SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Human action recognition using a time-delayed photonic reservoir computing

Chao KAI[1], Pu LI[1,2*], Yi YANG[1], Bingjie WANG[1], K.Alan SHORE[3] & Yuncai WANG[2]

[1]*Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan* 030024*, China;*
[2]*Guangdong Provincial Key Laboratory of Photonics Information Technology, School of Information Engineering, Guangzhou* 510006*, China;*
[3]*School of Computer Science and Electronic Engineering, Bangor University, Wales* LL57 1UT*, UK*

This supplementary document mainly includes three parts. In Appendix A, we introduce the theoretical model of TDPRC, including the details of the white chaos mask, the simulation model of the optical reservoir and the Tikhonov regression algorithm. In Appendix B, we analyze the effect of seven hyperparameters on the system states and the performance of our TDPRC. Then we give a set of optimal hyperparameters that enable the TDPRC to achieve a good recognition result. In Appendix C, we illustrate the typical recognition results, including the results after parameter optimization, the results based on only using the Tikhonov regression algorithm, and the results without using the HOG. In Appendix C, we also analyze the complexity and computational cost of our TDPRC by comparing it with some representative DNNs.

Appendix A Theoretical model of TDPRC

Appendix A.1 Details of the white chaos mask

The white chaos mask is experimentally generated by optical heterodyning of two external-cavity lasers. More details on white chaos can be found in [1].

Appendix A.2 Simulation model of the optical reservoir

In our method, the photonic reservoir is constructed with an optically injected semiconductor laser (R-laser) with a time delay loop. The transient response of the R-laser can be modeled as follows [2, 3]:

$$\begin{aligned}\frac{dE_r(t)}{dt} =& \frac{(1+i\alpha)}{2}\left[\frac{G_N(N_r(t)-N_0)}{1+\epsilon|E_r(t)|^2}-\frac{1}{\tau_p}\right]E_r(t)\\ &+\frac{\kappa}{\tau_i}E_r(t-\tau)\,exp(-i2\pi f_r\tau)\\ &+\frac{\kappa_{inj}}{\tau_i}E_d(t)\,exp(i2\pi\Delta f t)+\sqrt{2\beta N_r(t)}\chi(t)\end{aligned} \tag{A1}$$

$$\frac{dN_r(t)}{dt} = I_r-\frac{N_r(t)}{\tau_s}-\frac{G_N(N_r(t)-N_0)}{1+\epsilon|E_r(t)|^2}|E_r(t)|^2 \tag{A2}$$

$$E_d(t) = \sqrt{\frac{G_N\tau_p\left(N_0+\frac{1}{G_N\tau_p}\right)(I_d-1)}{G_N\tau_s+\epsilon}}exp(i\pi S(t)) \tag{A3}$$

Note, E_{RL} and E_{DL} are the complex electric field amplitude of the R-laser and D-laser, respectively. N_{RL} denotes the average carrier density carrier density of the R-laser. $\chi(t)$ is the Gaussian noise with zero mean and unity variance to represent the noise originate from the detection process and environmental change. β is the noise strength set as its typical value of 4.5×10^{-4}. The other parameters and their associated values are listed in Table A1. We point that the initial value of four structural parameters such as I_{RL}, κ_{inj}, κ, and Δf are set empirically and will be optimized in the training process.

Appendix A.3 Tikhonov regression algorithm

During training, the output vector of the TDPRC $\boldsymbol{y(n)}$ should as closely as feasible to the target vector $\boldsymbol{y}_{target}\boldsymbol{(n)}$. Thus, we use the Tikhonov regression algorithm to train the output connection weights $\boldsymbol{W}_i$, as shown in Eq. (A4). Note, λ in Eq. (A4) represents the ridge parameter whose typical value is 10^{-6}.

$$\boldsymbol{W}_i = argmin||\boldsymbol{y}_{target}\boldsymbol{(n)}-\boldsymbol{y(n)}||^2+\lambda||\boldsymbol{W}_i||^2 \tag{A4}$$

* Corresponding author (email: lipu8603@126.com)

Table A1 The laser parameters used in TDPRC

Parameter	Symbol	Value
Linewidth enhancement factor	α	5
Gain coefficient	G	5.2×10^{-12} m^3s^{-1}
Transparent carrier density	N_0	1×10^{24} m^{-3}
Saturation coefficient	ε	5×10^{-23}
Internal round-trip time	t_{in}	7.38×10^{-12} s
Photon lifetime	t_p	1.6×10^{-12} s
Carrier lifetime	t_s	2.5×10^{-9} s
Delayed feedback time	τ	5×10^{-9} s
Injection current of R-laser	I_{RL}	1.3 I_{th}
Injection strength	κ_{inj}	0.4
Feedback strength	κ	0.06
Frequency detuning	Δf	-10 GHz

Appendix B Hyperparameters optimization of TDPRC

The recognition performance of the TDPRC is closely related to the system states. In this part, we investigate the influence of seven hyperparameters on the recognition *ER* of the TDPRC: the sample size, the virtual node size, the mask standard deviation, the injection current of R-laser, the injection strength, the feedback strength, and the frequency detuning. By analyzing the influence of these hyperparameters, we obtain a set of optimal hyperparameters that enable the TDPRC to achieve a good recognition result.

Appendix B.1 Recognition *ER* of the TDPRC

The error rate (*ER*) is employed to evaluate the recognition performance of the TDPRC. The recognition *ER* is defined as Eq. (B1) and thus the recognition accuracy rate can be described as $(1-ER)\times100\%$. In Eq. (B1), q is the total number of the videos in the testing set, and p is the number of misidentified videos. In our work, we select the videos from 15 volunteers as the training set and the videos from another 10 volunteers that have never been used as the testing set. Based on the testing set, the recognition performance of the TDPRC can be verified.

$$ER=\frac{p}{q}\times100\% \tag{B1}$$

Appendix B.2 Effect of the sample size

Figure B1 shows the dependence of the recognition *ER* on the sample size. Here, about 1.5 s video stream (containing 35 frames of volunteers' video) is used for training. The triangles represent the specific values, and the solid line is the associated fitting curve. In principle, a large training sample size is helpful for the reservoir to sufficiently learn the characteristics of the recognition target. However, excess training samples will sharply accelerate the calculation complexity. Thus, when we set the sample size, the resource efficiency should also be taken into account. From Fig. B1, we found that the recognition *ER* decreases monotonically with increasing sample size and would converge to a stationary value of about 10%. But it should be noticed that once the sample size is larger than 10, the growth of recognition performance becomes significantly slower. Considering the calculation efficiency, we finally set the sample size to be 15. In addition, we point out that the recognition performance will be improved in a very small range when we use a sample size larger than 15.

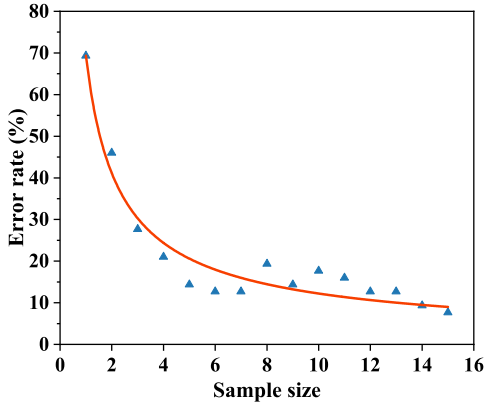

Figure B1 Dependence of the recognition *ER* on the training sample size.

Appendix B.3 Effect of the virtual node size

Then, we investigate the dependence of the recognition *ER* on the virtual node size. Figure B2 illustrates the associated results. In this simulation, we fix the interval time θ at 0.05 ns and change the virtual node size by adjusting the cycle times of the input data in the feedback loop. When the virtual node size is set as 100 (corresponding to one cycle in the feedback loop), the TDPRC shows

a poor performance with a recognition *ER* near 20%. With the increase of the number of circulating cycles, the *ER* will decrease correspondingly. The lowest recognition *ER* of 7.5% is obtained when the virtual node size is 900 (corresponding to nine cycles in the feedback loop). It should be emphasized that, in contrast to traditional ANNs, the virtual node size of our TDPRC can be readily expanded without varying the physical structure and the transient response of the reservoir can be extended to arbitrary dimensions [4]. From this point, the TDPRC with a simple structure shows great potential for ANN-based applications.

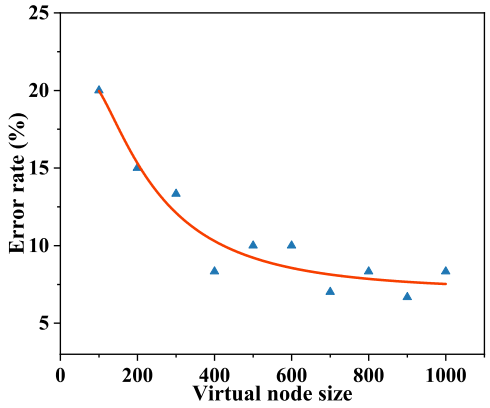

Figure B2 Dependence of the recognition *ER* on the virtual node size.

Appendix B.4 Effect of the mask standard deviation

The random mask signal plays a significant role in the TDPRC. On one hand, the random mask plays a role of connection weights between the input and reservoir layer. On the other hand, the input data can be nonlinearly mapped into a high-dimensional space by processing of the mask signals. In this work, we employ a chaos mask that can significantly enrich the nonlinear transient response of the TDPRC [5]. In Fig. B3, we investigate the effect of the standard deviation of the chaos mask on the recognition *ER*. The chaos sequence is collected in the experiment as in Ref. [6] and its standard deviation is rescaled to the value 1. The standard deviation of the chaos sequence is changed by multiplying it with a constant σ. From Fig. B3, we can see that the recognition *ER* first decreases rapidly and then slowly increases as the standard deviation increases. Finally, the chaos mask with a standard deviation of 0.3 is selected for our TDPRC.

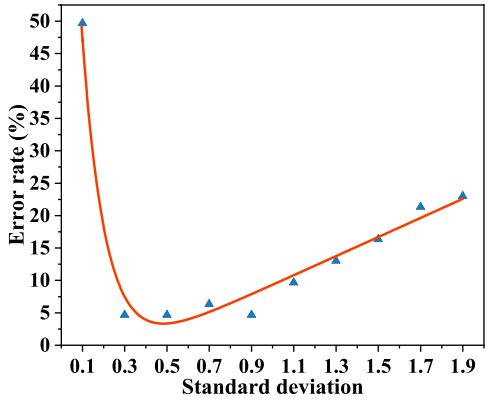

Figure B3 Dependence of the recognition *ER* on the standard deviation of chaos mask.

Appendix B.5 Effect of four structure parameters in the reservoir

In this section, we investigate the effect of four structure parameters of the reservoir layer on the TDPRC performance, namely the bias current I_{RL}, injection intensity κ_{inj}, feedback strength κ, and frequency detuning Δf. These parameters directly affect the nonlinear output of the R-laser in our TDPRC.

Appendix B.5.1 *Effect of the bias current*

Figure B4(a) gives the dependence of the recognition performance on the bias current I_{RL}. The other structural parameters are empirically set at $\kappa_{inj} = 0.4$, $\kappa = 0.06$, and Δf = -10 GHz, respectively. We find that the system performance first slowly improves and then declines with increasing R-laser bias current. The lowest recognition *ER* of 3.67% is achieved at I_{RL}=1.35 I_{th}. Figure B4(b) depicts the associated bifurcation diagram of the R-laser. When I_{RL} is biased at 1.35 I_{th}, the R-laser works near the boundary between the chaos oscillation region and the quasi-periodic oscillation region. In this condition, the system can provide a rich nonlinear transient response of input signal.

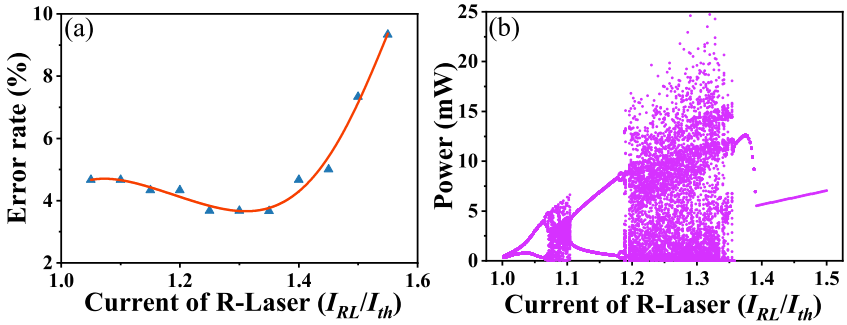

Figure B4 (a) The effect of current of R-laser on the recognition *ER*. (b) Bifurcation diagram with increase of the current I_{RL} for κ_{inj} = 0.4, κ = 0.06, and Δf = -10 GHz.

Appendix B.5.2 *Effect of the injection strength*

Figure B5 depicts the associated influence of injection strength κ_{inj}. Here, the bias current I_{RL} is set at 1.35 I_{th}, while other parameters remain unchanged. From Fig. B5(a), one can observe that with increased injection strength, the recognition *ER* firstly decreases and then converges to a stationary value. The lowest *ER* of 2.67% can be obtained when the injection strength κ_{inj} is set to 0.5. From the associated bifurcation diagram in Fig. B5(b), we can determine that when the injection strength is set as 0.5, the R-laser works in the injection-locked state and is locked by the D-laser. For the TDPRC based on semiconductor laser subject to optical injection and feedback, optical injection from D-laser is necessary to keep the R-laser at a dynamically stable steady state.

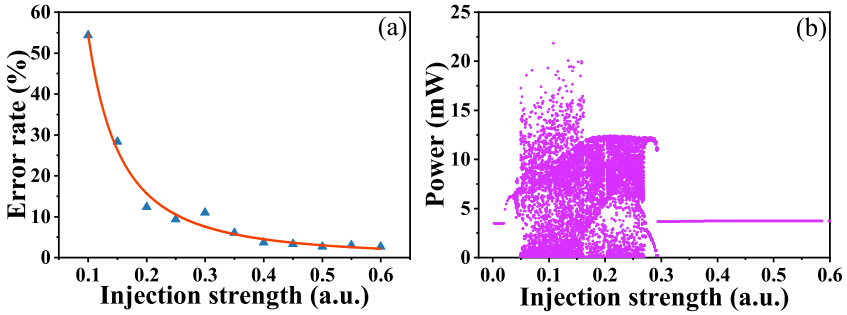

Figure B5 (a) The effect of injection strength on the recognition *ER*. (b) Bifurcation diagram with increase of the injection strength κ_{inj} for I_{RL} = 1.35 I_{th}, κ = 0.06, and Δf = -10 GHz.

Appendix B.5.3 *Effect of the feedback strength*

Figure B6(a) shows the effect of feedback strength κ on recognition performance. The bias current I_{RL} and the injection strength κ_{inj} are set to be 1.35 I_{th} and 0.5, respectively. When κ is equivalent to 0.1125, a lower *ER* rate can be obtained. From the associated bifurcation diagram [Fig. B6(b)], we can find that the output state of R-laser enters the chaotic state after a quasi-periodic state as the feedback strength κ increases. When κ is 0.1125, the R-laser operates in a chaotic state, where a rich nonlinear dynamic response can be provided. After optimizing the feedback strength κ, we can finally reduce the recognition *ER* to 2.33%.

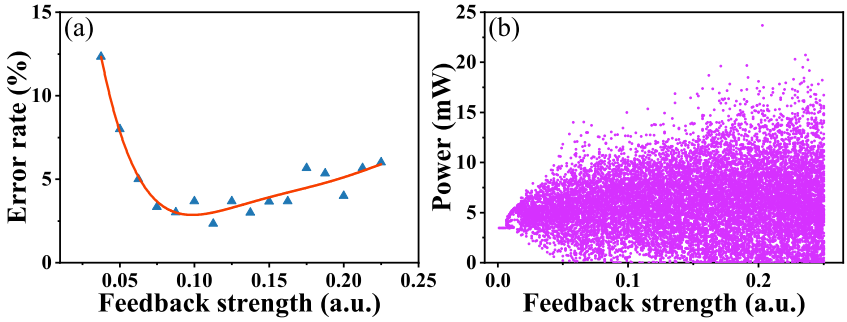

Figure B6 (a) The effect of feedback strength on the recognition *ER*. (b) Bifurcation diagram with increase of the feedback strength κ for I_{RL} = 1.35 I_{th}, κ_{inj} = 0.5, and Δf = -10 GHz.

Appendix B.5.4 *Effect of the frequency detuning*

Figure B7(a) shows the dependence of the TDPRC performance on the frequency detuning Δf. From it, we can see that the variation trend of the *ER* is approximately symmetrical about the frequency detuning Δf = -15 GHz, where the lowest recognition

ER of 2%. Through observing the associated bifurcation diagram [Fig. B7(b)], we find that the lowest recognition *ER* is obtained when the R-laser works at the edge of the chaos region. This is consistent with that in Ref. [7].

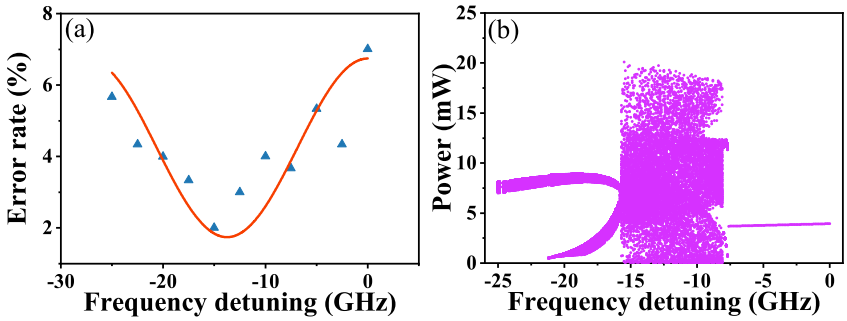

Figure B7 (a) The effect of frequency detuning on the recognition *ER*. (b) Bifurcation diagram with difference of the frequency detuning Δf for $I_{RL} = 1.35\ I_{th}$, $\kappa_{inj} = 0.5$, and $\kappa = 0.1125$.

Appendix C Typical recognition results

In this part, we utilize the optimized TDPRC to identify the actions of the other 10 volunteers in the testing set. Table C1 illustrates typical recognition results, where a confusion matrix is used. From it, we observe that handclapping and running can be perfectly recognized. For other actions, the recognition accuracies are slightly smaller than 100%. For example, 2% of the jogging behavior videos are incorrectly recognized as running, and the remaining 98% of videos are correctly identified as jogging. This may be caused by one human's action of running being similar to another human's jogging. Therefore, the results show several errors between these two actions. Overall, the total recognition accuracy is 98% for all human behavior classes using the TDPRC. We point out that the current recognition accuracy can be further improved by increasing the training sample size of relevant actions, when the resource efficiency is not considered.

Table C1 Typical recognition results after parameters optimization

	Actual Box[1)]	Actual HC	Actual HW	Actual Jog	Actual Run	Actual Walk
Recognized Box	98%	0	2%	0	0	0
Recognized HC	0	100%	0	0	0	0
Recognized HW	0	0	96%	0	0	0
Recognized Jog	0	0	2%	98%	0	4%
Recognized Run	0	0	0	2%	100%	0
Recognized Walk	2%	0	0	0	0	96%

Then, we use only the Tikhonov regression algorithm to calculate the output connection weight directly without the reservoir layer. The associated recognition results are shown in Table C2. In this case, a relatively high recognition accuracy is achieved for the running and jogging. For other actions, the associated recognition accuracies decrease by more than 20%. For example, 60% walking behavior videos are correctly identified, and the remaining 40% are wrongly identified as running or jogging. This means that the human action features are difficult to be classified only using the ridge regression in the low-dimensional space. That is, our TDPRC can indeed improve the performance of the HAR, because our proposed TDPRC classifies the input data by mapping the features from low-dimensional space to high-dimensional space.

Table C2 Recognition results based on only using Tikhonov regression algorithm

	Actual Box	Actual HC	Actual HW	Actual Jog	Actual Run	Actual Walk
Recognized Box	74%	0	0	0	0	0
Recognized HC	0	68%	0	0	0	0
Recognized HW	0	0	74%	0	0	0
Recognized Jog	22%	10%	16%	100%	2%	28%
Recognized Run	4%	22%	6%	0	98%	12%
Recognized Walk	0	0	4%	0	0	60%

1) Note, the Box, HC, HW, Jog, Run, and Walk stand for the actions of boxing, handclapping, handwaving, jogging, running, and walking, respectively.

Next, we supplement a human action recognition result without using the HOG as shown in Table C3. In this case, the pixel vector of the human action video is directly multiplied by the chaos mask to generate the input signal and fed into the photonic reservoir, after undergoing the same frame extraction and size cropping. After calculation, we find that the overall recognition accuracy without the HOG is as high as 94%. Comparing with the recognition result in Table C1, we can confirm that the HOG technique only can improve the recognition accuracy in a very small range by reducing redundant features.

Table C3 Recognition results without using the HOG

	Actual Box	Actual HC	Actual HW	Actual Jog	Actual Run	Actual Walk
Recognized Box	96%	0	0	0	0	0
Recognized HC	0	98%	0	0	0	0
Recognized HW	0	0	96%	2%	0	0
Recognized Jog	4%	0	4%	88%	6%	6%
Recognized Run	0	2%	0	6%	94%	2%
Recognized Walk	0	0	0	4%	0	92%

At last, we consider the complexity and computational cost of our TDPRC by comparing with other typical DNNs for HAR, as shown in Table C4. There are various of DNNs have been reported for HAR, but most of them can be categorized into convolutional neural networks (CNNs) [8–16]. Among them, the visual geometry group network-16 (VGG-16) [12], long short-term memory-3DCNN (LSTM-3DCNN) [13] and CNN-Poisson distribution along with univariate measure (CNN-PDaUM) [14] are three representatives with the state-of-the-art performance. From the view of complexity, ones can find that our TDPRC with only one hidden layer and one physical node can reach a level of accuracy comparable or even superior to the CNNs with dozens of hidden layers and tens of thousands of physical nodes. On the other hand, the computational cost can be estimated using the number of multiply-adds per recognition [16]. The more the number of multiply-adds per recognition is, the higher the computational cost is. From Table C4, we can confirm that our TDPRC only needs 220500 multiply-adds per recognition for a 98% accuracy rate. That is at the same level of accuracy rate as the CNN-PDaUM where at least 4.2×10^8 multiply-adds per recognition are required. This indicates that our method is computationally efficient in the HAR without sacrificing the recognition accuracy rate.

Table C4 Comparison of TDPRC with various other deep learning methods

Methods	VGG-16	LSTM-3DCNN	CNN-PDaUM	TDPRC
Accuracy rate	91.83%	93.6%	98.7%	98%
Number of hidden Layers	16	22	24	1
Number of physical nodes	13416	273650	22144	1
Number of multiply-adds	$\sim1.38\times10^8$	$\sim1.39\times10^8$	$\sim4.2\times10^8$	220500

References

1 Wang A B, Wang B J, Li L, et al. Optical heterodyne generation of high-dimensional and broadband white chaos. IEEE J Sel Top Quantum Electron, 2015, 21: 531-540
2 Lang R, Kobayashi K. External optical feedback effects on semiconductor injection laser properties. IEEE J Quantum Electron, 1980, 16: 347-355
3 Jiang N, Wang Y J, Zhao A K, et al. Simultaneous bandwidth-enhanced and time delay signature-suppressed chaos generation in semiconductor laser subject to feedback from parallel coupling ring resonators. Opt Express, 2020, 28: 1999-2009
4 Nguimdo R M, Verschaffelt G, Danckaert J, et al. Delay-based reservoir computing using semiconductor ring lasers. Nonlinear Opt Its Appl VIII; Quantum Opt III, 2014, 1: 292-297
5 Nakayama J, Kanno K, Uchida A. Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal. Opt Express, 2016, 24: 8679-8692
6 Wang A B, Wang B J, Li L, et al. Optical heterodyne generation of high-dimensional and broadband white chaos. IEEE J Sel Top Quantum Electron, 2015, 21: 531-540
7 Zhang H, Feng X, Li B X, et al. Integrated photonic reservoir computing based on hierarchical time-multiplexing structure. Opt Express, 2014, 22: 31356-31370
8 Chakraborty S, Mondal R, Singh P K, et al. Transfer learning with fine tuning for human action recognition from still images. Multimed Tools Appl, 2021, 80: 20547-20578
9 Guha R, Khan A H, Singh P K, et al. CGA: A new feature selection model for visual human action recognition. Neural Comput Appl, 2021, 33: 5267-5286
10 Huang L J, Huang Y, Ouyang W L, et al. Part-level graph convolutional network for skeleton-based action recognition. in: IEEE Transactions on Circuits and Systems for Video Technology, 2020. 11045-11052
11 Ye F F, Pu S L, Zhong Q Y, et al. Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition. in: Proceedings of the 28th ACM International Conference on Multimedia, 2020. 55-63
12 Liu X, Qi D Y, Xiao H B. Construction and evaluation of the human behavior recognition model in kinematics under deep learning. J Amb Intel Hum Comp, 2020, 1: 1-9
13 Ouyang X, Xu S J, Zhang C Y, et al. A 3D-CNN and LSTM based multi-task learning architecture for action recognition. IEEE Access, 2019, 7: 40757-40770
14 Khan M A, Zhang Y D, Khan S A, et al. A resource conscious human action recognition framework using 26-layered deep convolutional neural network. Multimed Tools Appl, 2021, 80: 35827-35849

15 Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. in: Advances in Neural Information Processing Systems, 2014. 568–576
16 Sun L, Jia K, Yeung D Y, et al. Human action recognition using factorized spatio-temporal convolutional networks. in: IEEE International Conference on Computer Vision, 2015. 45–65
17 LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86: 2278-2324