2024 IEEE International Conference on Acoustics, Speech and Signal Processing

Inducing Inductive Bias in Vision Transformer for EEG Classification

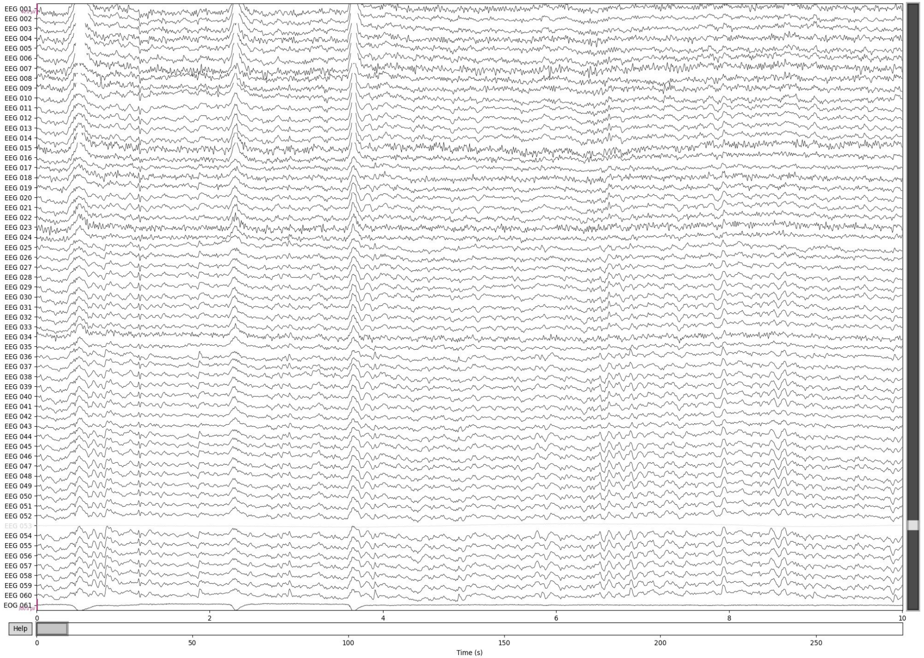
Rabindra Khadka, Pedro Lind, Gustavo Mello, Michael Riegler, Anis Yazidi

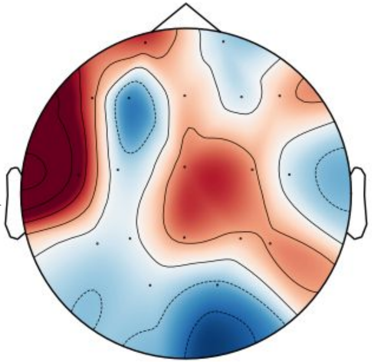
rabindra@oslomet.no

Overview

- EEG and its characteristics
- Vision Transformers and Inducing inductive bias
- Experimental Results of Brain Signal Vision Transformer (BSVT)

Context





- ❖ EEG provides direct measure of electrical brain activity
- ❖ It is characterized by low spatial (on the order of cm) and high temporal resolutions (on the order of ms)

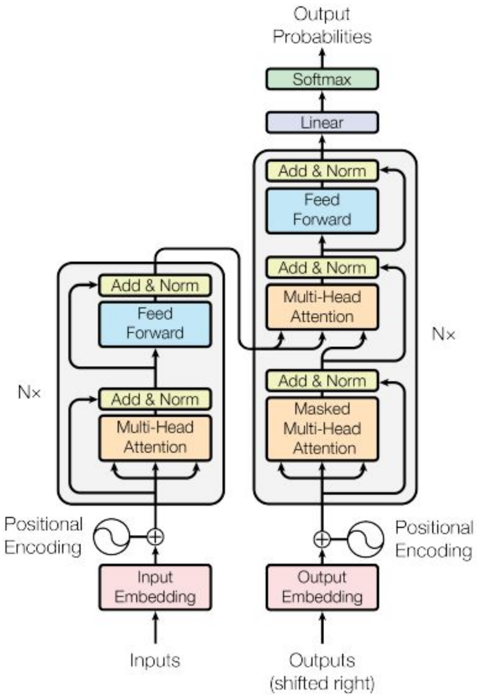
Motivation

Transformers lack inductive bias inherent to CNN .

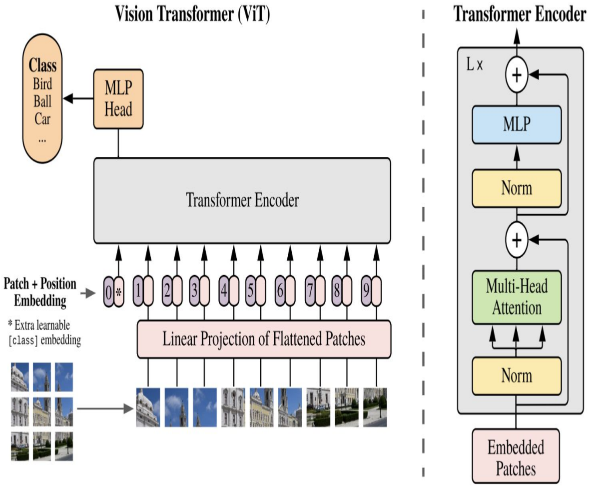
Transformer focus on global context through attention mechanism.

Is it possible to inject inductive bias to ViT models such that it is light and incorporate some prior knowledge related to the characteristics of EEG data?

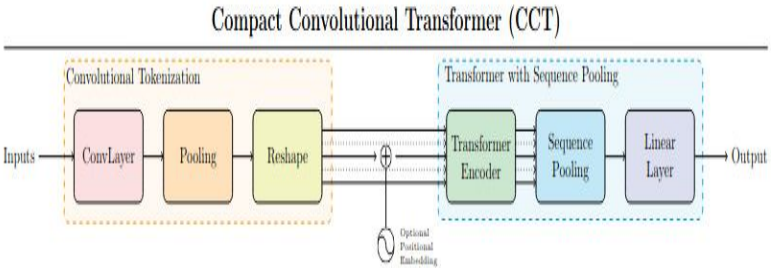
Related Works



Attention is All you Need
(Vaswani et al. 2017)

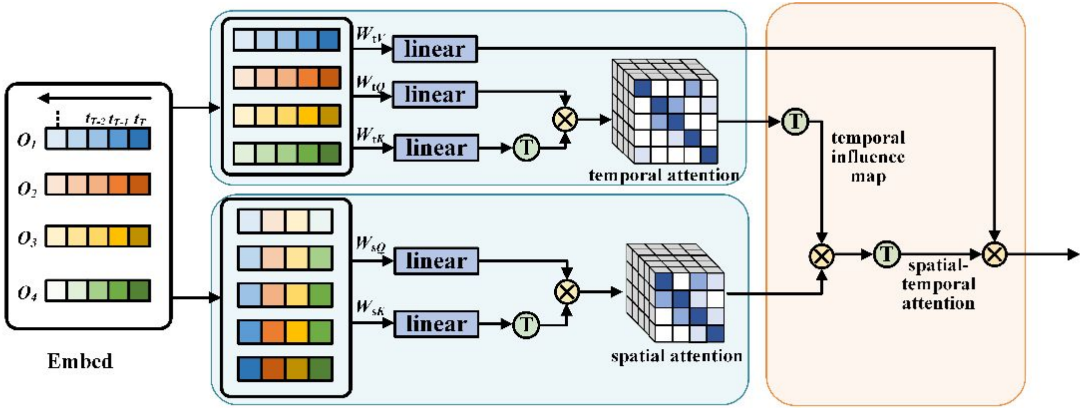


An image is worth 16 x16 words: Transformers for image recognition at scale
(Dosovitskiy, et al. 2020)

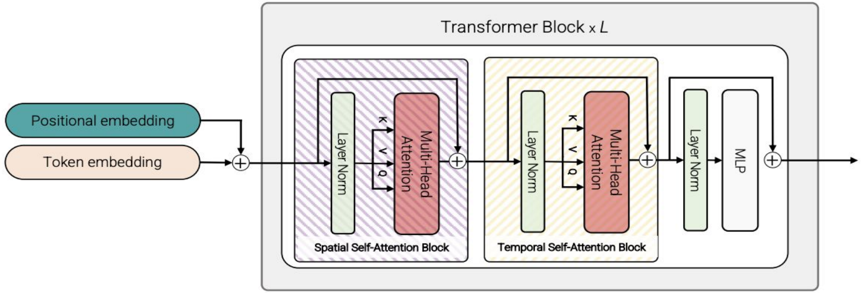


Escaping the Big Data Paradigm with Compact Transformers
(Hassani et al. 2022)

Related Works

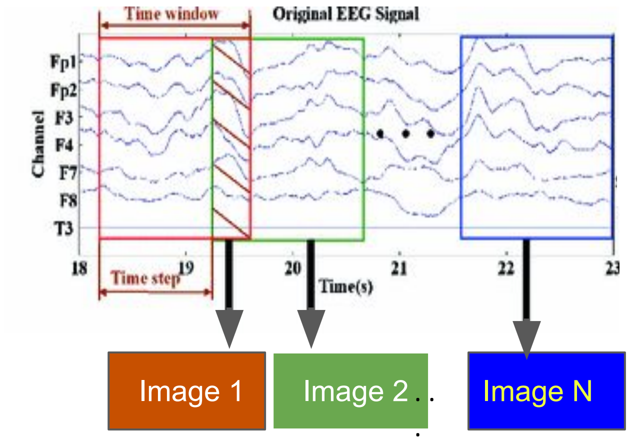


NAST: Non-Autoregressive Spatial-Temporal Transformer for Time Series Forecasting (Chen et al. 2021)



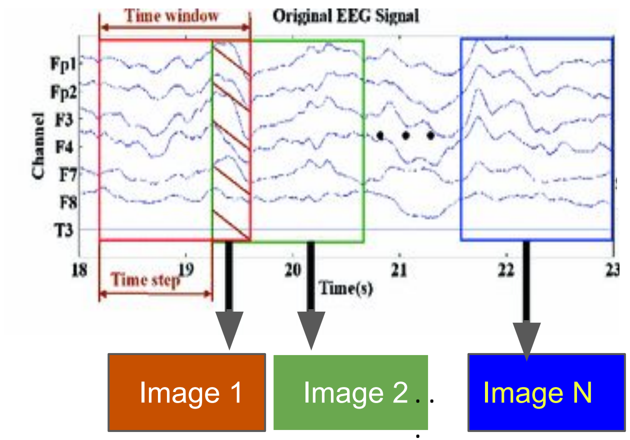
ViViT: A Video Vision Transformer (Hassani et al. 2022)

Patchifying and Embeddings

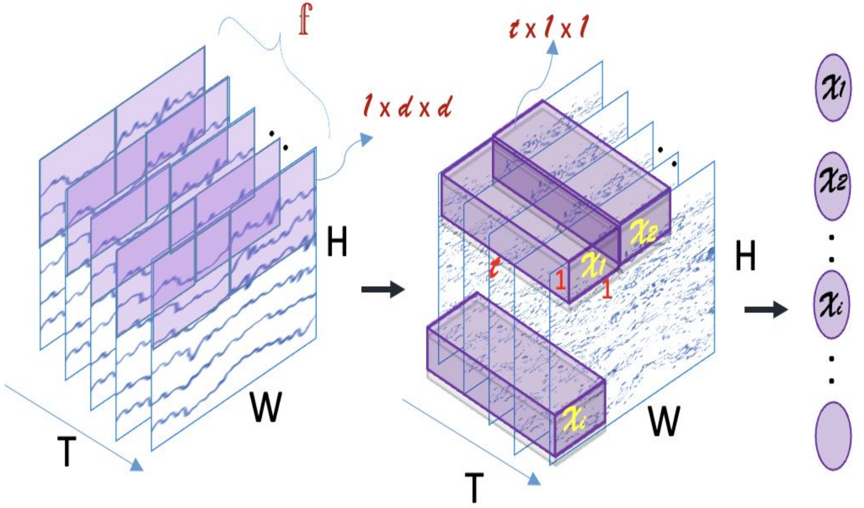


# of frames x channels x timesteps

Patchifying and Embeddings



# of frames x channels x timesteps

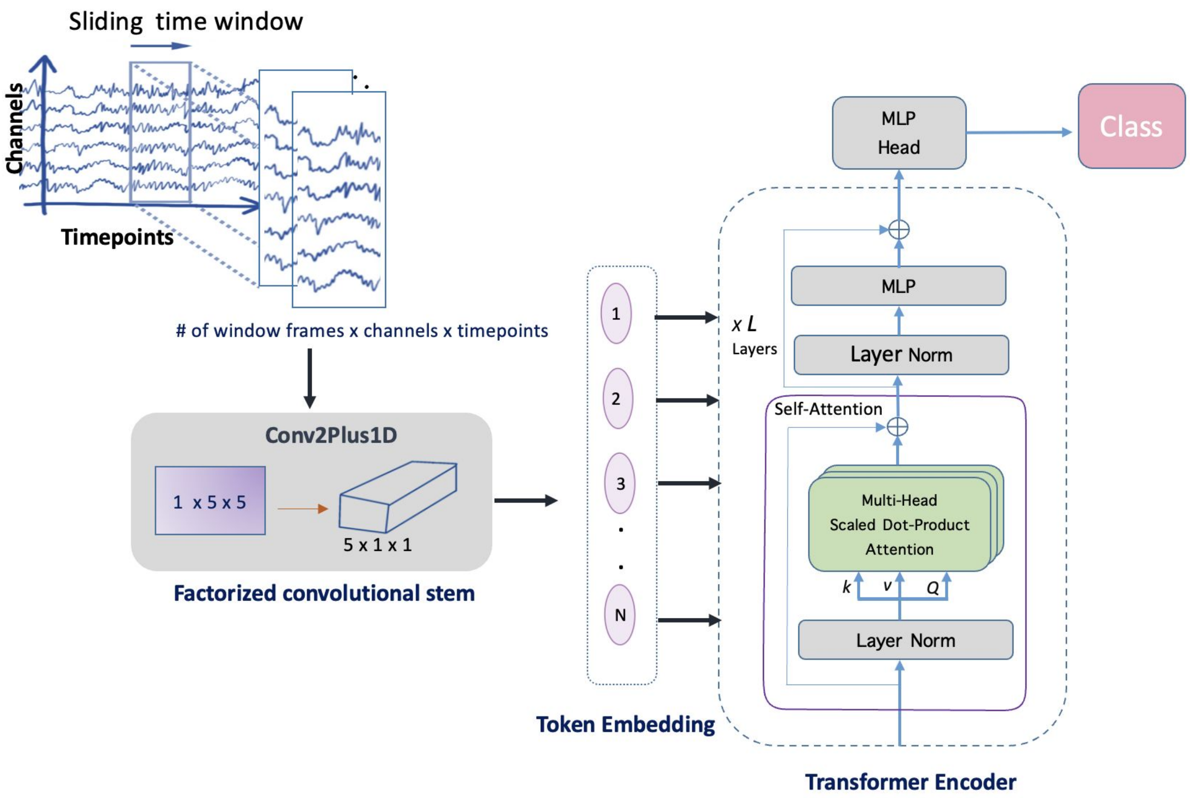


$\mathcal{X} \in \mathbb{R}^{T \times H \times W}$

$\tilde{\mathcal{X}} \in \mathbb{R}^{p_t \times p_h \times p_w \times d}$

- Non-overlapping spatio-temporal embeddings
- Additional non-linearity

BSVT Architecture



Ref: Tran et al . 2018.

Dataset

TUH Abnormal EEG Corpus (v2.0.0)	Non-pathological		Pathological	
	Recordings	Patients	Recordings	Patients
Development set	1371	1237	1346	893
Final evaluation set	150	148	126	105
Total	1521	1385	1472	998

Ref: https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml

Experiments

PERFORMANCE COMPARISON ON ABNORMAL EEG DATA.

Model	Accuracy	F1-score	# Params	MACs
ViT [4]	69.17	66.40	35.70 M	0.25 G
CCT [10]	73.46	68.20	0.21 M	0.09 G
BSVT(ours)	82.67	77.32	0.75 M	1.34 G

Ablation Study

ViT´s stem	F1-score	Accuracy
2D-convolution stem (3x3)	65.34	72.60
Fact. conv. stem(ours)	67.20	75.42
Fact. conv. stem with ReLU(ours)	77.32	82.67

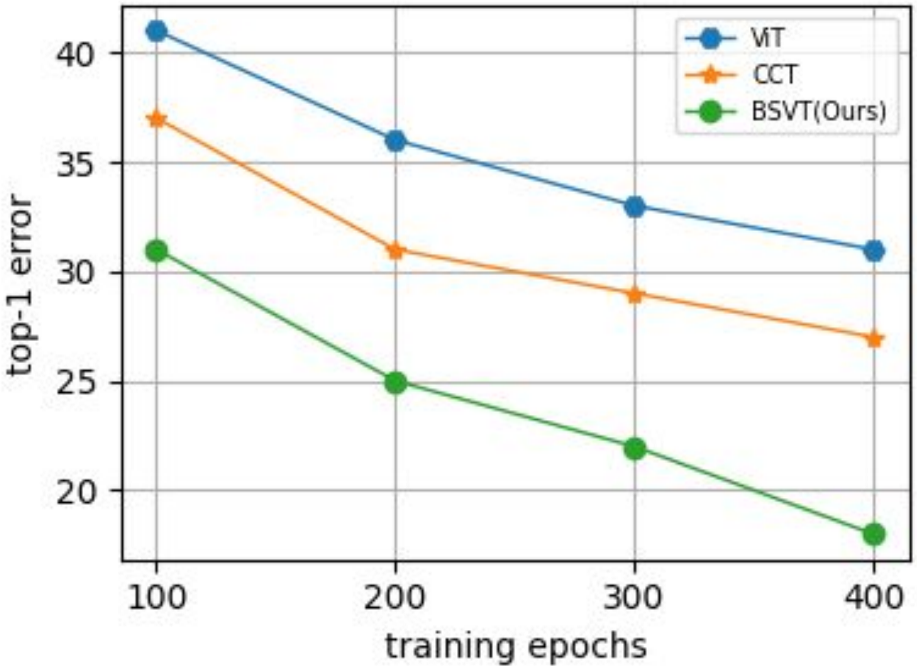
Ablation Study

ViT´s stem	F1-score	Accuracy
2D-convolution stem (3x3)	65.34	72.60
Fact. conv. stem(ours)	67.20	75.42
Fact. conv. stem with ReLU(ours)	77.32	82.67

ABLATION STUDY OF FILTER SIZES

Fact. conv. stem (2+1)D filter sizes	F1-score	Accuracy
BSVT - $(1 \times 5 \times 7), (9 \times 1 \times 1)$	76.14 ± 0.57	79.428 ± 0.53
BSVT - $(1 \times 7 \times 10), (12 \times 1 \times 1)$	75.57 ± 0.27	78.826 ± 0.64

Convergence behaviour



Conclusions

- **Leveraging Inductive Bias:** By integrating inductive bias, drawn from an understanding of EEG data characteristics (temporal and spatial attributes), we have developed a streamlined Vision Transformer (ViT) model. This approach not only reduces the model's parameter count but also decreases computational complexity, measured in FLOPs.

- **BSVT Model Advantages:** Our BSVT (Brain Signal Vision Transformer) model outpaces existing transformer-based benchmarks. This efficiency stems from the innovative use of factorized convolution, which fuses spatial and temporal features into embeddings. Furthermore, this method introduces enhanced non-linearity and elevates model complexity, without the penalty of additional parameters.

Future Work

- ➢ Going beyond EEG signal classification and delving into more intricate tasks
- ➢ Employing the model in the context of low data regime under self-supervised setting or transfer learning
- ➢ Analysis of learned visual concepts from EEG signals and mapping to EEG biomarkers