TaylorSwiftNet: Taylor Driven Temporal Modeling for Swift Future Frame Prediction

Saber Pourheydari[1,*], Emad Bahrami[1,*], Mohsen Fayyaz[1,4*], Gianpiero Francesca[2], Mehdi Noroozi[3], Jürgen Gall[1]

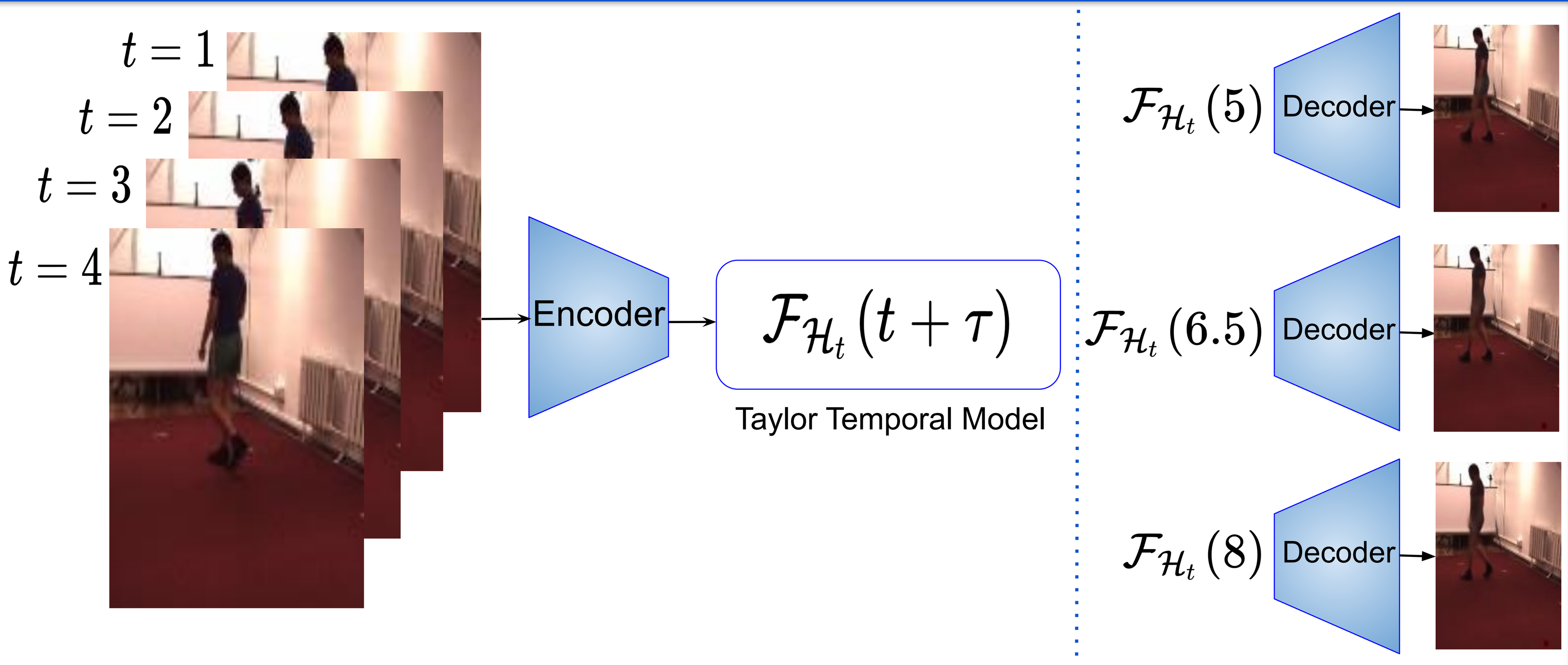
[1]University of Bonn, [2]Toyota Motor Europe, [3]Bosch Center for Artificial Intelligence, [4] Microsoft Research,
*Contributed equally to this work

UNIVERSITÄT BONN

BMVC 2022

Introduction

- Methods based on Partial Differential Equations (PDE) are capable of modeling in continuous time space
- However, the PDE-based methods discretize the PDEs for training and inference
- We propose an approach capable of learning temporally continuous representation that does not need any discretization
- To forecast future frames at $t+\tau$ the learnt function $\mathcal{F}_{\mathcal{H}_t}$ can be evaluated for any positive value τ in parallel

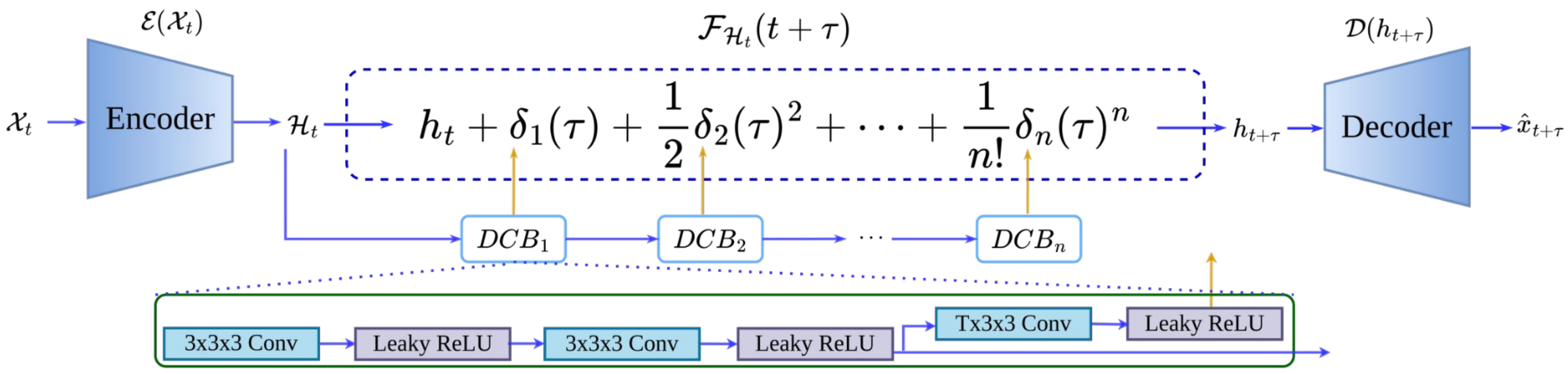


Temporal Dynamics Modeling Through Taylor Series

- The model consists of three parts that are learned end-to-end, Encoder ($\mathcal{E}$), Decoder ($\mathcal{D}$), and Temporal model ($\mathcal{F}_{\mathcal{H}_t}$)

$$\hat{x}_{t+\tau} = \mathcal{D}(h_{t+\tau})\ ;\ h_{t+\tau} = \mathcal{F}_{\mathcal{H}_t}(t+\tau)\ ;\ \mathcal{H}_t = \mathcal{E}(\mathcal{X}_t)$$

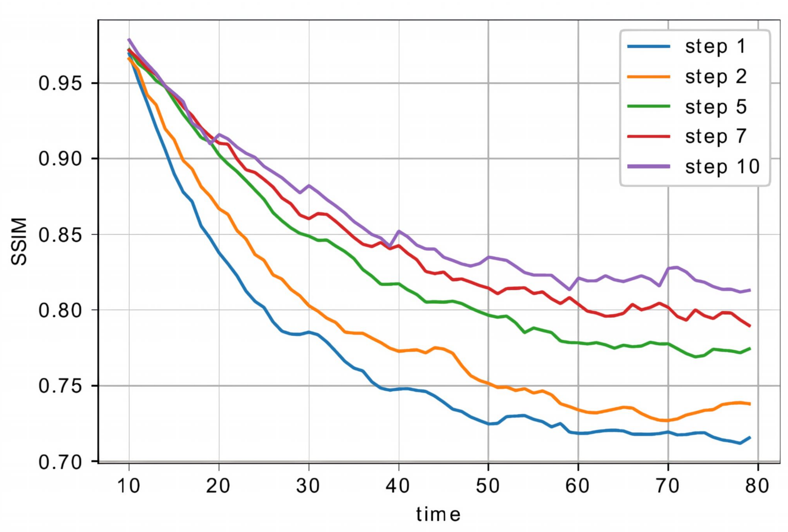
- The Temporal model, $\mathcal{F}_{\mathcal{H}_t}$, is approximated using Taylor series: $\mathcal{F}_{\mathcal{H}_t}(t+\tau) \simeq \sum_{n=0}^{\gamma} \frac{\mathcal{F}_{\mathcal{H}_t}^{(n)}(t)}{n!}\tau^n\ ;\ \mathcal{F}_{\mathcal{H}_t}^{(n)} = \frac{\partial^n \mathcal{F}_{\mathcal{H}_t}}{\partial \tau^n}$
- The derivatives $\delta_n = \mathcal{F}_{\mathcal{H}_t}^{(n)}$ are approximated with Delta Convolutional Block (DCB)



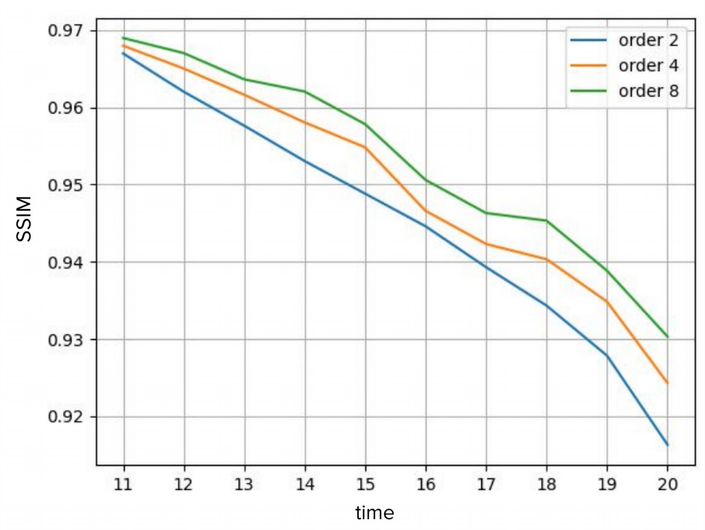
Quantitative Results

	Moving MNIST			Traffic BJ			Sea Surface Temperature			Human 3.6M		
Method	MSE ↓	MAE ↓	SSIM ↑	MSE ×100	MAE	SSIM	MSE ×10	MAE	SSIM	MSE /10	MAE /100	SSIM
Advection-diffusion	-	-	-	-	-	-	34.1	54.1	0.966	-	-	-
DDPAE	38.9	90.7	0.922	-	-	-	-	-	-	-	-	-
ConvLSTM	103.3	182.9	0.707	48.5	17.7	0.978	45.6	63.1	0.949	50.4	18.9	0.776
PredRNN	56.8	126.1	0.867	46.4	17.1	0.971	41.9	62.1	0.955	48.4	18.9	0.781
Causal LSTM	46.5	106.8	0.898	44.8	16.9	0.977	39.1	62.3	0.929	45.8	17.2	0.851
MIM	44.2	101.1	0.910	42.9	16.6	0.971	42.1	60.8	0.955	42.9	17.8	0.790
E3D-LSTM	41.3	86.4	0.920	43.2	16.9	0.979	34.7	59.1	0.969	46.4	16.6	0.869
PhyDNet	24.4	70.3	0.947	41.9	16.2	0.982	31.9	53.3	0.972	36.9	16.2	0.901
SimVP	23.8	68.9	0.948	41.4	16.2	0.982	-	-	-	31.6	**15.1**	0.904
TaylorSwiftNet (ours)	**21.2**	**60.8**	**0.952**	**35.3**	**13.7**	**0.992**	**29.8**	**52.2**	**0.978**	**23.1**	15.8	**0.910**

Comparison to the state-of-the-art on four datasets

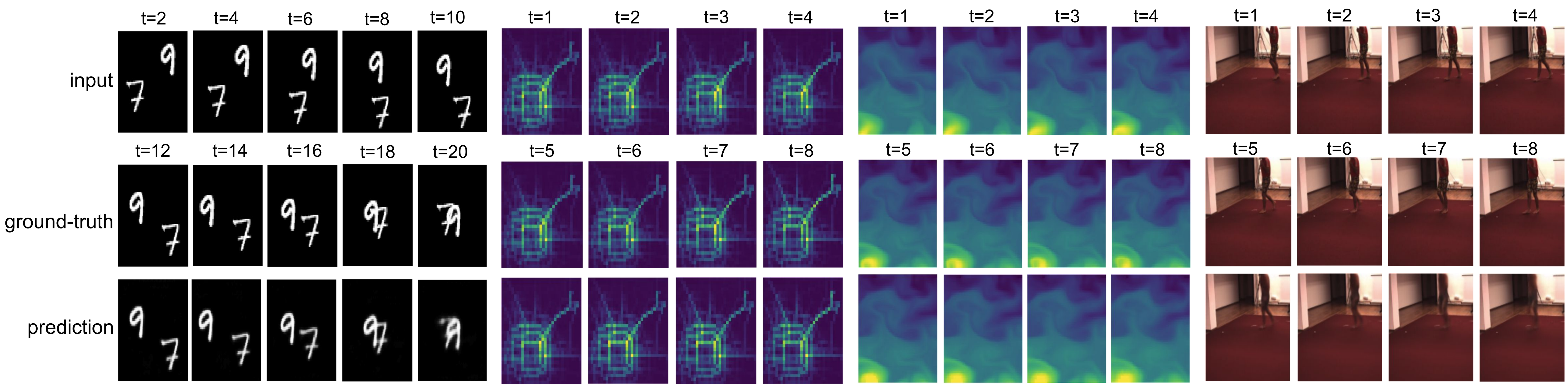


Comparing different orders of our temporal model



Comparison of future predictions for time horizons that are larger than the 10 frames used for training

Qualitative Results



(a) Moving MNIST (b) Traffic BJ (c) Sea Surface Temperature (d) Human 3.6M