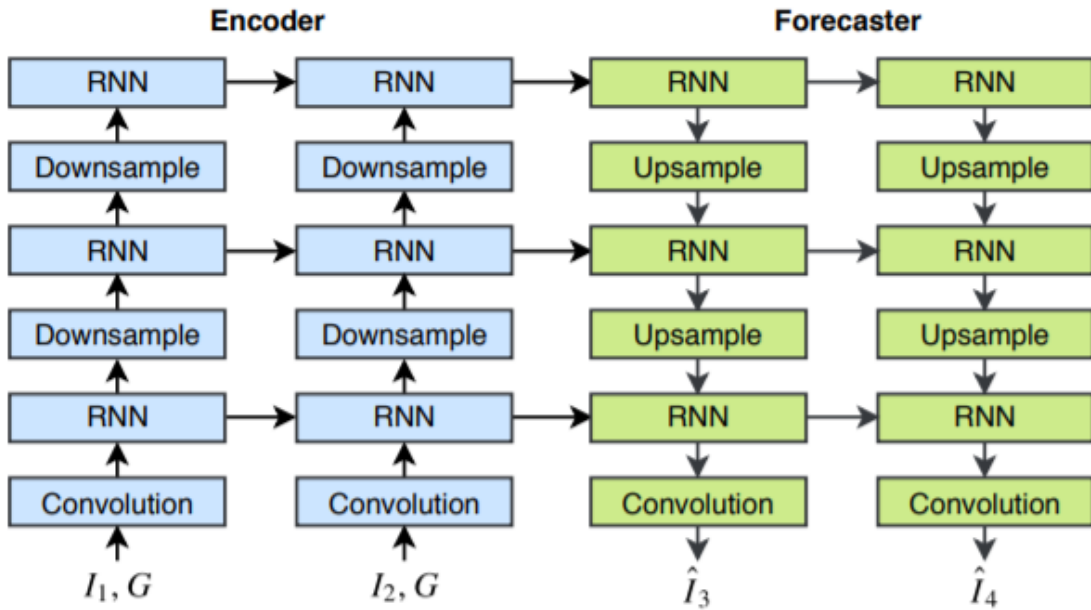


Figure 1: Network description about encoder predictor in which $n = 3, J = 2, K = 2$

Note that, in the predictor process, the flow is reversed because ‘The reason to reverse the order of the forecasting network is that the high-level states, which have captured the global spatialtemporal representation, could guide the update of the low-level states. Moreover, the low-level states could further influence the prediction.’

2.3 Convolutional GRU

The main formulas of the convGRU are given as follows:

$$\begin{aligned} Z_t &= \sigma(W_{xz} * X_t + W_{hz} * H_{t-1}) \\ R_t &= \sigma(W_{xr} * X_t + W_{hr} * H_{t-1}) \\ H'_t &= \text{LeakyRELU}(0.2)(W_{xh} * X_t + R_t \circ (W_{hh} * H_{t-1})) \\ H_t &= (1 - Z_t) \circ H'_t + Z_t \circ H_{t-1} \end{aligned} \quad (4)$$

Among these equations, \circ is the Hadamard product which also known as element-wise product, while $*$ stands for matrix convolution.

2.4 Trajectory GRU

The author takes the location into account by warp the specific location e.g. $i, j \in N_{i,j}^h$ to new location i, j in the hidden state. This is achieved by the following formula.

$$H'_{y, :, i, j} = f(W_{hh} \text{concat}(< H_{t-1, :, p, g} | (p, g) \in N_{i,j}^h >)) = f\left(\sum_{l=1}^{|N_{i,j}^h|} W_{hh}^l H_{t-1, :, p_l, :, i, j, g_l, :, i, j}\right) \quad (5)$$

$N_{i,j}^h$ is the ordered neighborhood set at location (i, j) defined by the hyperparameters of the state-to-state convolution. L is the number of links connecting both location before warped and warped. The author plotted the links illustrated below.

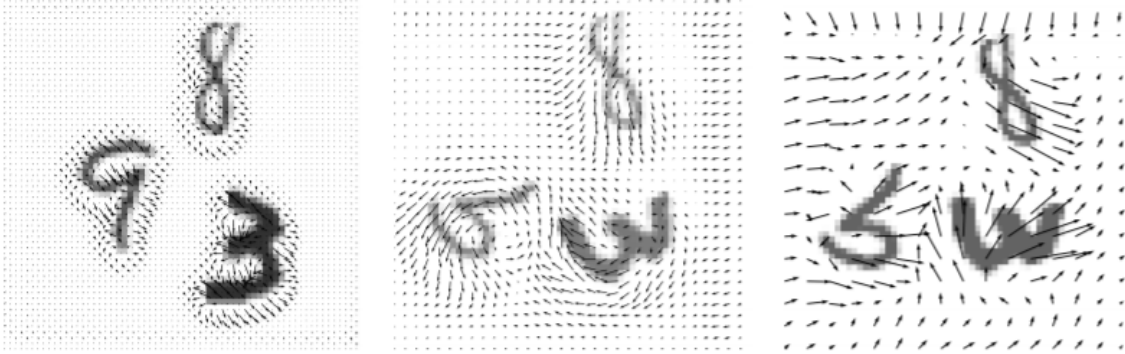


Figure 2: The plotted one of 13 links of arrow starting from each pixel to the pixel that is referenced by the link. From left to right are the first, second and third layer of the encoder. The link displayed has learned behavior of rotations