

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

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan  
Yeung

Department of Computer Science and Engineering

Hong Kong University of Science and Technology  
{xshiab,zgaoag,lelausen,hwangaz,dyyeung}@cse.ust.hk

Wai-kin Wong, Wang-chun Woo

Hong Kong Observatory

Hong Kong, China

{wkwong,wcwoo}@hko.gov.hk

# Contents

<b>Contents.....</b>	<b>2</b>
<b>Abstract .....</b>	<b>3</b>
<b>I. Introduction .....</b>	<b>4</b>
<b>II. Related Work.....</b>	<b>7</b>
<b>III. Model.....</b>	<b>8</b>
1. Encoding-forecasting Structure (Cấu trúc dự báo mã hóa) .....	8
2. Convolutional GRU (mô hình ConvGRU).....	9
3. Trajectory GRU (mô hình TrajGRU) .....	9
<b>IV. Experiments on MovingMNIST++ .....</b>	<b>13</b>
<b>V. Benchmark for Precipitation Nowcasting .....</b>	<b>14</b>
1. HKO-7 Dataset (tập dữ liệu HKO-7).....	14
2. Evaluation Methodology (đánh giá phương pháp) 15	
3. Evaluated Algorithms (đánh giá thuật toán) ....	16
4. Evaluation Results (đánh giá kết quả) .....	17
<b>VI. Conclusion and Future Work .....</b>	<b>19</b>

***Giảng viên hướng dẫn: TS. Nguyễn Thiên Bảo***

**Nhóm tìm hiểu: Nhóm 6**

**Thành viên**

Nguyễn Bá Lê An - 15110001

Nguyễn Hữu Khang - 15110062

Lê Bảo Châu - 15110078

Nguyễn Hồng Châu - 15110018

Vũ Thị Lệ - 15110071

# Abstract

Với mục tiêu đưa ra các dự báo có độ phân giải cao về lượng mưa trong khu vực, **precipitation nowcasting** đã trở thành một công nghệ quan trọng và cơ bản của nhiều dịch vụ công cộng khác nhau, từ cảnh báo mưa bão đến an toàn bay. Gần đây, mô hình Convolutional LSTM (ConvLSTM) đã được chứng minh là vượt trội so với các phương pháp dựa trên dòng quang truyền thống cho precipitation nowcasting, cho thấy các mô hình học sâu có tiềm năng rất lớn để giải quyết vấn đề. Tuy nhiên, cấu trúc tái phát tích chập trong các mô hình dựa trên ConvLSTM là bất biến vị trí trong khi chuyển động và biến đổi tự nhiên (ví dụ: xoay) nói chung là biến thể vị trí. Hơn nữa, vì since deep-learning-based precipitation nowcasting là một lĩnh vực mới nổi lên, các giao thức đánh giá rõ ràng vẫn chưa được thiết lập. Để giải quyết những vấn đề này, chúng tôi đề xuất cả một mô hình mới và một chuẩn mực cho precipitation nowcasting. Cụ thể, chúng tôi vượt xa ConvLSTM và đề xuất mô hình quỹ đạo GRU (TrajGRU) có thể chủ động tìm hiểu cấu trúc biến thể vị trí cho các kết nối định kỳ. Bên cạnh đó, chúng tôi cung cấp một benchmark bao gồm một bộ dữ liệu quy mô lớn trong hiện thực từ Đài thiên văn Hồng Kông, một training loss mới và một giao thức đánh giá toàn diện để tạo điều kiện cho nghiên cứu trong tương lai và đánh giá tình trạng của nghệ thuật.

## I. Introduction

Precipitation nowcasting đề cập đến vấn đề cung cấp dự báo trong phạm vi rất ngắn (ví dụ: 0-6 giờ) về cường độ mưa trong khu vực địa phương dựa trên bản đồ tiếng radar echo, máy đo mưa và dữ liệu quan sát khác cũng như các mô hình Dự báo thời tiết số (NWP). Nó tác động đáng kể đến cuộc sống hàng ngày của nhiều người và đóng một vai trò quan trọng trong nhiều ứng dụng trong hiện thực. Trong số các khả năng khác, nó giúp tạo điều kiện cho người lái xe bằng cách dự đoán tình trạng đường, tăng cường an toàn chuyến bay bằng cách cung cấp hướng dẫn thời tiết cho hàng không khu vực và tránh thương vong bằng cách đưa ra cảnh báo lượng mưa trên toàn thành phố. Ngoài sự phức tạp vốn có của bầu khí quyển và các quá trình động lực có liên quan, nhu cầu ngày càng tăng đối với lượng mưa thời gian thực, quy mô lớn và hạt mịn hiện nay đặt ra thêm những thách thức đối với cộng đồng khí tượng và đã làm dấy lên mối quan tâm nghiên cứu trong cộng đồng máy học [23, 25].

Các cách tiếp cận thông thường để đo mưa là sử dụng các hệ thống vận hành hiện có dựa trên dòng quang [28]. Trong một hệ thống phát sóng hiện đại ngày nay, các chuyển động của đám mây đối lưu được ước tính trước tiên từ các bản đồ tiếng vang radar quan sát được bằng dòng quang và sau đó được sử dụng để dự đoán các bản đồ tiếng vang radar trong tương lai bằng cách sử dụng tiến trình bán Lagrangian. Tuy nhiên, các phương pháp này không được giám sát theo quan điểm học máy ở chỗ chúng không tận dụng được lượng lớn dữ liệu tiếng vang radar hiện có. Gần đây, sự tiến bộ đã được thực hiện bằng cách sử dụng các kỹ thuật học sâu được giám sát [15] cho precipitation nowcasting. Shi et al. [23] phát biểu precipitation nowcasting như là vấn đề dự báo trình tự không gian và đề xuất mô hình Convolutional Long Short-Term Memory (ConvLSTM), nó mở rộng LSTM [7] bằng cách có cấu trúc tích chập trong cả quá trình trạng thái đầu vào và trạng thái đầu ra, để giải quyết vấn đề. Sử dụng liên tục tiếng vang radar cho việc training model, các tác giả chứng minh rằng ConvLSTM tốt hơn trong việc bắt các mối tương quan không gian hơn là fully-connected LSTM và đưa ra dự đoán chính xác hơn thuật toán Real-time Optical flow by Variational methods for Echoes of Radar (ROVER) hiện tại được sử dụng bởi đài quan sát Hồng Kông (HKO).

Tuy nhiên, mặc dù nỗ lực tiên phong của theo hướng hấp dẫn này, báo cáo vẫn có nhiều sự thiếu sót. Đầu tiên, mô học học sâu là chỉ đánh giá trên một tập dữ liệu tương đối nhỏ bao gồm 97 ngày và điểm phát sóng chỉ đạt ngưỡng tốc độ mưa là 0.5mm/h được đem ra so sánh. Như hệ thống phát sóng lượng mưa thế giới thực cần phải chú ý thêm ở mưa lớn như là ngày bão là thứ gây ra nhiều mối đe dọa cho xã hội, hiệu suất ngưỡng 0.5 mm/h (biểu thị đang mưa hoặc không) là không đủ cho chứng minh hiệu suất tổng thể thuật toán [28]. Trên thực tế, vì khu vực học sâu cho Precipitation Nowcasting vẫn nằm ở giai đoạn đầu, nó vẫn chưa rõ ràng những mô hình nên được đánh giá như thế nào để đáp ứng nhu cầu ứng dụng trong thế giới thực. Thứ hai, mặc dù cấu trúc tái phát tích chập sử dụng ở ConvLSTM tốt hơn cấu trúc định kỳ fully-connected trong việc nắm bắt các mối tương quan không gian, nó không phải là tối và để lại sự cải thiện. Đối với các mẫu chuyển động như xoay và nhân rộng, cấu trúc tương quan cục bộ của frame liên tiếp sẽ khác nhau so với các vị trí không gian và thời gian khác nhau. Do đó, nó không hiệu quả khi sử dụng tích chập, nó sử dụng filter vị trí bất biến để thể hiện mối quan hệ vị trí bất biến này. Những nỗ lực trước đây đã cố gắng giải quyết vấn đề bằng cách sửa đổi đầu ra của RNN từ những dự đoán thô để trở thành chuyển đổi vị trí bất biến của đầu vào, như là một luồng quang hoặc filter động cục bộ [5, 3]. Tuy nhiên, không nhiều nghiên cứu đã công bố để đưa ra vấn đề bằng cách sửa đổi cấu trúc định kỳ chính nó.

Trong bài báo này, chúng tôi hướng đến việc giải quyết hai vấn đề này bằng cách đề xuất cả điểm chuẩn (benchmark) và mô hình mới cho lượng mưa hiện tại (Precipitation Nowcasting). Đối với điểm chuẩn mới, chúng tôi

xây dựng bộ dữ liệu HKO-7 chứa dữ liệu tiếng vang radar từ năm 2009 đến 2015 gần Hồng Kông. Do các bản đồ tiếng vang của radar đến một luồng trong kịch bản trong thế giới thực, các thuật toán phát sóng có thể áp dụng phương pháp học trực tuyến để thích ứng với các mô hình mới nổi một cách linh hoạt. Để tính đến cài đặt này, chúng tôi sử dụng hai giao thức thử nghiệm trong điểm chuẩn của mình: cài đặt ngoại tuyến trong đó thuật toán chỉ có thể sử dụng một cửa sổ cố định của bản đồ tiếng vang radar trước đó và cài đặt trực tuyến trong đó thuật toán được sử dụng miễn phí tất cả lịch sử dữ liệu và bất kỳ thuật toán học tập trực tuyến. Một vấn đề khác đối với Precipitation Nowcasting là tỷ lệ các sự kiện mưa ở các ngưỡng mưa khác nhau rất mất cân bằng. Lượng mưa lớn hơn xảy ra ít thường xuyên hơn nhưng có tác động trong thế giới thực cao hơn. Do đó, chúng tôi đề xuất các biện pháp Lỗi bình phương trung bình (B-MSE) và Lỗi tuyệt đối trung bình cân bằng (B-MAE) để đào tạo và đánh giá, trong đó chỉ định nhiều trọng lượng hơn cho các trận mưa lớn hơn trong tính toán của MSE và MAE. Chúng tôi thực nghiệm thấy rằng các biến thể cân bằng của các hàm mất mát phù hợp hơn với hiệu suất phát sóng tổng thể ở nhiều ngưỡng tốc độ mưa hơn so với các hàm mất mát ban đầu. Hơn nữa, các thí nghiệm của chúng tôi cho thấy rằng đào tạo với các chức năng mất cân bằng là điều cần thiết cho các mô hình học sâu để đạt được hiệu suất tốt ở ngưỡng tỷ lệ mưa cao hơn. Đối với mô hình mới, chúng tôi đề xuất mô hình Đơn vị lặp lại theo quỹ đạo (TrajGRU) sử dụng mạng con để xuất cấu trúc kết nối trạng thái trước khi chuyển trạng thái. TrajGRU cho phép trạng thái được tổng hợp dọc theo một số quỹ đạo đã học và do đó linh hoạt hơn GRU Convolutional (ConvGRU) [2] có cấu trúc kết nối được cố định. Chúng tôi cho thấy TrajGRU vượt trội so với ConvGRU, Mạng lưới lọc động (DFN) [3] cũng như Mạng thần kinh chuyển đổi 2D và 3D (CNNs) [20, 27] trong cả bộ dữ liệu MoveMNIST ++ tổng hợp và bộ dữ liệu HKO-7.

Sử dụng bộ dữ liệu mới, các giao thức thử nghiệm, tổn thất đào tạo và mô hình, chúng tôi cung cấp đánh giá thực nghiệm trên bảy mô hình, bao gồm một mô hình cơ sở đơn giản luôn dự đoán khung cuối cùng, hai mô hình dựa trên dòng quang (ROVER và biến thể phi tuyến của nó) mô hình học tập sâu (TrajGRU, ConvGRU, 2D CNN và 3D CNN). Chúng tôi cũng cung cấp một chuẩn mực quy mô lớn cho lượng mưa hiện tại. Xác thực thử nghiệm của chúng tôi cho thấy (1) tất cả các mô hình học sâu đều vượt trội so với các mô hình dựa trên dòng quang, (2) TrajGRU đạt hiệu suất tổng thể tốt nhất trong số tất cả các mô hình học sâu và (3) sau khi áp dụng tinh chỉnh trực tuyến, các mô hình đã được thử nghiệm trong cài đặt trực tuyến luôn vượt trội so với cài đặt ngoại tuyến. Theo hiểu biết tốt nhất của chúng tôi, đây là điểm chuẩn toàn diện đầu tiên của các mô hình học tập sâu cho vấn đề phát sóng kết tủa. Ngoài ra, vì hiện tượng mưa rào có thể được xem là một vấn đề dự đoán video [22, 27], công việc của chúng tôi là lần đầu tiên cung cấp bằng chứng và biện minh rằng việc học trực tuyến có thể hữu ích cho dự đoán video nói chung.

## II. Related Work

### *Học sâu về Precipitation Nowcasting và dự đoán video*

Đối với vấn đề Precipitation Nowcasting, các yếu tố phản xạ trong bản đồ tiếng vang radar lần đầu tiên được chuyển thành hình ảnh thang độ xám trước khi được đưa vào thuật toán dự đoán [23]. Do đó, Precipitation Nowcasting có thể được xem như là một loại vấn đề dự đoán video với một camera cố định, đó là radar thời tiết. Do đó, các phương pháp được đề xuất để dự đoán các khung hình trong tương lai trong các video tự nhiên cũng có thể áp dụng cho Precipitation Nowcasting và có liên quan đến bài báo của chúng tôi. Có ba loại kiến trúc chung để dự đoán video: mô hình dựa trên RNN, mô hình dựa trên CNN 2D và mô hình dựa trên CNN 3D. Ranzato et al. [22] đề xuất mô hình dựa trên RNN đầu tiên cho dự đoán video, sử dụng RNN tích chập với hạt nhân trạng thái  $1 \times 1$  để mã hóa các khung hình quan sát được. Srivastava et al. [24] đề xuất mạng bộ giải mã-mã hóa LSTM sử dụng một LSTM để mã hóa các khung đầu vào và một LSTM khác để dự đoán nhiều khung hình phía trước. Mô hình đã được khái quát hóa trong [23] bằng cách thay thế LSTM được kết nối đầy đủ bằng ConvLSTM để nắm bắt các mối tương quan không gian tốt hơn. Sau đó, Finn et al. [5] và De Brabandere et al. [3] mở rộng mô hình trong [23] bằng cách làm cho mạng dự đoán sự biến đổi của khung đầu vào thay vì dự đoán trực tiếp các pixel thô. Ruben và cộng sự. [26] đề xuất sử dụng cả RNN thu được chuyển động và CNN để nắm bắt nội dung để tạo dự đoán. Cùng với các mô hình dựa trên RNN, các mô hình dựa trên CNN 2D và 3D đã được đề xuất trong [20] và [27]. Mathieu và cộng sự. [20] đã xử lý chuỗi khung là nhiều kênh và áp dụng CNN 2D để tạo dự đoán trong khi [27] coi chúng là độ sâu và áp dụng CNN 3D. Cả hai bài báo đều cho thấy Mạng Đối thủ Sáng tạo (GAN) [6] rất hữu ích để tạo dự đoán sắc nét.

### *Kết nối định kỳ có cấu trúc cho mô hình không gian vũ trụ*

Từ góc độ cấp cao hơn, dự báo lượng mưa (Precipitation Nowcasting) và dự đoán video về bản chất là các vấn đề dự báo trình tự không gian trong đó cả hai đầu vào và đầu ra là các chuỗi không gian [23]. Gần đây, có một xu hướng thay thế cấu trúc được kết nối đầy đủ trong các kết nối định kỳ của RNN bằng các cấu trúc liên kết khác để tăng cường khả năng Mạng lưới để mô hình hóa mối quan hệ cộng đồng. Khác với ConvLSTM thay thế kết nối đầy đủ bằng tích chập và được thiết kế cho các video dày đặc, SocialLSTM [1] và Architectural-RNN (S-RNN) [11] đã được đề xuất chia sẻ một khái niệm tương tự. SocialLSTM định nghĩa cấu trúc liên kết dựa trên khoảng cách giữa những người khác nhau và được thiết kế để dự đoán quỹ đạo của con người trong khi S-RNN xác định cấu trúc dựa trên biểu đồ spatiotemporal đã cho. Tất cả các mô hình này khác với TrajGRU của chúng tôi ở chỗ mô

hình của chúng tôi chủ động tìm hiểu cấu trúc kết nối định kỳ. Liang et al. [17] đã đề xuất LSTM phát triển cấu trúc, cũng có khả năng tìm hiểu cấu trúc kết nối của RNNs. Tuy nhiên, mô hình của chúng được thiết kế cho nhiệm vụ phân tích đối tượng ngữ nghĩa và học cách tự động hợp nhất các nút biểu đồ. Do đó, nó khác với TrajGRU nhằm mục đích tìm hiểu cấu trúc tương quan cục bộ cho dữ liệu không gian vũ trụ.

### III. Model

*Trong phần này, chúng tôi trình bày mô hình mới để phát hiện mưa. Trước tiên chúng tôi giới thiệu cấu trúc dự báo mã hóa chung được sử dụng trong bài viết này. Sau đó, chúng tôi xem xét mô hình ConvGRU và trình bày mô hình TrajGRU mới.*

#### 1. Encoding-forecasting Structure (Cấu trúc dự báo mã hóa)

Chúng tôi áp dụng một công thức tương tự của vấn đề precipitation nowcasting như trong [23]. Giả sử rằng các bản đồ tiếng vang radar tạo

thành một chuỗi không gian  $\langle \mathcal{I}_1, \mathcal{I}_2, \dots \rangle$ . Tại một mốc thời gian nhất định, mô hình của chúng tôi tạo ra các dự đoán bước K có khả năng nhất

$\hat{\mathcal{I}}_{t+1}, \hat{\mathcal{I}}_{t+2}, \dots, \hat{\mathcal{I}}_{t+K}$ , dựa trên các quan sát J trước đó bao gồm cả quan sát hiện tại:  $\mathcal{I}_{t-J+1}, \mathcal{I}_{t-J+2}, \dots, \mathcal{I}_t$ . Mạng dự báo mã hóa của chúng tôi trước tiên mã hóa các quan sát thành n lớp trạng thái RNN:

$$\mathcal{H}_t^1, \mathcal{H}_t^2, \dots, \mathcal{H}_t^n = h(\mathcal{I}_{t-J+1}, \mathcal{I}_{t-J+2}, \dots, \mathcal{I}_t)$$

và sau đó sử dụng n lớp RNN khác để tạo dự đoán dựa trên các trạng thái được mã hóa này:  $\hat{\mathcal{I}}_{t+1}, \hat{\mathcal{I}}_{t+2}, \dots, \hat{\mathcal{I}}_{t+K} = g(\mathcal{H}_t^1, \mathcal{H}_t^2, \dots, \mathcal{H}_t^n)$ .

Hình 1 minh họa cấu trúc dự báo mã hóa của chúng tôi cho  $n = 3, J = 2, K = 2$ . Chúng tôi chèn các lớp downsampling và upampling giữa các RNN, được thực hiện bởi convolution and deconvolution với stride. Lý do để đảo ngược thứ tự của mạng dự báo là các states cấp cao, đã chiếm được đại diện toàn không gian, có thể hướng dẫn cập nhật các states cấp thấp. Hơn nữa, các states cấp thấp có thể ảnh hưởng đến dự đoán. Cấu trúc này hợp lý hơn cấu trúc [23] không đảo ngược liên kết của mạng dự báo trước đây vì chúng tôi có thể tự do thêm các lớp RNN trên đầu và không bắt buộc skip-connection để tổng hợp thông tin cấp thấp. Người ta có thể chọn bất kỳ loại RNN nào như ConvGRU hoặc TrajGRU mới được đề xuất của chúng tôi trong cấu trúc dự báo mã hóa chung này miễn là trạng thái của chúng tương ứng với các tenxơ.



## 2. Convolutional GRU (mô hình ConvGRU)

Các công thức chính của ConvGRU được sử dụng trong bài viết này được đưa ra như sau:

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

Các thuật ngữ bias được bỏ qua để đơn giản ký hiệu. '\*' là hoạt động tích chập và 'o' là sản phẩm Hadamard. Ở đây,  $\mathcal{H}_t, \mathcal{R}_t, \mathcal{Z}_t, \mathcal{H}'_t \in \mathbb{R}^{C_h \times H \times W}$  lần lượt là trạng thái bộ nhớ, cổng đặt lại, cổng cập nhật và thông tin mới.

$\mathcal{X}_t \in \mathbb{R}^{C_i \times H \times W}$  là đầu vào và  $f$  là kích hoạt, được chọn là leaky ReLU với độ dốc âm bằng 0,2 [18] trên giấy.  $H, W$  là chiều cao và chiều rộng của trạng thái và các tenxơ đầu vào và  $C_h, C_i$  lần lượt là các kích thước kênh của trạng thái và các tenxơ đầu vào. Mỗi khi có đầu vào mới, cổng đặt lại sẽ kiểm soát xem có xóa trạng thái trước đó không và cổng cập nhật sẽ kiểm soát lượng thông tin mới sẽ được ghi vào trạng thái.

## 3. Trajectory GRU (mô hình TrajGRU)

Khi được sử dụng để nắm bắt các mối tương quan về không gian, sự thiếu hụt của ConvGRU và các ConvRNN khác là cấu trúc kết nối và trọng lượng được cố định cho tất cả các vị trí. Hoạt động tích chập về cơ bản áp dụng bộ lọc bất biến vị trí cho đầu vào. Nếu các đầu vào đều bằng 0 và các cổng đặt lại đều là một, chúng ta có thể viết lại quy trình tính toán của thông tin mới tại một vị trí cụ thể  $(i, j)$  tại dấu thời gian  $t$ , tức là,  $\mathcal{H}'_{t, :, i, j}$ , như sau :

$$\mathcal{H}'_{t, :, i, j} = f(\mathbf{W}_{hh} \text{concat}(\langle \mathcal{H}_{t-1, :, p, q} \mid (p, q) \in \mathcal{N}_{i, j}^h \rangle)) = f\left(\sum_{l=1}^{|\mathcal{N}_{i, j}^h|} \mathbf{W}_{hh}^l \mathcal{H}_{t-1, :, p_l, i, j, q_l, i, j}\right). \quad (2)$$

Ở đây,  $\mathcal{N}_{i, j}^h$  là vùng lân cận được đặt hàng tại vị trí  $(i, j)$  được xác định bởi các siêu đường kính của tích chập trạng thái như kích thước hạt nhân, độ giãn nở và phần đệm [30].  $(p_l, i, j, q_l, i, j)$  là vị trí lân cận thứ  $l$  của vị trí  $(i, j)$ . Hàm concat (·) nối các vectơ bên trong trong tập hợp và  $\mathbf{W}_{hh}$  là biểu diễn ma trận của các trọng số tích chập trạng thái.

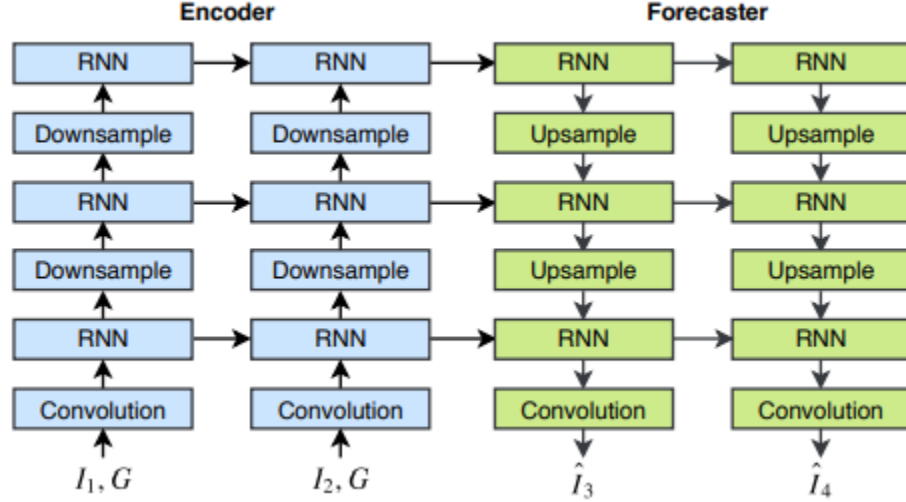
Khi siêu tham số của tích chập được cố định, vùng lân cận đặt

$\mathcal{N}_{i, j}^h$  giữ nguyên cho tất cả các vị trí. Tuy nhiên, hầu hết các mẫu chuyển động có các bộ lân cận khác nhau cho các vị trí khác nhau. Ví dụ, xoay và chia tỷ lệ tạo ra các trường dòng chảy với các góc khác

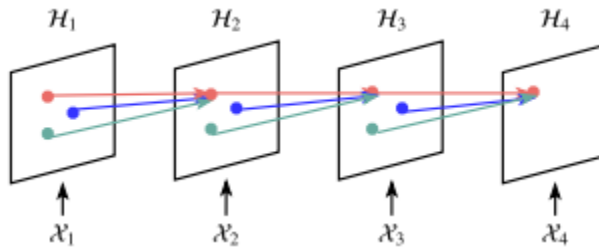
nhau chỉ đến các hướng khác nhau. Do đó, sẽ hợp lý hơn khi có cấu trúc kết nối biến thể vị trí là

$$\mathcal{H}'_{t,:,i,j} = f\left(\sum_{l=1}^L \mathbf{W}_{hh}^l \mathcal{H}_{t-1,:,p_{l,i,j}(\theta),q_{l,i,j}(\theta)}\right), \quad (3)$$

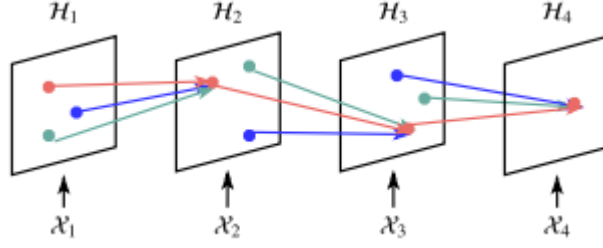
trong đó  $L$  là tổng số liên kết cục bộ,  $(p_{l,i,j}(\theta), q_{l,i,j}(\theta))$  là vùng lân cận thứ  $l$  được tham số hóa bởi  $\theta$ .



Hình 1: Ví dụ về cấu trúc dự báo mã hóa được sử dụng trong bài báo. Trong hình, chúng tôi sử dụng ba RNN để dự đoán hai khung tương lai  $\hat{I}_3, \hat{I}_4$  cho hai khung đầu vào  $I_1, I_2$ . Các tọa độ không gian  $G$  được nối với khung đầu vào để đảm bảo mạng biết các quan sát từ các vị trí khác nhau. Các RNN có thể là ConvGRU hoặc TrajGRU. Số không được cung cấp làm đầu vào cho RNN nếu thiếu liên kết đầu vào.



(a) Đối với RNN tích chập, các kết nối định kỳ được cố định theo thời gian.



4. (b) Đối với RNN quỹ đạo, các kết nối định kỳ được xác định động.

Hình 2: So sánh cấu trúc kết nối của RNN tích chập và RNN quỹ đạo. Liên kết với cùng màu chia sẻ cùng trọng lượng chuyển tiếp.

Dựa trên quan sát này, chúng tôi đề xuất TrajGRU, sử dụng đầu vào hiện tại và trạng thái trước đó để tạo vùng lân cận cục bộ được đặt cho từng vị trí tại mỗi mốc thời gian. Do các chỉ số vị trí là rời rạc và không khác biệt, chúng tôi sử dụng một tập hợp các luồng quang liên tục để thể hiện các chỉ số này. Các công thức chính của TrajGRU được đưa ra như sau:

$$\begin{aligned}
 \mathcal{U}_t, \mathcal{V}_t &= \gamma(\mathcal{X}_t, \mathcal{H}_{t-1}), \\
 \mathcal{Z}_t &= \sigma(\mathcal{W}_{xz} * \mathcal{X}_t + \sum_{l=1}^L \mathcal{W}_{hz}^l * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})), \\
 \mathcal{R}_t &= \sigma(\mathcal{W}_{xr} * \mathcal{X}_t + \sum_{l=1}^L \mathcal{W}_{hr}^l * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})), \\
 \mathcal{H}'_t &= f(\mathcal{W}_{xh} * \mathcal{X}_t + \mathcal{R}_t \circ (\sum_{l=1}^L \mathcal{W}_{hh}^l * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l}))), \\
 \mathcal{H}_t &= (1 - \mathcal{Z}_t) \circ \mathcal{H}'_t + \mathcal{Z}_t \circ \mathcal{H}_{t-1}.
 \end{aligned} \tag{4}$$

Ở đây,  $L$  là tổng số liên kết được phép.  $\mathcal{U}_t, \mathcal{V}_t \in \mathbb{R}^{L \times H \times W}$  là các trường lưu trữ cấu trúc kết nối cục bộ được tạo bởi mạng tạo cấu trúc  $y$ .  $\mathcal{W}_{hz}^l, \mathcal{W}_{hr}^l, \mathcal{W}_{hh}^l$  là các trọng số để chiếu các kênh, được thực hiện bằng các convolutions  $1 \times 1$ .

Hàm  $\text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})$  chọn các vị trí được chỉ ra bởi  $\mathcal{U}_{t,l}, \mathcal{V}_{t,l}$  từ  $\mathcal{H}_{t-1}$  thông qua hạt nhân lấy mẫu song tuyến [10, 9].

$$\mathcal{M}_{c,i,j} = \sum_{m=1}^H \sum_{n=1}^W \mathcal{I}_{c,m,n} \max(0, 1 - |i + \mathbf{V}_{i,j} - m|) \max(0, 1 - |j + \mathbf{U}_{i,j} - n|). \tag{5}$$

Nếu chúng ta biểu thị  $\mathcal{M} = \text{warp}(\mathcal{I}, \mathbf{U}, \mathbf{V})$  trong đó  $\mathcal{M}, \mathcal{I} \in \mathbb{R}^{C \times H \times W}$  và  $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{H \times W}$ , chúng ta có:

Ưu điểm của cấu trúc như vậy là chúng ta có thể tìm hiểu cấu trúc liên kết bằng cách tìm hiểu các tham số của mạng con  $y$ . Trong các thử nghiệm của chúng tôi,  $y$  có sự kết hợp của  $\mathcal{X}_t$  và  $\mathcal{H}_{t-1}$  làm đầu vào và được cố định là mạng nơ ron tích chập một lớp với kích thước kernel 5x5 và 32 feature maps. Do đó, chỉ có một số lượng nhỏ các tham số và thêm gần như không mất chi phí cho tính toán tổng thể. So với một ConvGRU có tích chập trạng thái  $K \times K$ , TrajGRU có thể tìm hiểu cấu trúc kết nối hiệu quả hơn với  $L < K^2$ . Đối với ConvGRU và TrajGRU, số lượng tham số mô hình bị chi phối bởi kích thước của các trọng số trạng thái, là  $O(L \times C_h^2)$  đối với TrajGRU và  $O(K^2 \times C_h^2)$  đối với ConvGRU. Nếu  $L$  được chọn nhỏ hơn  $K^2$ , số lượng tham số của TrajGRU cũng có thể nhỏ hơn ConvGRU và mô hình TrajGRU có thể sử dụng các tham số hiệu quả hơn. Minh họa cấu trúc kết nối định kỳ của ConvGRU và TrajGRU được đưa ra trong Hình 2. Gần đây, Jeon & Kim [12] đã sử dụng các ý tưởng tương tự để mở rộng các hoạt động tích chập trong CNN. Tuy nhiên, Đơn vị tích hợp hoạt động (ACU) được đề xuất của họ tập trung vào các hình ảnh trong đó nhu cầu về bộ lọc biến thể vị trí bị hạn chế. TrajGRU của chúng tôi tập trung vào các video trong đó các bộ lọc biến thể vị trí rất quan trọng để xử lý các mẫu chuyển động như xoay. Ngoài ra, chúng tôi đang sửa đổi cấu trúc của kết nối định kỳ và đã kiểm tra số lượng liên kết khác nhau trong khi [12] sửa số liên kết thành 9.

*Bảng 1:* So sánh TrajGRU và các mô hình cơ sở trong bộ dữ liệu MoveMNIST ++. 'Conv-K $\alpha$ -D $\beta$ ' dùng để chỉ ConvGRU với kích thước hạt nhân  $\alpha \times \alpha$  và độ giãn  $\beta \times \beta$ . 'Traj-L $\lambda$ ' dùng để chỉ TrajGRU với  $\lambda$

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

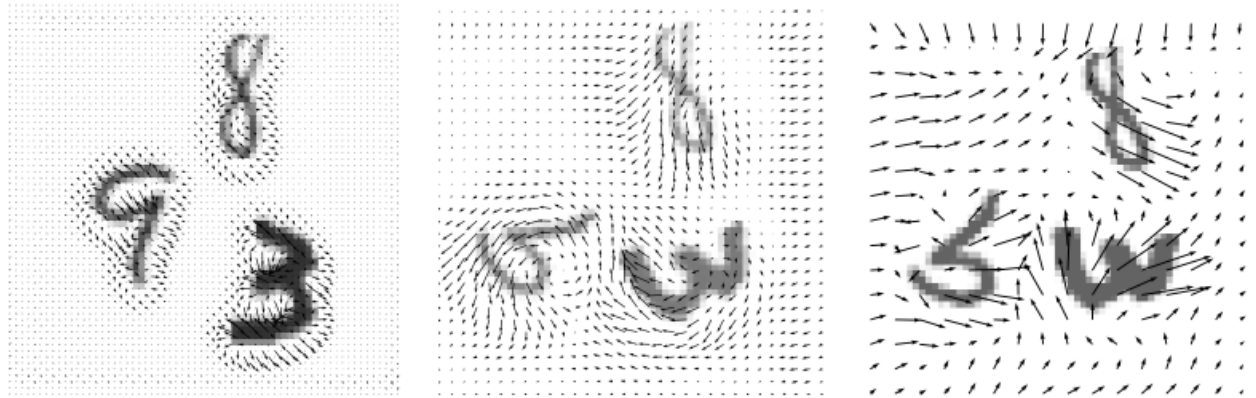
liên kết. Chúng tôi thay thế lớp đầu ra của mô hình ConvGRU-K5-D1 để lấy DFN.

## IV. Experiments on MovingMNIST++

Trước khi đánh giá mô hình của chúng tôi về precipitation nowcasting khó khăn hơn, trước tiên chúng tôi so sánh TrajGRU với ConvGRU, DFN và 2D / 3D trên bộ dữ liệu dự đoán video tổng hợp để chứng minh tính hiệu quả của nó.

Bộ dữ liệu MoveMNIST trước đây [24, 23] chỉ di chuyển các chữ số với tốc độ không đổi và không phù hợp để đánh giá khả năng của các mô hình khác nhau trong việc chụp các mẫu chuyển động phức tạp hơn. Do đó, chúng tôi thiết kế bộ dữ liệu MoveMNIST ++ bằng cách mở rộng MoveMNIST để cho phép xoay ngẫu nhiên, thay đổi tỷ lệ và thay đổi chiều sáng. Mỗi khung có kích thước  $64 \times 64$  và chứa ba chữ số chuyển động. Chúng tôi sử dụng 10 khung làm đầu vào để dự đoán 10 khung tiếp theo. Vì các khung có thay đổi chiều sáng, chúng tôi sử dụng MSE thay vì entropy chéo để đào tạo và đánh giá. Chúng tôi đào tạo tất cả các mô hình bằng cách sử dụng trình tối ưu hóa Adam [14] với tỷ lệ học bằng  $10^{-4}$  và động lượng bằng 0,5. Đối với các mô hình RNN, chúng tôi sử dụng cấu trúc dự báo mã hóa được giới thiệu trước đây với ba lớp RNN. Tất cả các RNN đều là ConvGRU hoặc TrajGRU và tất cả đều sử dụng cùng một bộ siêu đường kính. Đối với TrajGRU, chúng tôi khởi tạo trọng số của lớp đầu ra của mạng tạo cấu trúc thành không. Các bước của các lớp ghép xuống và lấy mẫu ở giữa được chọn là 2. Số lượng bộ lọc cho ba RNN lần lượt là 64, 96, 96. Đối với mô hình DFN, chúng tôi thay thế lớp đầu ra của ConvGRU bằng bộ lọc cục bộ  $11 \times 11$  và chuyển đổi khung trước đó để có được dự đoán. Đối với các mô hình RNN, chúng tôi đào tạo cho họ 200.000 lần lặp với ngưỡng cắt định mức bằng 1 và kích thước lô bằng 4. Đối với các mô hình CNN, chúng tôi đào tạo cho họ 100.000 lần lặp với ngưỡng cắt định mức bằng 50 và kích thước lô bằng 32. Cấu hình thử nghiệm chi tiết của các mô hình cho thử nghiệm MoveMNIST ++ có thể được tìm thấy trong phần phụ lục. Chúng tôi cũng đã cố gắng sử dụng GAN có điều kiện cho các mô hình 2D và 3D nhưng không đạt được kết quả hợp lý.

Bảng 1 đưa ra kết quả của các mô hình khác nhau trên cùng một bộ thử nghiệm có chứa 10.000 chuỗi. Chúng tôi đào tạo tất cả các mô hình bằng cách sử dụng ba hạt giống khác nhau để báo cáo độ lệch chuẩn. Chúng ta có thể thấy rằng TrajGRU chỉ có 5 liên kết vượt trội hơn ConvGRU với kích thước state-to-state kernel  $3 \times 3$  và độ giãn nở  $2 \times 2$  (9 liên kết). Ngoài ra, hiệu suất của TrajGRU được cải thiện khi số lượng liên kết tăng lên. TrajGRU với  $L = 13$  vượt trội so với ConvGRU với state-to-state kernel  $7 \times 7$  và vẫn có ít tham số hơn. Một quan sát khác từ bảng là DFN không hoạt động tốt trong bộ dữ liệu tổng hợp này. Điều này là do DFN sử dụng softmax để tăng cường độ thưa thớt của các bộ lọc cục bộ đã học, không thể mô hình hóa thay đổi chiều sáng vì giá trị tối đa luôn nhỏ hơn sau khi kết hợp với hạt nhân dương có trọng số lên tới 1. Đối với DFN, khi giá trị pixel nhận được nhỏ hơn, không thể tăng trở lại. Hình 3 trực quan hóa các cấu trúc đã học của TrajGRU. Chúng ta có thể thấy rằng mạng đã học được các mẫu liên kết cục bộ hợp lý.



*Hình 3:* Các liên kết được chọn của TrajGRU-L13 ở các khung và lớp khác nhau. Chúng tôi chọn một trong 13 liên kết và vẽ một mũi tên bắt đầu từ mỗi pixel đến pixel được liên kết bởi liên kết. Từ trái sang phải chúng ta hiển thị cấu trúc đã học ở lớp thứ nhất, thứ hai và thứ ba của bộ mã hóa. Các liên kết hiển thị ở đây đã học hành vi cho các phép quay. Chúng tôi lấy mẫu phụ các liên kết được hiển thị cho lớp đầu tiên để dễ đọc hơn. Chúng tôi bao gồm hình ảnh động cho tất cả các lớp và các liên kết trong tài liệu bổ sung.

## V. Benchmark for Precipitation Nowcasting

### 1. HKO-7 Dataset (tập dữ liệu HKO-7)

Tập dữ liệu HKO-7 sử dụng điểm chuẩn gồm dữ liệu radar echo từ năm 2009 đến năm 2015 thu thập bởi HKO. Các hình ảnh phản xạ của radar CAPPI có độ phân giải  $480 \times 480$ , được chụp từ độ cao 2km và bao phủ một khu vực  $512 \times 512$  diện tích trung tâm ở Hồng Kông. Dữ liệu được chụp lại cứ



mỗi 6 phút và do đó có 240 hình ảnh mỗi ngày. Các hệ số phản xạ radar logarit ban đầu được chuyển đổi tuyến tính thành các giá trị pixel thông qua  $\text{pixel} = 255 \times (\text{dBZ} + 10/70) + 0,5$  và được cắt ở đoạn giữa 0 và 255. Các hình ảnh phản xạ của radar ban đầu được tạo ra bởi radar thời tiết Doppler bị nhiễu do các yếu tố như sự lộn xộn trên mặt đất, sự lộn xộn trên biển, sự lan truyền dị thường và nhiễu điện từ. Để giảm bớt tác động của nhiễu trong huấn luyện và đánh giá, chúng tôi lọc các pixel nhiễu trong tập dữ liệu và tạo mặt nạ nhiễu theo quy trình hai giai đoạn được mô tả trong phụ lục.

Khi tình trạng mưa thừa thớt, chúng tôi chọn những ngày mưa dựa trên thông tin ở rain barrel để tạo thành tập dữ liệu cuối cùng, với 812 ngày để huấn luyện, 50 ngày để đánh giá và 131 ngày để kiểm tra. Cách xử lý hiện tại gần với kịch bản thực tế vì chúng tôi có thể huấn luyện một mô hình bổ sung để phân loại xem trời có mưa vào ngày hôm sau hay không và áp dụng mô hình dự báo lượng mưa của chúng tôi nếu mô hình cấp độ thô (coarser-level) này dự đoán rằng trời sẽ mưa. Các giá trị phản xạ radar được chuyển đổi thành giá trị cường độ mưa (mm / h) bằng mối quan hệ Z-R:  $\text{dBZ} = 10 \log a + 10b \log R$  trong đó R là mức độ tốc độ mưa (rain-rate level),  $a = 58,53$  và  $b = 1,56$ . Thống kê tổng thể và phân phối lượng mưa trung bình hàng tháng của bộ dữ liệu HKO-7 được nêu trong phần phụ lục.

## 2. Evaluation Methodology (đánh giá phương pháp)

Khi ánh xạ radar echo đến cùng một luồng, các thuật toán dự báo tức thời có thể áp dụng việc học trực tuyến để thích ứng với các mẫu mới xuất hiện. Chúng tôi đề xuất hai cài đặt trong giao thức đánh giá của chúng tôi: (1) cài đặt ngoại tuyến trong đó thuật toán luôn dùng 5 hình ảnh làm input và dự đoán 20 hình ảnh phía trước và (2) cài đặt trực tuyến trong đó thuật toán dùng các đoạn (segment) liên tục có độ dài là 5 và dự đoán 20 hình ảnh phía trước cho mỗi phân đoạn mới nhận được. Giao thức đánh giá được mô tả có hệ thống hơn trong phần phụ lục. Môi trường thử nghiệm đảm bảo rằng cùng một tập hợp được kiểm tra trong cả cài đặt ngoại tuyến và trực tuyến để so sánh công bằng.

Đối với cả hai cài đặt, chúng tôi đánh giá skill scores cho nhiều ngưỡng tương ứng với các mức mưa khác nhau để đưa ra đánh giá toàn diện về hiệu suất hiện tại của thuật toán. Bảng 2 cho thấy sự phân bố lượng mưa khác nhau trong tập dữ liệu của chúng tôi. Chúng tôi chọn sử dụng các ngưỡng 0,5, 2, 5, 10, 30 để tính skill scores cho CSI và Heidke (HSS). Để tính skill scores ở ngưỡng  $\tau$  cụ thể, là 0,5, 2, 5, 10 hoặc 30, đầu tiên chúng tôi chuyển đổi các giá trị pixel trong dự đoán và ground truth thành 0/ 1 bởi ngưỡng là  $\tau$ . Sau đó, chúng tôi tính toán TP (prediction=1, truth=1), FN (prediction=0, truth=1), FP (prediction=1, truth=0), và TN (prediction=0, truth=0). Điểm CSI được tính bằng  $\text{TP} / (\text{TP} + \text{FN} + \text{FP})$  và điểm HSS được

tính bằng  $(TP \times TN - FN \times FP) / (TP + FN)(FN + TN) + (TP + FP)(FP + TN)$ . Trong quá trình tính toán, các điểm bị che (masked point) bị bỏ qua.

**Table 2: Rain rate statistics in the HKO-7 benchmark.**

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

Như trong Bảng 2, tần số của các mức mưa khác nhau rất mất cân bằng. Chúng tôi đề xuất sử dụng trọng số trong hàm mất mát để giúp giải quyết vấn đề này. Cụ thể, chúng tôi chỉ định một trọng số  $w(x)$  đến từng pixel theo cường độ mưa của nó  $x$ :

$$w(x) = \begin{cases} 1, & x < 2 \\ 2, & 2 \leq x < 5 \\ 5, & 5 \leq x < 10 \\ 10, & 10 \leq x < 30 \\ 30, & x \geq 30 \end{cases}$$

Ngoài ra, pixel bị che có trọng số là 0.

Kết quả điểm B-MSE và B-MAE được tính là

$$B - MSE = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{480} \sum_{j=1}^{480} w_{n,i,j} (x_{n,i,j} - \hat{x}_{n,i,j})^2$$

$$B - MAE = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{480} \sum_{j=1}^{480} w_{n,i,j} |x_{n,i,j} - \hat{x}_{n,i,j}|$$

Trong đó  $N$  là tổng số hình ảnh và  $w_{n,i,j}$  là trọng số tương ứng với pixel  $(i, j)$  trong hình ảnh thứ  $n$ . Đối với các độ đo MSE và MAE thông thường, chúng tôi chỉ cần đặt tất cả các trọng số thành 1 ngoại trừ các điểm bị che.

### 3. Evaluated Algorithms (đánh giá thuật toán)

Chúng tôi đã đánh giá bảy thuật toán dự báo, bao gồm mô hình đơn giản nhất luôn dự đoán hình ảnh cuối cùng, hai phương pháp dựa trên dòng



quang (optical flow) (ROVER và biến thể phi tuyến của nó) và bốn phương pháp học sâu (TrajGRU, ConvGRU, 2D CNN và 3D CNN). Cụ thể, chúng tôi đã đánh giá hiệu suất của các mô hình học sâu trong cài đặt trực tuyến bằng cách tinh chỉnh các thuật toán bằng AdaGrad [4] với learning rate bằng  $10^{-4}$ . Chúng tôi tối ưu hóa tổng số B-MSE và B-MAE trong quá trình huấn luyện ngoại tuyến và tinh chỉnh trực tuyến. Trong quá trình đào tạo huấn luyện, tất cả các mô hình đều được tối ưu hóa bởi thuật toán tối ưu adam với learning rate bằng  $10^{-4}$  và momentum bằng 0,5. và chúng tôi huấn luyện các mô hình này với việc early-stopping trên tổng B-MSE và B-MAE. Đối với các mô hình RNN, batch size được đặt thành 4. Đối với các mô hình CNN, batch size được đặt thành 8. Đối với mô hình TrajGRU và ConvGRU, chúng tôi sử dụng 3-layer encoding-forecasting structure với số với số lượng bộ lọc cho RNN được đặt thành 64, 192, 192. Chúng tôi sử dụng kernel size bằng  $5 \times 5$ ,  $5 \times 5$ ,  $3 \times 3$  cho các mô hình ConvGRU trong khi số lượng liên kết được đặt thành 13, 13, 9 cho mô hình TrajGRU. Chúng tôi cũng huấn luyện mô hình ConvGRU với mất mát MSE và MAE ban đầu, được đặt tên là “ConvGRU-nobal”, để đánh giá sự cải thiện bằng cách đào tạo với mất mát B-MSE và B-MAE. Các cấu hình mô hình khác bao gồm ROVER, ROVER-phi tuyến và deep model được bao gồm trong phụ lục.

#### 4. Evaluation Results (đánh giá kết quả)

Các kết quả đánh giá tổng thể được tóm tắt trong Bảng 3. Để phân tích độ tin cậy của kết quả, chúng tôi huấn luyện các mô hình 2D CNN, 3D CNN, ConvGRU và TrajGRU bằng ba random seed khác nhau và báo cáo độ lệch chuẩn trong Bảng 4. Chúng tôi thấy rằng huấn luyện với balanced loss functions là điều cần thiết để hiệu suất dự đoán tốt hơn với lượng mưa lớn. Mô hình ConvGRU được huấn luyện mà không có balanced loss, đại diện tốt nhất cho các mô hình, có điểm dự đoán kém hơn so với các phương pháp dựa trên dòng quang ở ngưỡng 10 mm / h và 30 mm / h. Ngoài ra, chúng tôi thấy rằng tất cả các mô hình deep learning được huấn luyện với balanced loss đều vượt trội so với các mô hình dựa trên dòng quang. Trong số các mô hình deep learning, TrajGRU thực hiện CNN 3D tốt nhất và vượt trội so với 2D CNN, điều này cho thấy cấu trúc mạng phù hợp là rất quan trọng để đạt được hiệu suất tốt. Sự cải thiện của TrajGRU so với các mô hình khác có ý nghĩa thống kê vì sự khác biệt về B-MSE và B-MAE lớn hơn ba lần độ lệch chuẩn của chúng. Hơn nữa, hiệu suất với tinh chỉnh trực tuyến được kích hoạt luôn tốt hơn so với hiệu suất không tinh chỉnh trực tuyến, điều này xác minh tính hiệu quả của việc học trực tuyến ít nhất là cho nhiệm vụ này.

Bảng 3: Kết quả điểm chuẩn HKO-7. Chúng tôi đánh dấu kết quả tốt nhất trong một cài đặt cụ thể bằng cách in đậm nó và kết quả tốt thứ hai bằng cách gạch chân. Mỗi ô chứa điểm trung bình của 20 hình ảnh được dự đoán. Trong cài đặt trực tuyến, tất cả các thuật toán đã sử dụng chiến lược học trực tuyến được mô tả trong bài báo. ↑ có nghĩa là điểm càng cao càng

tốt trong khi  $\downarrow$  có nghĩa là điểm càng thấp càng tốt.  $r \geq \tau$  có nghĩa là skill score ở ngưỡng lượng mưa mm / h. Đối với các mô hình 2D CNN, 3D CNN, ConvGRU và TrajGRU, chúng tôi huấn luyện các mô hình với ba random seed khác nhau và báo cáo điểm trung bình.

Algorithms	CSI $\uparrow$					HSS $\uparrow$					B-MSE $\downarrow$ B-MAE $\downarrow$	
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$		
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	<u>0.5489</u>	<u>0.4731</u>	<u>0.3720</u>	<u>0.2789</u>	<u>0.1776</u>	<u>0.6701</u>	<u>0.6104</u>	<u>0.5163</u>	<u>0.4159</u>	<u>0.2893</u>	<u>5951</u>	<u>15000</u>
TrajGRU	<b>0.5528</b>	<b>0.4759</b>	<b>0.3751</b>	<b>0.2835</b>	<b>0.1856</b>	<b>0.6731</b>	<b>0.6126</b>	<b>0.5192</b>	<b>0.4207</b>	<b>0.2996</b>	<b>5816</b>	<b>14675</b>
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	<u>0.5511</u>	<u>0.4737</u>	<u>0.3742</u>	<u>0.2843</u>	<u>0.1837</u>	<u>0.6712</u>	<u>0.6105</u>	<u>0.5183</u>	<u>0.4226</u>	<u>0.2981</u>	<u>5724</u>	<u>14772</u>
TrajGRU	<b>0.5563</b>	<b>0.4798</b>	<b>0.3808</b>	<b>0.2914</b>	<b>0.1933</b>	<b>0.6760</b>	<b>0.6164</b>	<b>0.5253</b>	<b>0.4308</b>	<b>0.3111</b>	<b>5589</b>	<b>14465</b>

Bảng 4: Khoảng tin cậy của các deep model được chọn trong điểm chuẩn HKO-7. Chúng tôi đào tạo 2D CNN, 3D CNN, ConvGRU và TrajGRU sử dụng ba random seed khác nhau và báo cáo độ lệch chuẩn của điểm kiểm tra.

Algorithms	CSI					HSS					B-MSE	B-MAE
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$		
Offline Setting												
2D CNN	0.0032	0.0023	0.0015	0.0001	0.0025	0.0032	0.0025	0.0018	0.0003	0.0043	90	95
3D CNN	0.0043	0.0027	0.0016	0.0024	0.0024	0.0042	0.0028	0.0018	0.0031	0.0041	44	26
ConvGRU	0.0022	0.0018	0.0031	0.0008	0.0022	0.0022	0.0021	0.0040	0.0010	0.0038	52	81
TrajGRU	0.0020	0.0024	0.0025	0.0031	0.0031	0.0019	0.0024	0.0028	0.0039	0.0045	18	32
Online Setting												
2D CNN	0.0002	0.0005	0.0002	0.0002	0.0012	0.0002	0.0005	0.0002	0.0003	0.0019	12	12
3D CNN	0.0004	0.0003	0.0002	0.0003	0.0008	0.0004	0.0004	0.0003	0.0004	0.0001	23	27
ConvGRU	0.0006	0.0012	0.0017	0.0019	0.0024	0.0006	0.0012	0.0019	0.0023	0.0031	30	69
TrajGRU	0.0008	0.0004	0.0002	0.0002	0.0002	0.0007	0.0004	0.0002	0.0002	0.0003	10	20

Bảng 5: Hệ số của Kendall giữa các skill score. Giá trị tuyệt đối cao hơn cho thấy mối tương quan mạnh mẽ hơn. Các số có giá trị tuyệt đối lớn nhất được đánh dấu bằng cách in đậm.

Skill Scores	CSI					HSS				
	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$	$r \geq 0.5$	$r \geq 2$	$r \geq 5$	$r \geq 10$	$r \geq 30$
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	<b>-0.60</b>	-0.55	-0.24	-0.26
B-MSE	-0.70	-0.57	<b>-0.61</b>	<b>-0.86</b>	-0.84	-0.62	-0.55	<b>-0.61</b>	<b>-0.86</b>	-0.84
B-MAE	<b>-0.74</b>	<b>-0.59</b>	-0.58	-0.82	<b>-0.92</b>	<b>-0.67</b>	-0.57	-0.59	-0.83	<b>-0.92</b>

Dựa trên kết quả đánh giá, chúng tôi cũng tính toán các hệ số của Kendall [13] giữa MSE, MAE, B-MSE, B-MAE và CSI, HSS ở các ngưỡng khác nhau. Được thể hiện trong Bảng 5, B-MSE và B-MAE có mối tương quan mạnh mẽ hơn với CSI và HSS trong hầu hết các trường hợp.

## VI. Conclusion and Future Work

Trong bài báo này, chúng tôi đã cung cấp điểm chuẩn quy mô lớn đầu tiên cho dự báo lượng mưa và đã đề xuất một mô hình TrajGRU mới với khả năng học recurrent connection structure. Chúng tôi đã chỉ ra TrajGRU hiệu quả hơn trong việc nắm bắt các mối tương quan về không gian so với ConvGRU. Đối với công việc trong tương lai, chúng tôi có kế hoạch kiểm tra xem TrajGRU có giúp cải thiện các nhiệm vụ học tập không gian khác như visual object tracking và video segmentation. Chúng tôi cũng sẽ cố gắng xây dựng một hệ thống dự đoán hoạt động bằng thuật toán đề xuất.

By

(NHÓM 6)



© Copyright (Insert year of publishing) by (Insert author or publisher name) - All rights reserved.

It is not legal to reproduce, duplicate, or transmit any part of this document in either electronic means or printed format. Recording of this publication is strictly prohibited.