Towards more efficient few-shot learning based human gesture recognition via dynamic vision sensors

Linglin Jing(Lboro), Yifan Wang(Lboro), Tailin Chen(NU), Shirin Dora(Lboro), Zhigang Ji(SJTU), Hui Fang(Lboro)

Problem

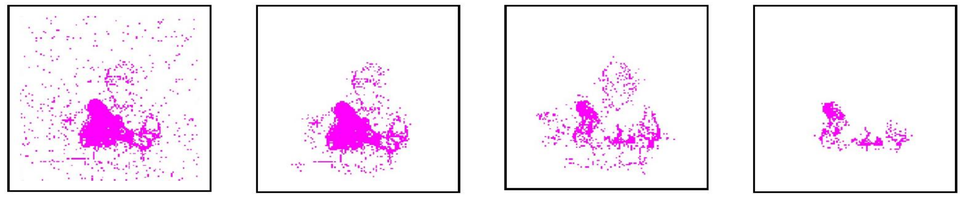
Real-time human gesture recognition on edge device

- ◆Generalised model for **new** gesture classes
- ◆High speed **motion** information recognition
- ◆**Power consumption** during pre-trained network

Main ideas

- ◆ Training with few-shot learning strategies
- ◆ Modeling on Dynamic Vision Sensor (DVS) data
- ◆ Use spiking neural network (SNN) as the backbone

Noise suppression



Before noise suppression | Spatial noise suppression | Spatial &Temporal noise suppression | Motion trajectory capture

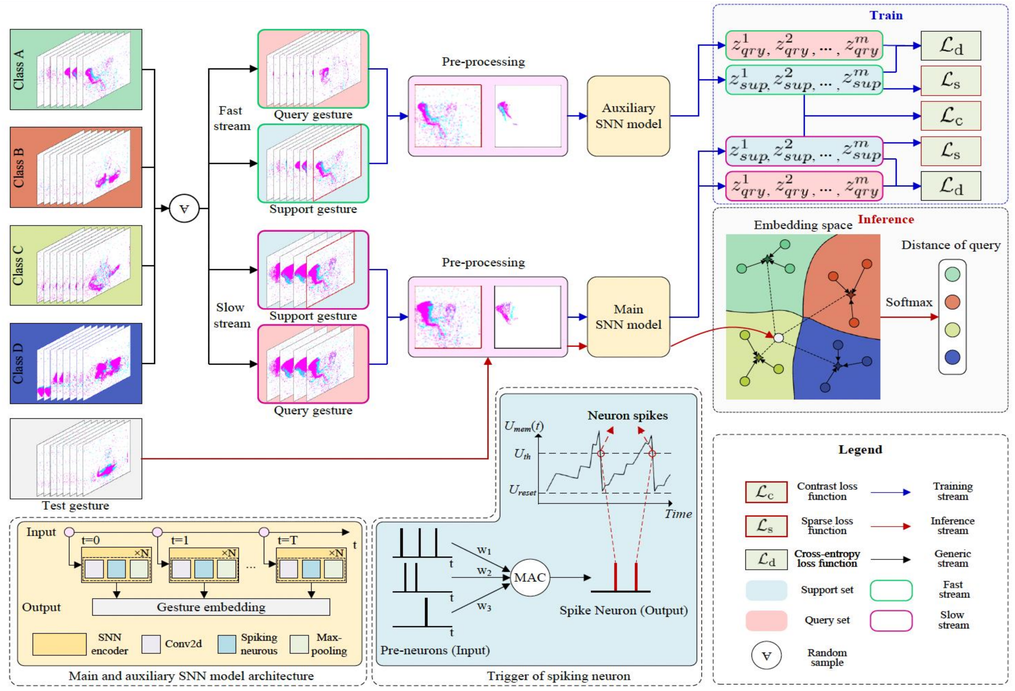
Loss terms

Dual-speed stream contrastive loss: $\mathcal{L}_c = -\frac{1}{N}\sum_{i=1}^{N} \cos\mathrm{sim}\left(f_\phi^F(x_i), f_\zeta^S(x_i)\right)$

Channel-wise sparsity loss: $\mathcal{L}_s = -\frac{1}{HW}\sum_{i=1}^{HW} \mathrm{MSE}\left(Max(f_\phi^F(x_i)), Avg(f_\phi^F(x_i))\right)$

Classification Loss: $\mathcal{L}_d^F = -\log p(y=k \mid m^F), \quad \mathcal{L}_d^S = -\log p(y=k \mid m^S)$

Network structure and training strategy



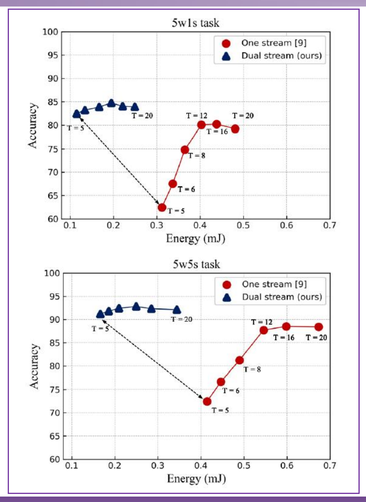
Experiments

Proposals	Methods	Fine Tune	Aug.	5w1s Acc.	5w5s Acc.
LPR [37]	Transfer	Y	Y	40.00%	43.30%
MAML [13]	Meta	Y	F	45.50%	53.70%
SOEL [38]	Transfer	Y	Y	64.70%	65.10%
MTO [15]	Meta	N	N	63.20%	73.30%
PLIF [10]	Metric	N	N	80.21%	88.53%
This Work	Metric	N	N	**84.75 %**	**92.82 %**

6+5-WAY Few-shot accuracy on DVS-gesture

Method	Operations	Latency(s)	Energy(mJ)	Energy ratio
Protot(TPU) [35]	$3.89E^{10}$	$5.40E^{-1}$	45.55	×240
Protot(Memristor) [35]	$3.89E^{10}$	$6.80E^{-3}$	1.56	×8
PLIF [10]	$2.04E^{7}$	$7.14E^{-2}$	0.48	×2.5
Ours	$1.29E^{7}$	$4.52E^{-2}$	0.31	×1.6
Ours(+N)	$8.94E^{6}$	$3.13E^{-2}$	0.21	×1.1
Ours(+N+S)	$\mathbf{8.23E^{6}}$	$2.88E^{-2}$	**0.19**	**×1**

Power consumption comparison



Contributions

- ◆ Construct a FSL-HGR SNN model based on **contrastive learning** strategy
- ◆ Proposed a data **pre-processing** strategy for energy efficiency
- ◆ Embed a **channel-wise spike sparsity loss** for energy efficiency