

# Supplementary Materials of *Unifying Spike Perception and Prediction: A Compact Spike Representation Model using Multi-scale Correlation*

Anonymous Authors

## 1 QUANTITATIVE DERIVATIONS OF $t_\theta$

Neurons generate electrical signals in response to light stimuli and transmit these signals by exchanging  $Na^+$  and  $K^+$  ions inside and outside the cell. We use the Leaky Integrate and Fire (LIF) model to analyze the duration of luminance effects on photo-receptor cells. Variables used in the derivation process and their meanings are illustrated in Table. 1.

The neuro-dynamical differential equation of the LIF model [12] [4] is formulated as

$$\nu \frac{dV_t}{dt} = -(V_t - E) + \alpha RI. \quad (1)$$

We decouple the persistent light stimuli into a sequence of transient luminance stimulus, each of which lasts for an extremely short time and thus the luminance can be approximated as constant. For the luminance stimulation defined as

$$I = \begin{cases} I_0, & 0 < t < t_0, \\ 0, & else \end{cases}, \quad (2)$$

The entire process can be divided into two dependent procedures, including the charging and the discharging process. which are the charging during  $[0, t_0)$  and the discharging during  $[t_0, t_\theta]$ .

The charging process persists during  $[0, t_0)$ , in which the Eq. 1 can be specified as

$$\nu \frac{dV_t}{dt} = -(V_t - E) + \alpha RI_0. \quad (3)$$

By separating variables, Eq. 3 is redrafted as

$$\nu \frac{dV_t}{V_t - E - \alpha RI_0} = -dt. \quad (4)$$

Through integrating in temporal series and regarding 0 and  $t_0$  as upper and lower bounds on the integral, it can be deduced from Eq. 4 that

$$\begin{aligned} \ln(V_t - E - \alpha RI_0) \Big|_0^{t_0} &= -\frac{t}{\nu} \Big|_0^{t_0}, \\ \ln \frac{V_0 - E - \alpha RI_0}{-RI_0} &= -\frac{t_0}{\nu}. \end{aligned} \quad (5)$$

From the above, the membrane potential at the end of this process can be derived as

$$V_0 = E + \alpha RI_0 [1 - \exp(-\frac{t_0}{\nu})]. \quad (6)$$

Similarly, the discharging process persists during  $[t_0, t_\theta]$ , in which the Eq. 1 can be specified as

$$\nu \frac{dV_t}{dt} = -(V_t - E). \quad (7)$$

By separating variables, Eq. 7 is redrafted as

$$\nu \frac{dV_t}{V_t - E} = -dt. \quad (8)$$

Variables	Meaning of Variables
$\nu$	The time constant
$t$	The current moment
$t_0$	The duration of luminance stimulation
$t_\theta$	The maximum moment which stimulation can affect
$V_t$	The membrane potential at moment $t$
$V_0$	The membrane potential at moment $t_0$
$E$	The resting membrane potential
$V_\theta$	The threshold potential that can distinguish between electrical signals or disturbances
$R$	The membrane resistance
$I_t$	The luminance stimulation at moment $t$
$I_0$	The luminance intensity
$P_0$	The photons received during temporal range $t_0$ with luminance intensity $I_0$
$\alpha$	The photovoltaic conversion efficiency

**Table 1: Variables in derivation and corresponding meanings.**

Through integrating in temporal series and regarding  $t_0$  and  $t_\theta$  as upper and lower bounds on the integral, it can be deduced from Eq. 4 that

$$\begin{aligned} \ln(V_t - E) \Big|_{t_0}^{t_\theta} &= -\frac{t}{\nu} \Big|_{t_0}^{t_\theta}, \\ \ln \frac{V_\theta - E}{V_0 - E} &= -\frac{t_\theta - t_0}{\nu}. \end{aligned} \quad (9)$$

Bringing in the results in Eq. 6, the membrane potential at the end of this process can be derived as

$$\begin{aligned} t_\theta &= t_0 + \nu \ln \frac{V_0 - E}{V_\theta - E} \\ &= t_0 + \nu \ln \frac{\alpha RI_0 [1 - \exp(-t_0/\nu)]}{V_\theta - E}. \end{aligned} \quad (10)$$

Since  $t_0$  is infinitesimal [5], Eq. 10 is approximated as

$$\begin{aligned} \lim_{t_0 \rightarrow 0} t_\theta &= \lim_{t_0 \rightarrow 0} [t_0 + \nu \ln \frac{\alpha RI_0 [1 - \exp(-t_0/\nu)]}{V_\theta - E}] \\ &= \nu \ln \frac{\alpha RI_0 [1 - \lim_{t_0 \rightarrow 0} \exp(-t_0/\nu)]}{V_\theta - E} \\ &= \nu \ln \frac{\alpha RI_0 t_0}{(V_\theta - E)\nu} \\ &= \nu \ln \frac{\alpha RP_0}{(V_\theta - E)\nu}. \end{aligned} \quad (11)$$

## 2 PLCC BETWEEN $t_\theta$ AND $\nu$

We numerically calculate Eq. 11 by approximation and further estimate the PLCC between  $t_\theta$  and  $\nu$  to illustrate the strong correlation.

For neurons, the resting membrane potential is  $-70mV$  with a noisy disturbance of  $\pm 10mV$ , and the membrane resistance is  $10 M\Omega$ . For cone cells, with a surface area of approximately  $1.77\mu m^2$ , the number of photons received per unit time per unit area is approximately  $2.75 \times 10^{21}/s/m^2$  under natural light conditions. Each photon can generate about 5 electrons, and the charge of each electron is approximately  $1.6 \times 10^{-19}C$ . Above all, coefficients in Eq. 11 are:

$$\alpha = 8 \times 10^{-19}$$

$$R = 1 \times 10^7$$

$$P_0 = 4.9 \times 10^9$$

$$V_\theta = -6 \times 10^{-2}$$

$$E = -7 \times 10^{-2}$$

Thus Eq. 11 can be numerically redrafted as

$$t_\theta = \nu \ln(3.894/\nu). \quad (12)$$

According to Eq. 12, PLCC [3] between  $t_\theta$  and  $\nu$  is calculated as **0.995**.

## 3 VISUALIZATION OF SPIKE VISUAL REPRESENTATION

We visualize the predicted visual representations through approaches with and without the proposed ICP method, as shown in Fig. 1. Without ICP, local low-intensity features are mistaken for noise and diffuse over time (Fig. 1(c)), while the establishment of ICP leads to stable prediction results. In the meanwhile, ICP can reserve detailed high-frequency information and rebuild more texture content, resulting in less blurring and more defined edges. These all illustrate that ICP can exploit the stronger temporal continuity of features at different scales, resulting in more significant performance enhancements for long-term predictions of scene sequences.

We further compare ICP-based approaches with MSTAU on quality of predicted visual representations, as demonstrating in Fig. 2. MSTAU is observed to accurately predict visual features in short-term prediction scenarios, including texture details and brightness intensity. However, as time progresses, inconsistencies arise in local brightness intensity compared to the ground truth (Fig. 2(d)). Influenced by rapidly changing high-frequency components, low-frequency information representing luminance is considered to be significantly changing during temporal extension. This contradicts the fact that low-frequency features change slowly, leading to distortion in the brightness intensity of areas with intense motion. Future work can address this issue by utilizing different time constant  $\nu$  for various temporal scales.

## 4 COMPARISON WITH OTHER PREDICTION METHODS ON SCENE RECONSTRUCTION

We compare the performance of the proposed MSTAU method with video prediction methods in scene and feature domain for scene reconstruction task as illustrating in Table. 2. On one hand, there is obvious distortion in reconstructed scene sequences, from which the

Processing Method		t+1	t+3	t+5	t+10
Scene Domain	ConvLSTM [6]	29.91	27.26	24.99	20.52
	PredRNN [9]	31.72	29.23	27.42	23.28
	PredRNN++ [10]	31.92	29.65	27.78	23.69
	MIM [11]	32.15	29.90	28.04	24.07
	E3D-LSTM [8]	32.20	29.95	28.10	24.93
	MAU [2]	32.27	30.02	28.16	25.79
	STAU [1]	33.41	31.10	29.41	27.83
Feature Domain	ConvLSTM [6]	31.04	29.32	27.57	23.98
	PredRNN [9]	32.67	30.86	29.02	25.24
	PredRNN++ [10]	32.77	30.95	29.11	25.75
	MIM [11]	32.98	31.15	29.30	26.18
	E3D-LSTM [8]	33.24	31.40	29.53	26.68
	MAU [2]	33.36	31.74	29.86	27.44
	STAU [1]	34.02	32.52	30.75	28.37
<b>MSTAU</b>		<b>34.43</b>	<b>33.60</b>	<b>32.18</b>	<b>30.01</b>

**Table 2: PSNR performance for reconstruction and prediction through approaches in scene-domain and feature-domain. Results indicate that the proposed MSTAU achieves more precise predictions in feature-domain, leading to reconstructed scenes with higher fidelity.**

	FLOPs (G)	Params (M)
PredRNN [9]	64.29	210.57
MAU [2]	31.72	112.28
<b>MSTAU</b>	<b><math>1.24 \times 4 + 1.38</math></b>	<b><math>2.14 \times 4 + 0.79</math></b>

**Table 3: Comparison on computational complexity between previous predictive approaches and the proposed MSTAU method.**

results predicted directly through previous video prediction methods are unsatisfactory. On the other hand, the proposed MSTAU method can finely perceive motion with assistance of intra-scale and inter-scale correlations compared to video prediction methods, further enabling the reconstruction of high-quality scenes. These demonstrate the best performance in spike perception and prediction, establishing an innovative perspective and a baseline for spike visual intelligence.

## 5 COMPARISON ON COMPUTATIONAL COMPLEXITY

We compare the computational complexity including floating point operations (FLOPs) and parameters between previous predictive approaches and the proposed MSTAU method. We downsample

the spike signals for 8 times to increase the informative density, while decreases the computational complexity in the meanwhile. As shown in Table. 3, the MSTAU method outperforms other approaches on both metrics significantly, illustrating competitive prospects for deployment on spike cameras. It’s vital to note that the statistical results for MSTAU method consider both 4 units at each temporal scale and the consequent compositor, formulated as *Single MSTAU*  $\times$  *Scales* + *Compositor*.

## 6 CODE

The implementation of ICP method and MSTAU is illustrated in the *code.py*.

## REFERENCES

- [1] Zheng Chang, Xinfeng Zhang, Shanshe Wang, Siwei Ma, and Wen Gao. 2022. STAU: a spatiotemporal-aware unit for video prediction and beyond. *arXiv preprint arXiv:2204.09456* (2022).
- [2] Zheng Chang, Xinfeng Zhang, Shanshe Wang, Siwei Ma, Yan Ye, Xiang Xinguang, and Wen Gao. 2021. Mau: A motion-aware unit for video prediction and beyond. *Advances in Neural Information Processing Systems* 34 (2021), 26950–26962.
- [3] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. *Noise reduction in speech processing* (2009), 1–4.
- [4] EM Harth, TJ Csermely, B Beek, and RD Lindsay. 1970. Brain functions and neural dynamics. *Journal of Theoretical Biology* 26, 1 (1970), 93–120.
- [5] Mohsen Pourahmadi. 1984. Taylor expansion of and some applications. *The American Mathematical Monthly* 91, 5 (1984), 303–307.
- [6] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28 (2015).
- [7] Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, and S Yu Philip. 2018. Predrnn+: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *International conference on machine learning*. PMLR, 5123–5132.
- [8] Yunbo Wang, Lu Jiang, Ming-Hsuan Yang, Li-Jia Li, Mingsheng Long, and Li Fei-Fei. 2018. Eidetic 3D LSTM: A model for video prediction and beyond. In *International conference on learning representations*.
- [9] Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S Yu. 2017. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Advances in neural information processing systems* 30 (2017).
- [10] Yunbo Wang, Haixu Wu, Jianjin Zhang, Zhifeng Gao, Jianmin Wang, S Yu Philip, and Mingsheng Long. 2022. Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 2 (2022), 2208–2225.
- [11] Yunbo Wang, Jianjin Zhang, Hongyu Zhu, Mingsheng Long, Jianmin Wang, and Philip S Yu. 2019. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9154–9162.
- [12] Jianhong Wu. 2011. *Introduction to neural dynamics and signal transmission delay*. Vol. 6. Walter de Gruyter.

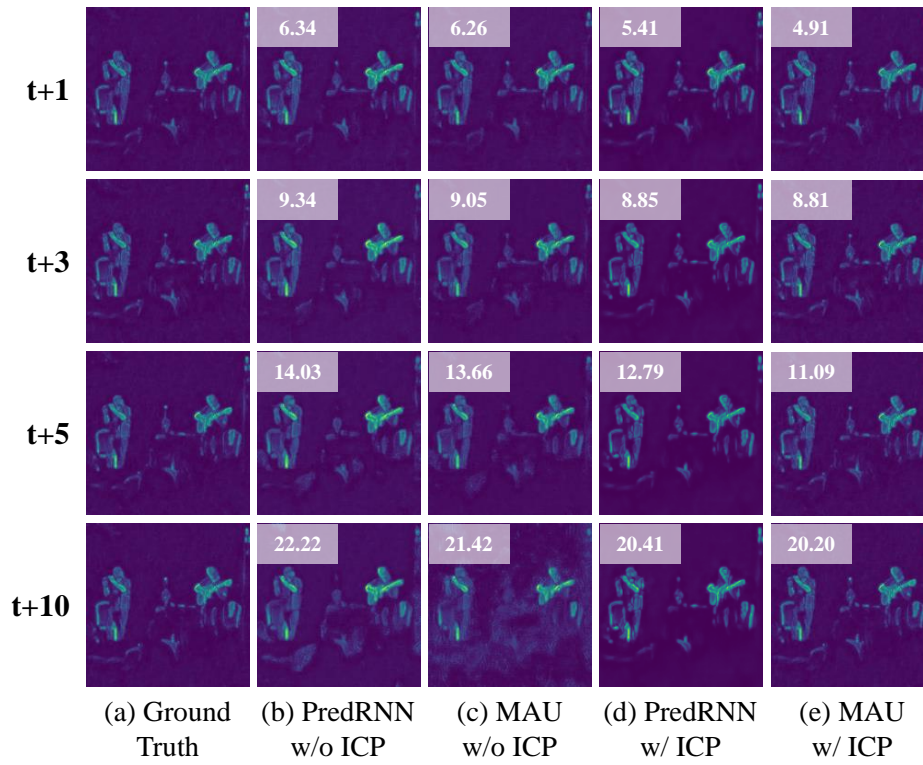


Figure 1: Visualization for predicted spike visual representation through different approaches. Results indicate that incorporating intra-scale correlation can decrease the prediction error measured by MSE significantly.

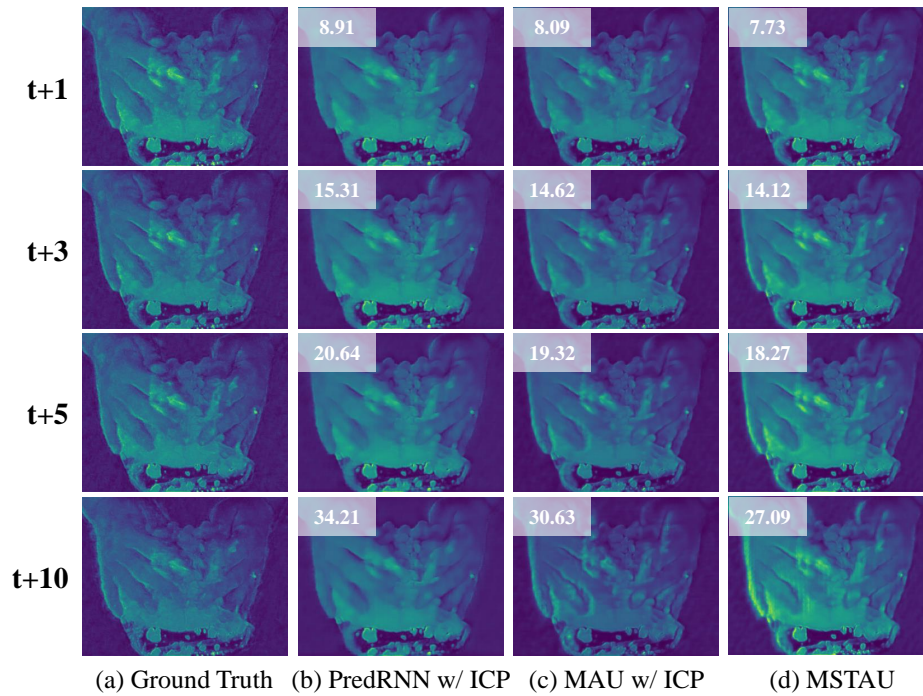


Figure 2: Visualization for predicted spike visual representation through different approaches. Results indicate that incorporating inter-scale correlation can further decrease the prediction error measured by MSE significantly.