ICIG 2025

Paper ID 119



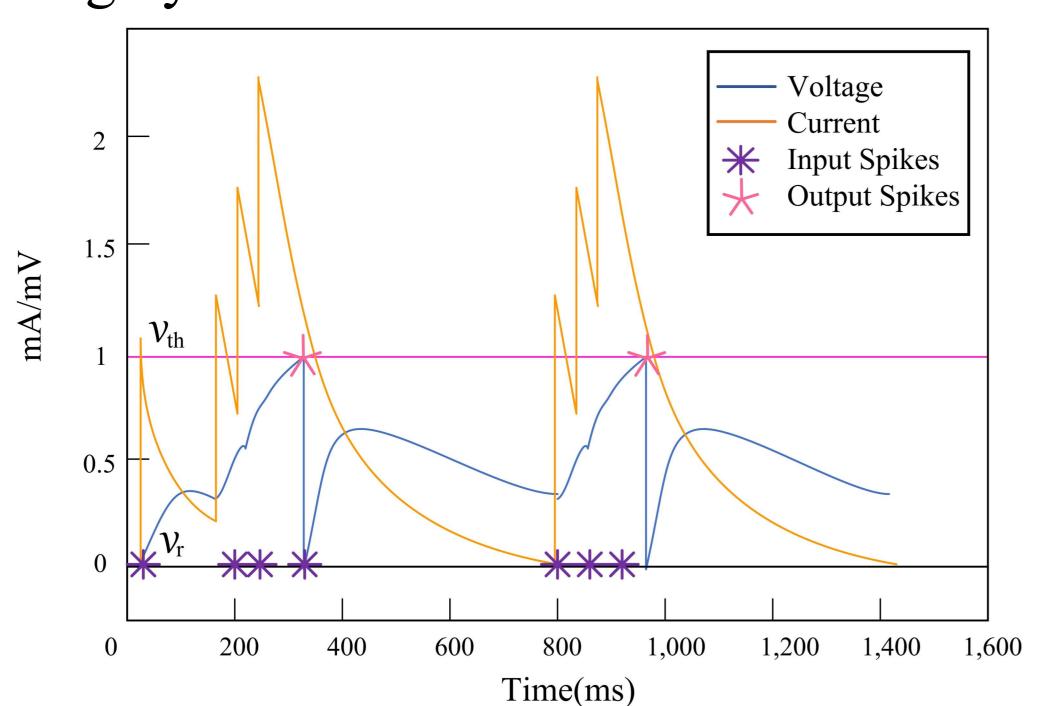
SNN-PAR: Energy Efficient Pedestrian Attribute Recognition via Spiking Neural Networks

Haiyang Wang, Qian Zhu, Mowen She, Yabo Li, Haoyu Song, Minghe Xu, Jin Tang, Xiao Wang

GitHub: https://github.com/Event-AHU/OpenPAR

Background

- 1. Current CNN and Transformer-based PAR models are computationally intensive with low energy efficiency, limiting their deployment.
- 2. SNNs achieve high energy efficiency through event-driven operation, core to which is the LIF neuron that processes information via bio-inspired spiking dynamics.



Methodology **Teacher MM-Foundation Network** (Used in Training Phase Only) Multi-modal **Transformer Vision Encoder Predicted Results** Short hair Holding a handbag KL Divergence between Wearing glasses $R(V_T, S)$ and $R(V_S, S)$ Text Encoder → Response-Level Age in (40, 50) **Attribute Set** → Attribute N **Predicted Results** Spiking Transformer Block xL **Student SNN Network** X' = SN(X),X' = SN(X), $Q = SN_Q(ConvBN_Q(X')),$ $K = SN_K(ConvBN_K(X')),$ $ConvBN_K(X'),$ $ConvBN_K(X$ $V = SN_V(ConvBN_V(X')),$ $SSA(Q, K, V) = ConvBN(SB(QK^TV * s)), \quad \mathcal{L} = \mathcal{L}_{CE} + \alpha \mathcal{L}_{respKD} + \beta \mathcal{L}_{featKD}$

Experiments

Table 1. Comparison with SOTA methods on PETA, PA100K and RAPv1 datasets.

Methods	PETA					PA100K					RAPv1				
Methods	mA	Acc	Prec	Recall	F1	mA	Acc	Prec	Recall	F1	mA	Acc	Prec	Recall	F1
SSCsoft [19]	86.52	78.95	86.02	87.12	86.99	81.87	78.89	85.98	89.10	86.87	82.77	68.37	75.05	87.49	80.43
IAA [37]	85.27	78.04	86.08	85.80	85.64	81.94	80.31	88.36	88.01	87.80	81.72	68.47	79.56	82.06	80.37
MCFL [28]	86.83	78.89	84.57	88.84	86.65	81.53	77.80	85.11	88.20	86.62	84.04	67.28	73.44	87.75	79.96
DRFormer [30]	89.96	81.30	85.68	91.08	88.30	82.47	80.27	87.60	88.49	88.04	81.81	70.60	80.12	82.77	81.42
VAC [13]	-	-	-	-	-	82.19	80.66	88.72	88.10	88.41	81.30	70.12	81.56	81.51	81.54
DAFL [20]	87.07	78.88	85.78	87.03	86.40	83.54	80.13	87.01	89.19	88.09	83.72	68.18	77.41	83.39	80.29
CGCN [10]	87.08	79.30	83.97	89.38	86.59	-	-	-	-	h (-	84.70	54.40	60.03	83.68	70.49
CAS [38]	86.40	79.93	87.03	87.33	87.18	82.86	79.64	86.81	87.79	85.18	84.18	68.59	77.56	83.81	80.56
VTB [4]	85.31	79.60	86.76	87.17	86.71	83.72	80.89	87.88	89.30	88.21	82.67	69.44	78.28	84.39	80.84
SNN-PAR (Ours)	80.58	73.55	81.76	82.79	81.96	73.86	71.70	83.03	81.30	81.67	75.43	63.06	74.67	78.28	75.94
Female Age 18 To Side Hand Ba Long Slee Trousers	60 g eve		Lor	18 To 60 Side Glasses ng Sleeve rousers			Age Lor Si	Female 2 18 To 6 Back ng Sleeve kirt And Dress	9200	The state of the s	Should	To 60 ack ler Bag Sleeve		Ag	Female ge 18 To 6 Side nort Sleeve Shorts
edestrian Back	Lowe strip	Tre	ousers	Boots	s S	horts	Ped	estrian	Bacl	<i>z</i>	ower tripe	Trous	sers]	Boots	Short
			M				(d)								
							(e)	G III							
							(f)	2	The same		2	1			24

Contributions

- 1. We propose an energy-efficient spiking transformer network for pedestrian attribute recognition, termed SNN-PAR.
- 2. To enhance the performance of SNN-PAR further, we adopt knowledge distillation from the artificial neural networks to guide the learning of spiking transformer networks.
- 3. Comprehensive experiments carried out on three publicly available datasets show that our proposed SNN-PAR model is effective for the PAR task.

References

- [1] Cheng et al., A simple visual-textual baseline forpedestrian attribute recognition. IEEE Transactions on Circuits and Systems forVideo Technology 32(10), 6994–7004
- [2] Zhou et al., Spikingformer:Spike-driven residual learning for transformer-based spiking neural network. arXivpreprint arXiv:2304.11954
- [3] Wang et al., Pedestrian attribute recognition: A survey[J]. Pattern Recognition, 2022, 121: 108220.

