Graduate Institute of Electronics Engineering, NTU

Undergraduate Group

Mid-term report – Transformer on sEMG

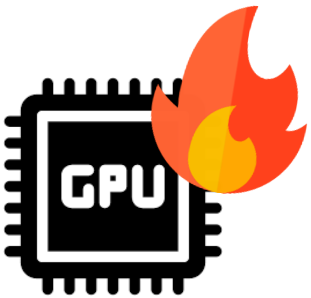
Presenter: Howard

Advisor: Prof. An-Yeu (Andy) Wu

Date: 2023/11/10

Outline

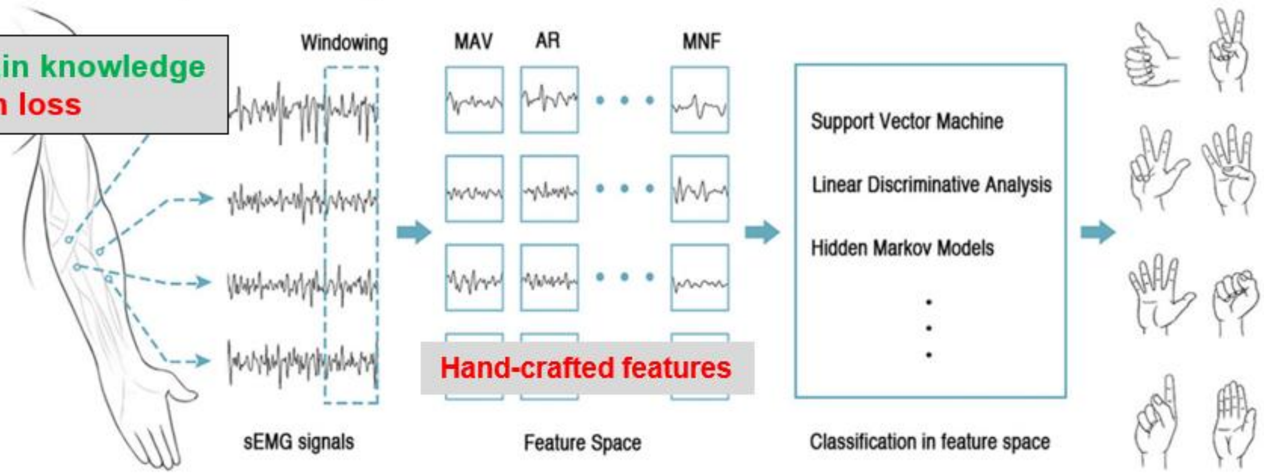
- Basic Transformer Procedure(ViT)
- Improved transformer
  - CViT
  - LST-EMG NET
- Future Work



Hand Gesture Recogntion

Feature-based*: Pre-processing + feature extractor + classifier*

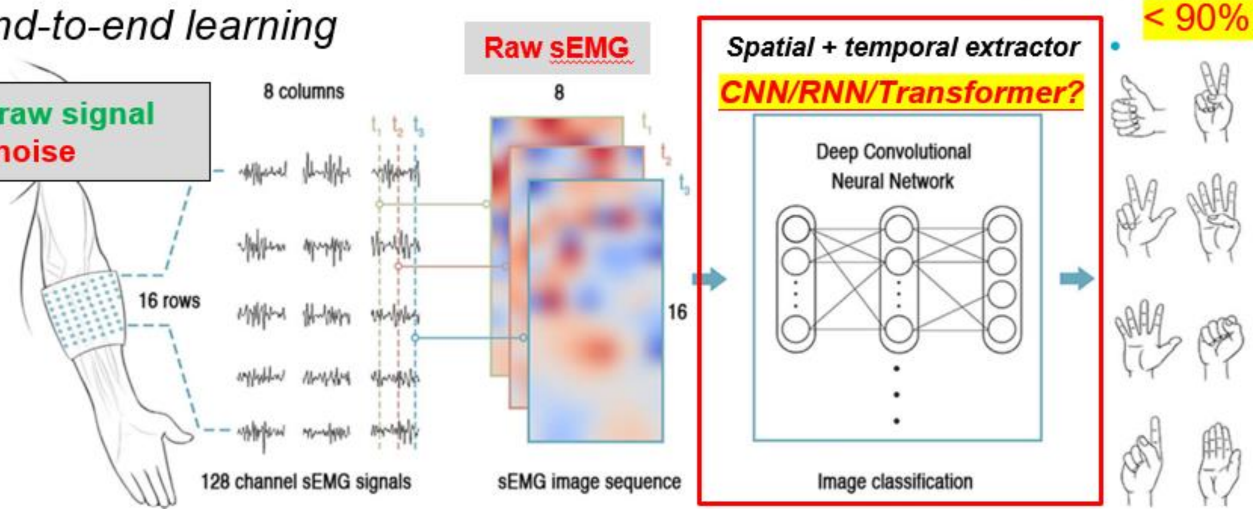
- Prior domain knowledge
- Information loss



DL-based*: End-to-end learning*

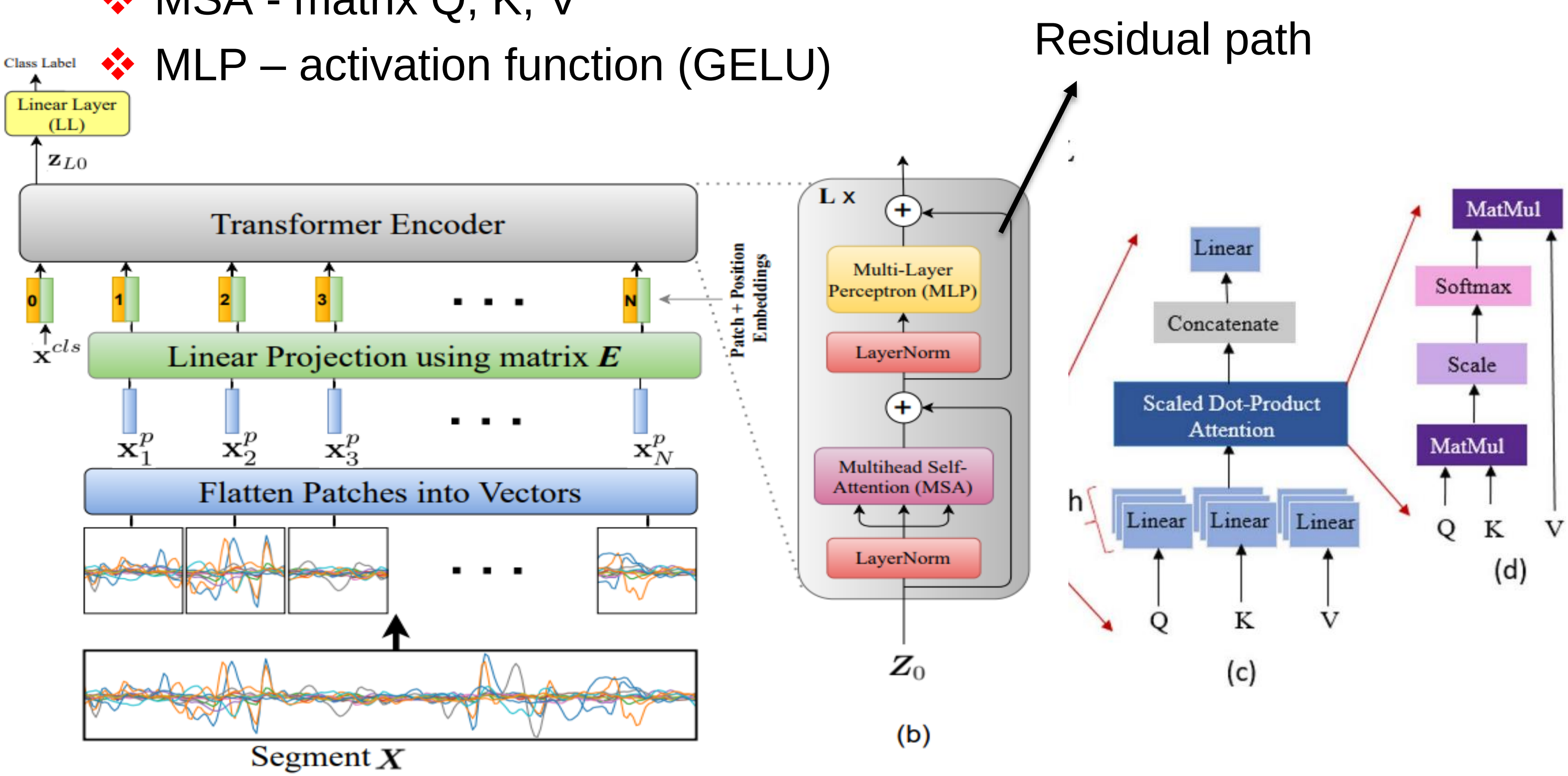
< 90% over 17 gestures

- Informative raw signal
- Redundant noise



TEMG : Basic Transformer Concept[2][3]

- x^{cls} - a trainable token
- Position embedding → encode the order of the input sequence
- Transformer encoder
  - MSA - matrix Q, K, V
  - MLP – activation function (GELU)

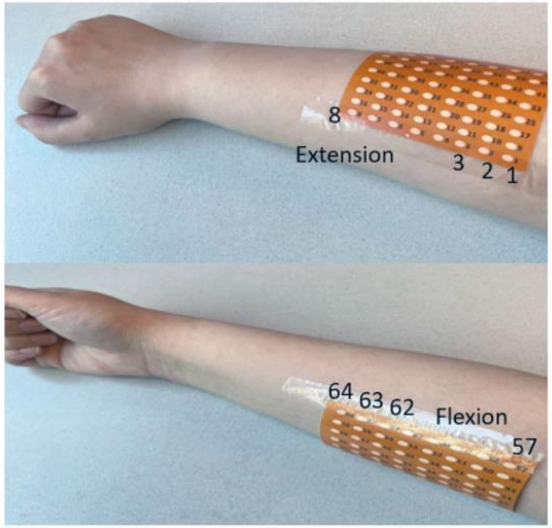


Result of Basic Transformer

- Performance improve while the computing complexity reduce 7 times [2]
- more electrodes (channel) → better performance [3]

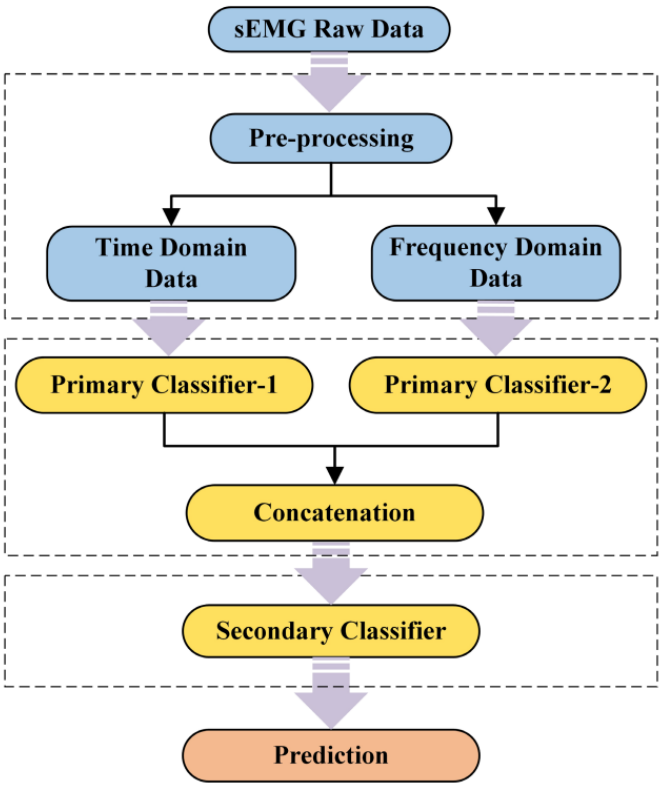
		300ms	
		Params	**Accuracy (%)**
Reference [8]	4-layer 3rd Order Dilation	466, 944	82.4
	4-layer 3rd Order Dilation (pure LSTM)	_	79.7
	SVM	_	30.7
Our Method	Model 1	**20,593**	80.88
	Model 4	65, 713	**82.93**

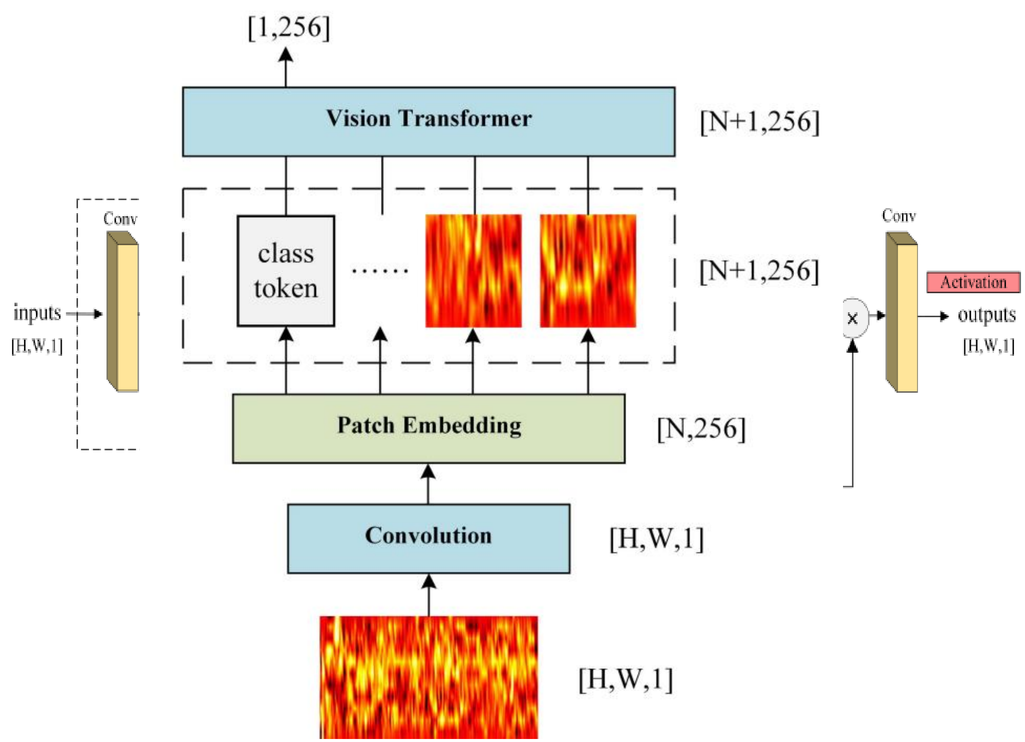
Reference	Window size (ms)	# Channels	Accuracy (%)	Train/Test Split
Ref[43]	200	128	84.6 (NA)	5-fold Cross Validation
CT-HGR-V1	250	128	91.98 (±2.22)	5-fold Cross Validation
CT-HGR-V2	250	128	92.88 (±2.10)	5-fold Cross Validation



Ensemble Learning - CViT[1]2022

- Time domain + frequency domain → better performance
- FFT transforms time domain → frequency domain
- Convolution → improve capability of generalization



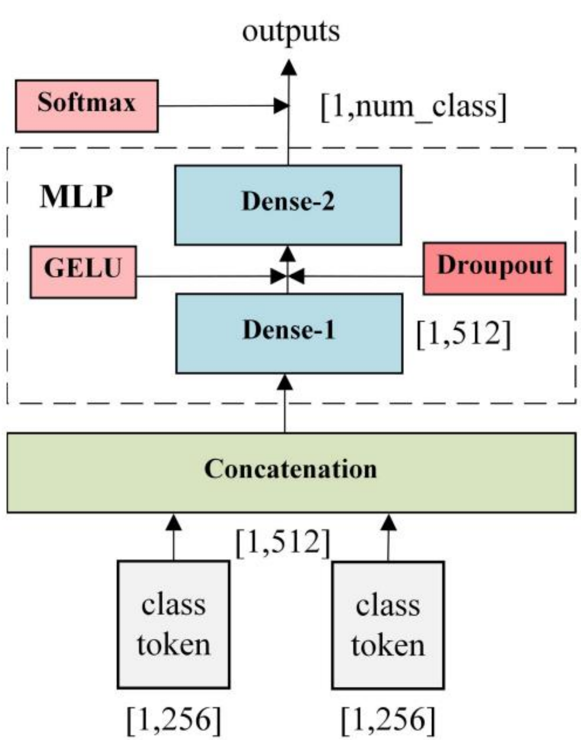


CViT : Convolution Vision Transformer

Primary classifier

Performance Between Stacking Strategies [1]2022

- Class token of time and frequency concatenation
- Dropout : prevents units from over-fitting too much



Secondary classifier

Stacking Strategies	Accuracy
Time	79.58%
Frequency	78.65%
Time + Time	81.06%
Frequency + Frequency	79.27%
Time + Frequency	**81.20%**

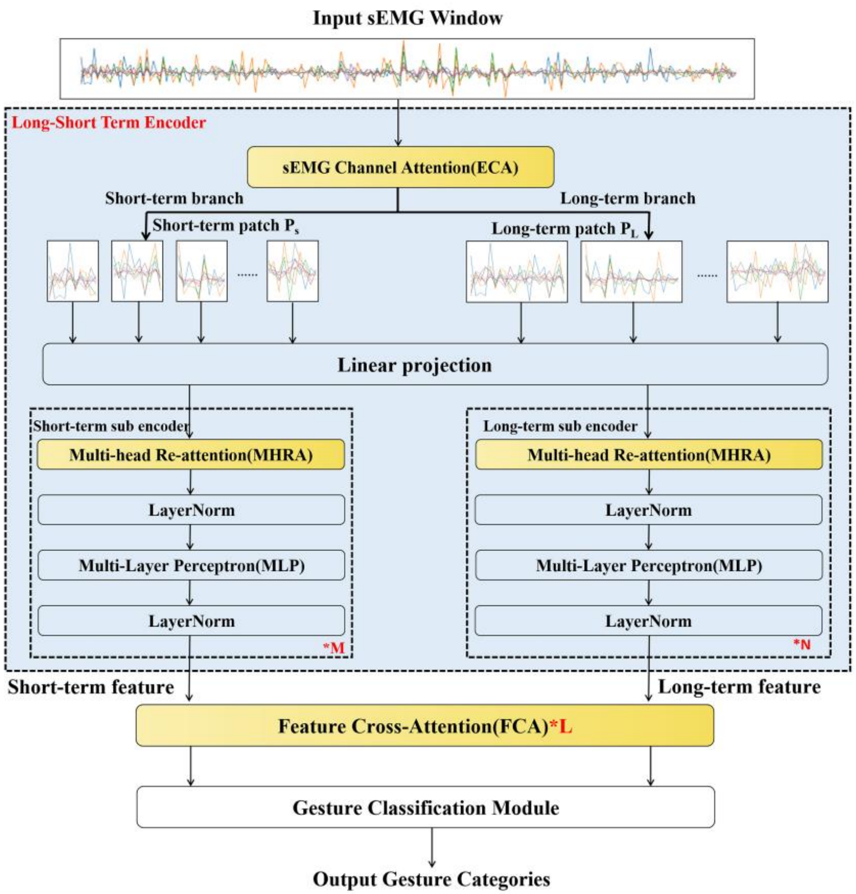
Method	Database	Number of movements	Window length	Accuracy
Random Forest [22]	NinaPro DB2	49	200ms	75.27%
CNN [4]	NinaPro DB2	49	200ms	78.71%
CviT	NinaPro DB2	49	200ms	**80.02%**
LSTM [8]	NinaPro DB2-E1	17	300ms	79.19%
ViT [13]	NinaPro DB2-E1	17	200ms / 300ms	82.05% / 82.93%
CviT	NinaPro DB2-E1	17	200ms / 300ms	**83.47% / 84.09%**
LDA [2]	NinaPro DB5-A/B	12 / 17	200ms	69.49% / 61.75%
SVM [2]	NinaPro DB5-A/B	12 / 17	200ms	67.9% / 58.27%
LSTM-CNN [10]	NinaPro DB5-A/B	12 / 17	200ms	71.66% / 61.4%
CviT	NinaPro DB5-A/B	12 / 17	200ms	**76.83% / 73.23%**

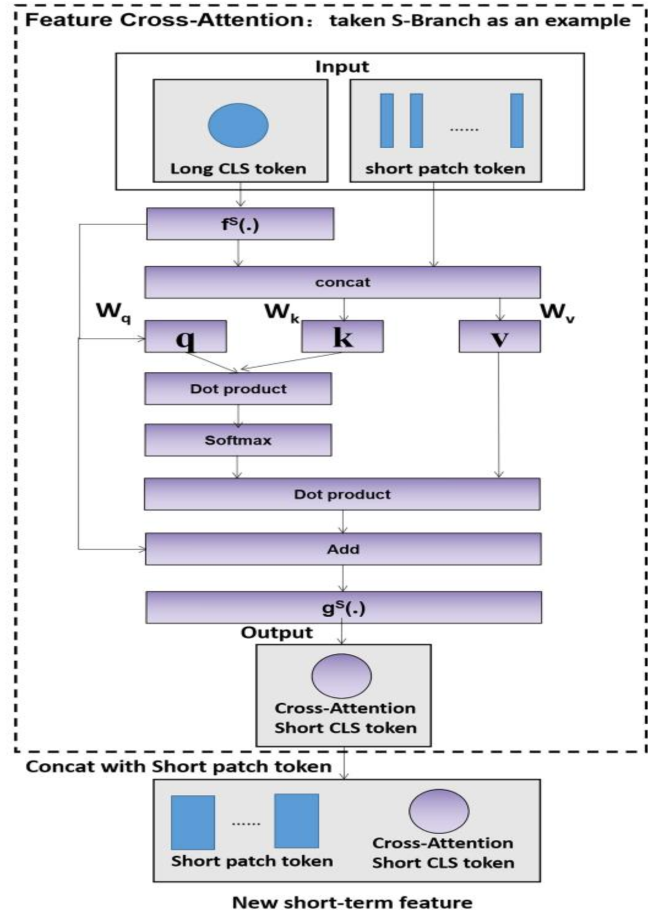
Best Performance→Time + Frequency

LST-EMG NET [5]2023

LST(long short term)

- Split raw segment into different length patches
- Feature Cross-Attention → long-term cls token + short-term patch





LST-EMG Result [5]2023

- Performance : LSTEMG > TEMG
- Inference time : LSTEMG > TEMG

Dataset	Model name	Accuracy	Inference time
DB2 exercise B	MSCNN	71.89%	5.60 ms
	BiTCN	65.79%	5.75 ms
	TEMG	78.77%	1.09 ms
	LSTEMGNet [ours]	**81.47%**	6.47 ms
DB5 exercise C	MSCNN	79.14%	7.27 ms
	BiTCN	83.75%	7.29 ms
	TEMG	68.18%	1.18 ms
	LSTEMGNet [ours]	**88.24%**	6.36 ms

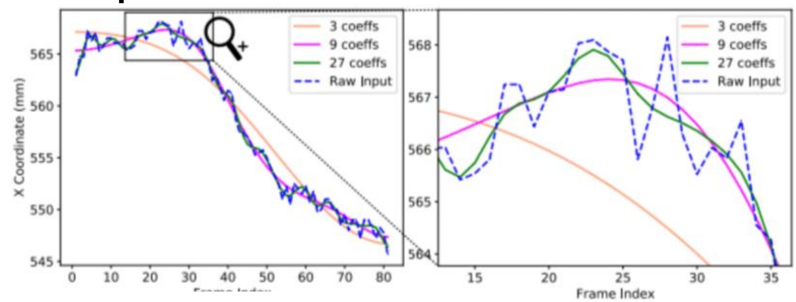
NinaPro DB2 DB5 - A public available multimodal database for machine learning research on human, robotic & prosthetic hands.

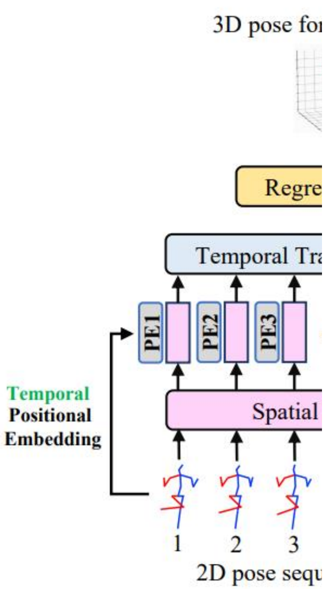
Future Work

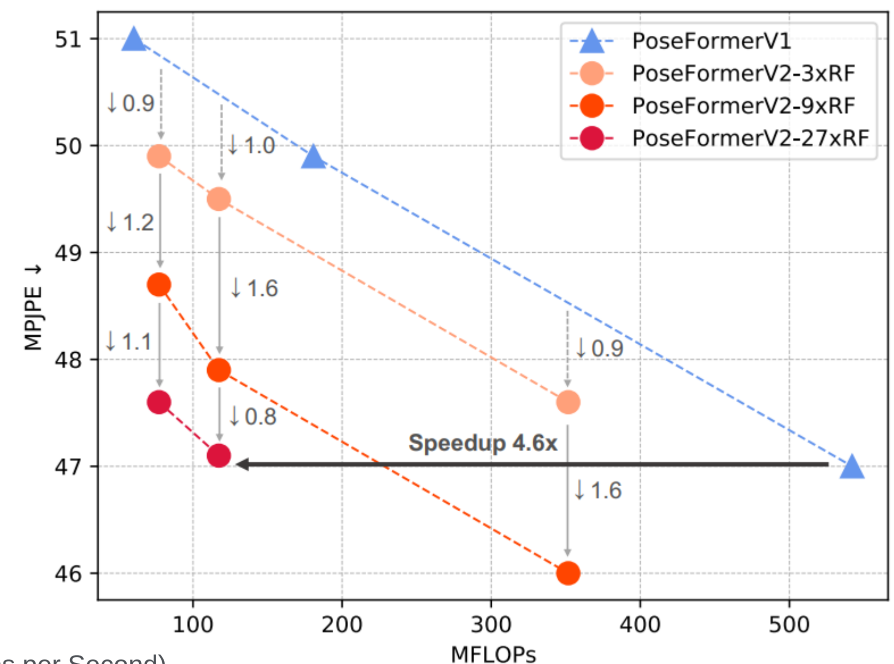
- Do some research on neutral network code(Pytorch)
- Replicate some transformer method on open-source data (NinaPro)

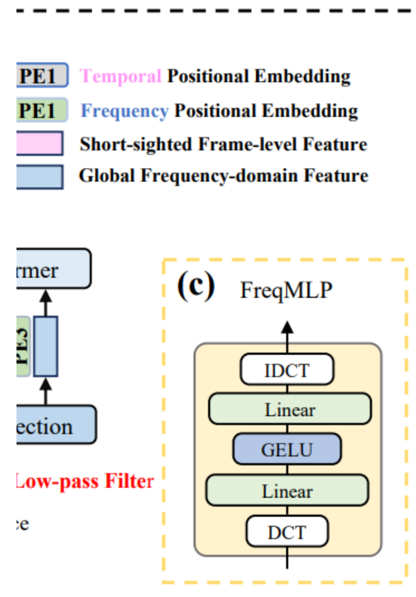
Complexity Reduction[4]2023

- Tradeoff between performance drop and computational cost
  - More sequence → high accuracy
  - Less detail → avoid high frequency noise









Original p

r v2

MPJPE (Mean Per Joint Position Error)
MFLOPs (Million Floating-point Operations per Second)

Reference – Overview

[1] S. Shen, X. Wang, F. Mao, L. Sun and M. Gu, "Movements Classification Through sEMG With Convolutional Vision Transformer and Stacking Ensemble Learning," in *IEEE Sensors Journal*, vol. 22, no. 13, pp. 13318-13325, 1 July1, 2022, doi: 10.1109/JSEN.2022.3179535.

[2] Rahimian, E.et al.Temgnet: Deep transformer-based decoding of upperlimb semg for hand gestures recognition.arXiv pre-printarXiv:2109.12379(2021).16.Toledo-Peral, C. L.et al.semg

[3] Montazerin, M., Rahimian, E., Naderkhani, F. et al. Transformer-based hand gesture recognition from instantaneous to fused neural decomposition of high-density EMG signals. Sci Rep 13, 11000 (2023). https://doi.org/10.1038/s41598-023-36490-w

[4] PoseFormerV2: Exploring Frequency Domain for Efficient and Robust 3D Human Pose Estimation

Qitao Zhao, Ce Zheng, Mengyuan Liu, Pichao Wang, Chen Chen.In CVPR 2023

[5] Zhang W, Zhao T, Zhang J, Wang Y. LST-EMG-Net: Long short-term transformer feature fusion network for sEMG gesture recognition. Front Neurorobot. 2023 Feb 28;17:1127338. doi: 10.3389/fnbot.2023.1127338. PMID: 36925629; PMCID: PMC10011454.

[6] Côté-Allard, Ulysse, Gabriel Gagnon-Turcotte, François Laviolette, and Benoit Gosselin. 2019. "A Low-Cost, Wireless, 3-D-Printed Custom Armband for sEMG Hand Gesture Recognition" *Sensors* 19, no. 12: 2811. https://doi.org/10.3390/s19122811

[7] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli and W. Geng, "Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 10, pp. 2964-2973, Oct. 2019, doi: 10.1109/TBME.2019.2899222.