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

Network structure and training strategy

Noise suppression

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

Loss terms

- Dual-stream contrastive loss: \( \mathcal{L}_c = - \frac{1}{N} \sum_{i=1}^{N} \cos \left( f_a^a (x_i), f_b^b (x_i) \right) \)
- Channel-wise sparsity loss: \( \mathcal{L}_s = -\frac{1}{HW} \sum_{i=1}^{HW} \text{MSE} \left( \text{Max}(f_a^a (x_i)), \text{Avg}(f_b^b (x_i)) \right) \)
- Classification Loss: \( \mathcal{L}_d^y = -\log p(y = k \mid m^x), \mathcal{L}_d^s = -\log p(y = k \mid m^s) \)

Experiments

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

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