Enhancing EEG-based Gaze Prediction with Transformers on EEGEyeNet

by Aniket Konkar

B.E. in Information Technology, October 2020, University of Mumbai

A Thesis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Master of Science

May 18, 2025

Thesis directed by

Xiaodong Qu
Assistant Professor of Computer Science

Acknowledgements

I would like to express my sincere gratitude to my advisor, Prof. Qu, for their invaluable guidance, support, and encouragement throughout this thesis. Without their insightful advice and unwavering mentorship, the completion of this work would not have been possible.

Abstract of Thesis

Enhancing EEG-based Gaze Prediction with Transformers on EEGEyeNet

Electroencephalography (EEG) provides a non-invasive method for capturing brain activity, but its complex, high-dimensional nature poses significant challenges for accurate gaze estimation. Deep learning, particularly Vision Transformer (ViT) architectures, offers promising capabilities for gaze prediction using Electroencephalography (EEG) signals. We build upon the EEGViT model, adapting it for EEG-based regression tasks by optimizing its architecture and training process to improve gaze position predictions. Our experiments were conducted on the Absolute Position Task from the EEGEyeNet dataset, which aims to predict the spatial coordinates of a subject's gaze based solely on EEG data. The study investigates the impact of architectural modifications, including kernel size adjustments and dropout rate tuning, leading to an RMSE of 50.77 mm—improving model performance by ~2% compared to the original version of EEGViT-TCN. These results position the optimized model within 3% of the state-of-the-art accuracy. Additionally, we examine the potential effects of pretraining across multiple cognitive tasks, evaluating its impact on the accuracy of gaze predictions.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background

Electroencephalography (EEG) is a widely used non-invasive neuroimaging technique that records the electrical activity generated within the brain through electrodes placed on the scalp. EEG data, characterized by its complex and multidimensional nature, contains extensive information about brain activity, enabling a deeper understanding of various neurological and cognitive phenomena [1].

One emerging application of EEG is in gaze prediction, wherein the analysis of neural signals linked to visual attention is used to infer the direction of a user's gaze [15]. Advances in machine learning and signal processing have further enhanced the capability of EEG-based systems to decode these neural patterns. However, the complex nature of EEG data presents significant challenges in developing predictive models that are both efficient and accurate, particularly given the high cost and effort associated with data collection [6]. For instance, the baseline results on the EEGEyeNet dataset [3], which focus on predicting a subject's gaze position from EEG signals, clearly highlight the limitations of traditional machine learning models in accurately interpreting and modeling the data.

In the past decade, the development and application of deep learning models have grown rapidly and have been able to address the inherent complexity of EEG data and extract meaningful patterns. Deep learning techniques excel at approximating non-linear functions and can learn complex relationships within high-dimensional data, which make them well-suited for EEG analysis. Particularly, Convolutional Neural Networks (CNNs) and the Self-Attention mechanism [23, 36] have shown strong potential in effectively handling EEG signal intricacies [1, 31], with many architectures adapted from successful models in the field of Computer Vision [4, 21, 38].

This study builds upon the EEGViT model [4], which adapts a hybrid Vision Transformer (ViT) [5] originally pretrained on the ImageNet dataset [6] for EEG-based regression tasks. The model repurposes architectural components and pretrained weights from the vision domain to effectively handle the high dimensionality and complexity of EEG signals.

1.2 Dataset

Research by [14] has demonstrated the feasibility of extracting position-specific information from ocular artifact components of EEG, thereby contributing to EEG-based virtual eye tracking. The EEGEyeNet dataset, introduced by Kastrati et al. (Kastrati et al., 2021), is a comprehensive multimodal resource that simultaneously records electroencephalography (EEG) and eye-tracking (ET) signals across 356 healthy adult participants, amounting to over 47 hours of data. It serves as a benchmark for evaluating gaze prediction methodologies, addressing both the methodological challenges and practical implications inherent in linking brain activity with eye movement data (Kastrati et al., 2021).

The EEGEyeNet dataset is accompanied by two levels of preprocessing—minimal and maximal—applied using the openly available toolbox by (Pedroni et al.). Minimal preprocessing includes basic filtering and correction of bad electrodes while retaining ocular artifacts, which can aid in gaze estimation tasks [3]. Maximal preprocessing goes further by using Independent Component Analysis (ICA) combined with IClabel to remove a wide range of artifacts, including muscle, heart, and eye-related noise, aiming to preserve only neurophysiological signals [3]. EEG and eye-tracking data are synchronized using the EYE-EEG toolbox, ensuring precise alignment with synchronization errors kept below 2 ms. These preprocessing strategies help manage the heavy contamination in EEG data from both environmental and physiological sources [3].

The absolute position task within the EEGEyeNet benchmark is designed to predict the precise spatial coordinates of a subject's gaze from EEG signals alone. EEG recordings were collected using a high-density 128-channel Geodesic Hydrocel system, recorded at a sampling rate of 500 Hz with a central reference configuration. Eye positions were simultaneously recorded using the EyeLink 1000 Plus system at the same sampling rate. Participants were seated 68 cm from a 24-inch monitor, with head movement minimized using a chin rest for stabilization.

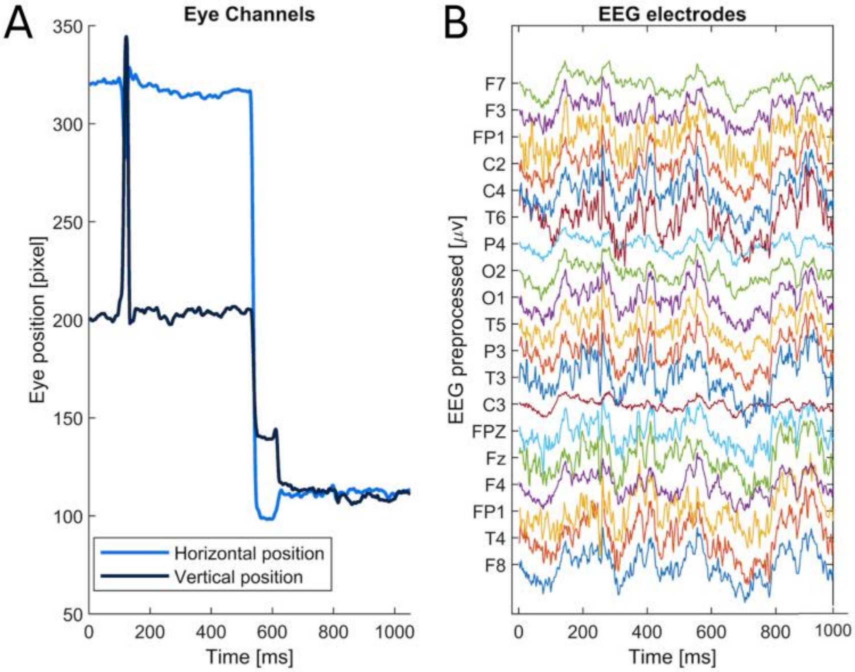

Figure 1.1: Visualization of a one-second EEG sample [3] - 500×129 dimensions with corresponding gaze data (A) and a subset of preprocessed EEG channels (B)

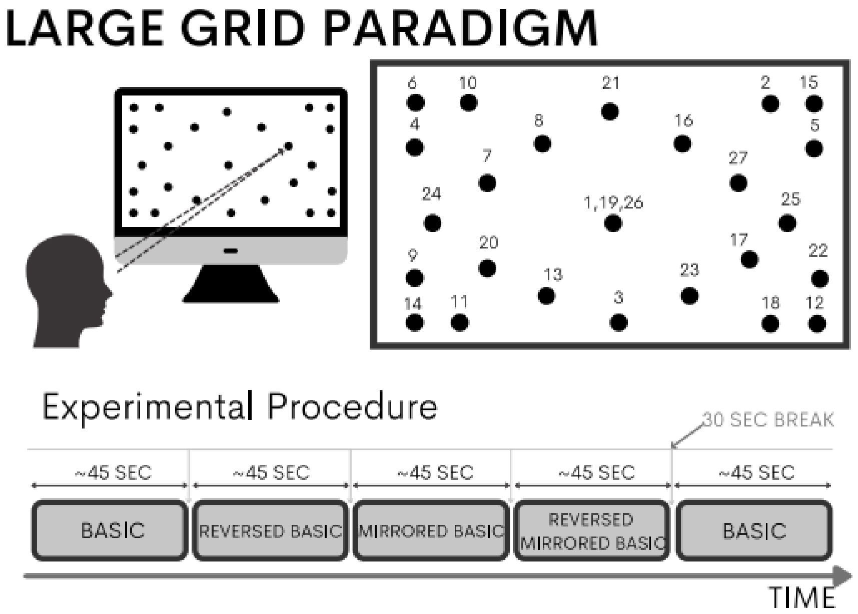

Figure 1.2: Large Grid Paradigm from EEGEyeNet [3]. Participants were instructed to fixate on specific dots during designated time intervals.

For determining the absolute gaze position, the data from the Large Grid Paradigm is used to benchmark proposed model variations. In this task, the goal is to estimate the subject's gaze location on the screen in terms of XY-coordinates, using only EEG data. Each data sample represents a one-second EEG segment during which the participant maintains a single fixation, ensuring clean, non-overlapping neural activity associated with visual attention. By formulating the problem as a regression task and evaluating performance based on the Euclidean distance between predicted and true gaze positions, this setup offers a robust framework for developing and testing deep learning models aimed at simulating a purely EEG-driven eye-tracking system.

1.3 Contributions

For this research, multiple experiments were conducted, where different architectures and techniques were combined and evaluated with the goal of improving gaze estimation performance on the Absolute Position Task from the EEGEyeNet Dataset. To summarize, this research produces the following findings:

- Modifying the kernel size and dropout rate in EEGViT-TCNet leading to achieving an RMSE of 50.77 mm - 2% better accuracy than the original EEGViT-TCNet. This result is comparable to state-of-the-art (within 3% or RMSE of 1.88 mm).
- Tested whether pretraining across the cognitive tasks (posing Task 2 as a Task 3 problem) from this dataset improve accuracy?

Chapter 2: Related Work

The original EEGEyeNet [3] dataset & benchmark paper provided baseline results for absolute gaze position estimation that showed classical machine learning models performed poorly, with results close to the naive baseline. In contrast, deep learning models achieved significantly better performance, with average Euclidean errors ranging from 70 to 80 mm. The benchmark established a baseline of 70.2 mm using a CNN, one of the simpler deep learning models. Although these results were not ideal, the improvement over traditional methods demonstrated the potential of EEG-based eye tracking. The authors suggested that incorporating temporal information through sequence models such as RNNs or Transformers [2] could further enhance performance. Overall, the evaluation confirmed that deep learning models were more effective for this task and highlighted EEGEyeNet as a valuable resource for advancing EEG-based gaze estimation.

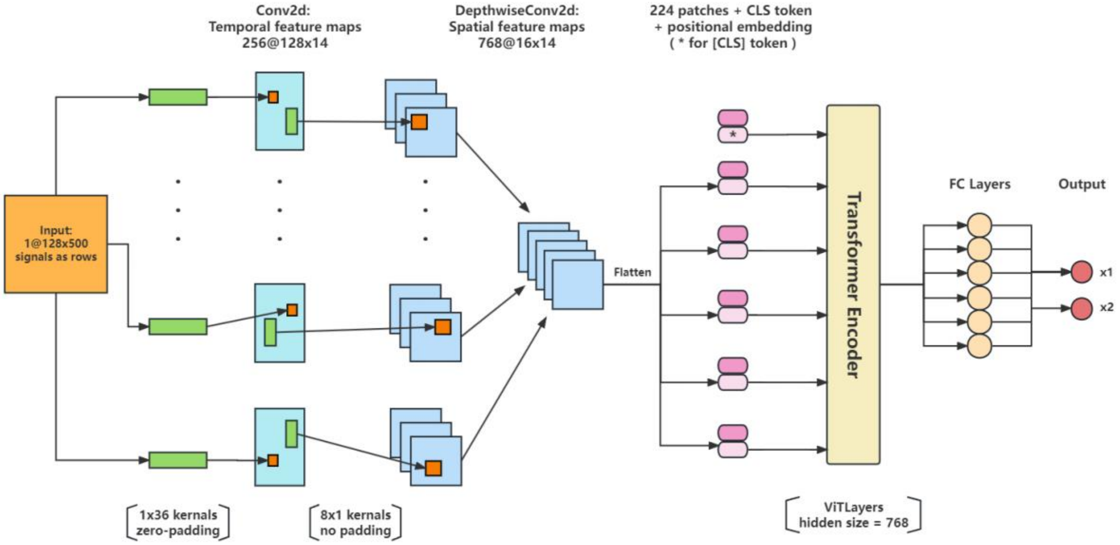

Figure 2.1: EEGViT [4] – specifically for EEG raw signal as input.

This pretrained EEGViT model [4] advanced the idea of adapting computer vision models for EEG tasks by proposing the use of a hybrid Vision Transformer (ViT), pretrained on the ImageNet dataset, for EEG regression. This approach repurposed a model initially designed for vision tasks to tackle the high-dimensional and complex nature of EEG data, drawing on the similarity between the structures of image and EEG data. The results demonstrated that the pretrained ViT outperformed previous deep learning models with an RMSE of 55.4 mm, with its success attributed not only to model architecture but also to the benefits of pretraining on large, readily available image datasets. This approach highlights the potential of leveraging pretraining across disciplines, especially in EEG analysis, where data collection is often challenging due to practical, financial, and ethical issues.

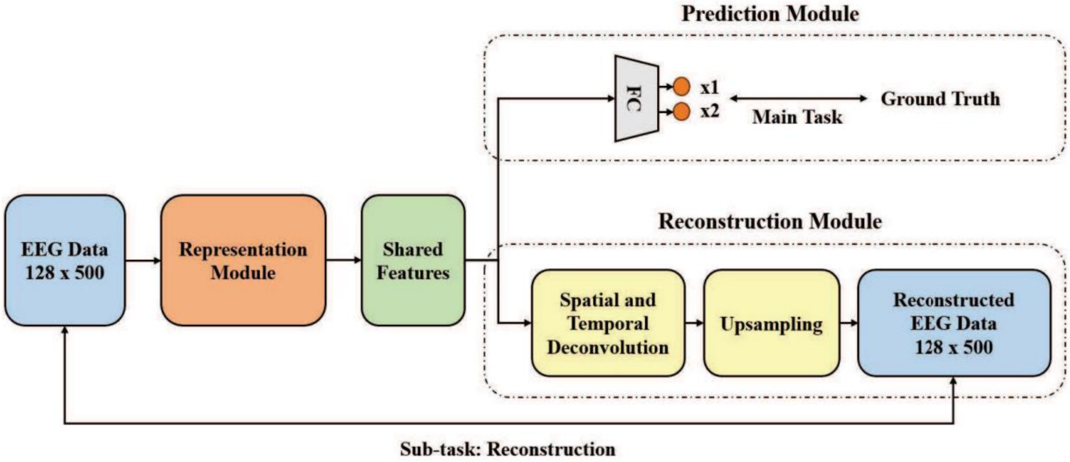

Figure 2.2: MTL-Transformer Architecture: Eye Tracking and Data Reconstructions [9]

This study [9] introduced a novel EEG signal reconstruction sub-module designed to integrate with any Encoder-Classifier-based model. The sub-module supported end-to-end training within a multi-task learning framework and operated under an unsupervised learning paradigm, enhancing its applicability across various tasks. The researchers incorporated this sub-module into advanced models, including Transformers and pretrained Transformers, and observed an RMSE of 54.1 mm. The success of this technique demonstrated the sub-module's effectiveness in enhancing encoder capabilities and highlighted its potential as a new paradigm for boosting deep learning performance in EEG-related applications.

A lightweight EEG regression model that combined a pre-trained MobileViT with Knowledge Distillation (KD) achieved RMSE of 53.6 mm while being 33% faster and 60% smaller, demonstrating its suitability for real-time, resource-constrained BCI applications [10].

This study [11] introduced an EEG-based gaze prediction algorithm and improved the RMSE to 53.06 mm compared to the previous state-of-the-art, while also reducing training time by over 67%, demonstrating both enhanced accuracy and efficiency.

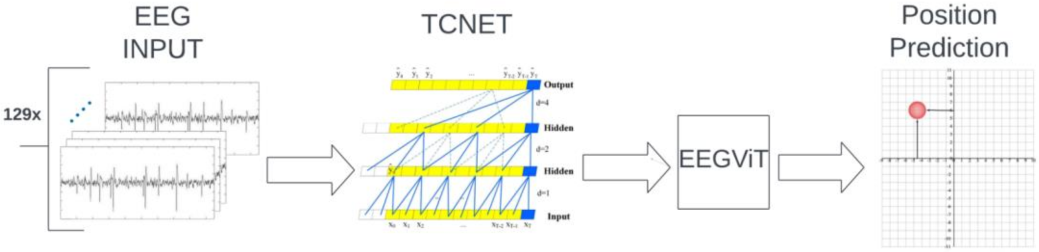

Figure 2.3: EEGViT-TCNet Model Diagram [12]

(Modesitt et al., 2024) integrated pretrained Vision Transformers (ViTs) with Temporal Convolutional Networks (TCNet). By leveraging ViTs for sequential modeling and TCNet for feature extraction, the approach achieved an RMSE of 51.8 mm [12]. It also achieved a speedup of up to 4.32x without compromising accuracy, setting a new benchmark and highlighting the potential of combining transformer architectures with specialized convolutional methods for EEG analysis.

This study [13] proposed EEG-DCViT, a model combining depthwise separable CNNs with Vision Transformers, enhanced by a clustering-based pre-processing strategy. The approach achieved an RMSE of 51.6 mm, highlighting the importance of both pre-processing and architectural refinement in improving EEG-based model performance.

Outlier Removal

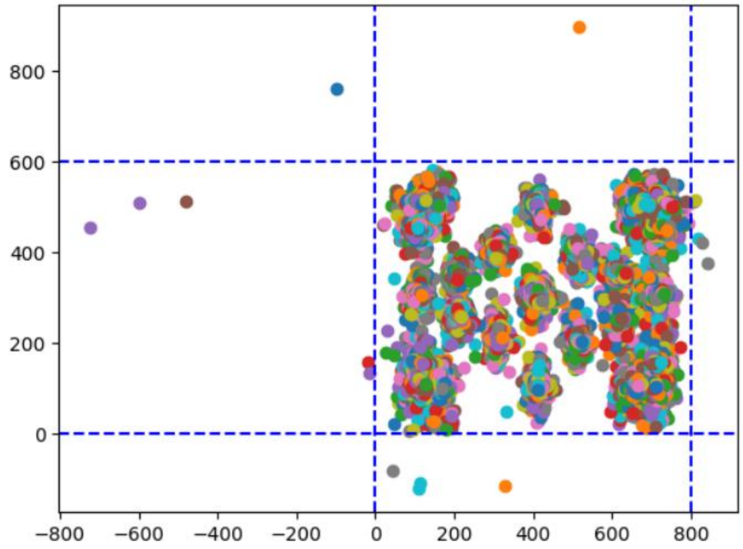

Figure 2.4: Visualization of the EEGEyeNet Absolute Position data

A recent approach [7] demonstrated that training the EEGViT-TCNet model on a pruned dataset—excluding 15 outlier samples with gaze coordinates outside the 800 × 600 screen resolution—achieved an RMSE of 48.9 mm, marking it as the current state-of-the-art (SOTA), to the best of our knowledge. While we were unable to replicate the same performance under identical setup conditions, the outlier removal did lead to slight improvements. Consequently, this filtering step was adopted in our experimental pipeline.

Chapter 3: Methodology

3.1 Modifying Kernel Size and Testing Dropout rate

The decision to adjust and optimize the kernel size within the EEGViT-TCNet architecture was motivated by the findings of [11], which highlighted the potential for modest accuracy improvements. After evaluation, it is confirmed that the accuracy benefits observed in [11] successfully transferred to the EEGViT-TCNet model achieving an RMSE of 50.77 mm (over the default version of EEGViT-TCNet, which was 51.8 mm)

In the first convolutional layer, we applied a 1 × 16 kernel to the 1-second EEG input of size 129 × 500, which was zero-padded to 129 × 512. These kernels act as band-pass filters on the raw signals. Our selected kernel size of 1 × 16 is smaller than that used in the EEGViT model(1 × 36). This was motivated by the kernel size used in [11], providing finer temporal resolution for feature extraction. Further reduction to a 1 × 4, 1 x 8 kernel did not yield accuracy improvements, suggesting that 1 × 16 is an optimal choice for both kernel size and stride in our configuration. Batch normalization was applied to the resulting 129 × 32 output. The second layer employed a depth-wise convolution with a 129 × 1 kernel, scanning across all EEG channels per temporal filter. Initially dropout of 70% produced the best results, however we were not able to replicate it for multiple runs, leading us to believe that A dropout rate of 70-75% was found to produce the most optmial results. The image size for the Vision Transformer was set to (1, 32). The remaining architectural components followed the EEGViT-TCNet default settings, and outlier removal (15 samples) was implemented as recommended by [7].

3.2 Exploring Pretraining

To investigate the performance of the EEGViT-pretrained model on the Absolute Position prediction task (Task 3) of the EEGEyeNet dataset, we adopted a two-phase approach involving pretraining on Task 2 and then fine-tuning on task 3. The theoretical motivation to believe this would produce comparable results to the previous models was because Task 2 and Task 3 while have the same underlying data, they have different data preparation strategy and as a result have completely different data used for benchmarking (Task 2 has saccades onset from middle ensures there's atleast one saccade in a sample as opposed to the data from Task 3 that only has fixation samples after pre-processing). Another motivation to try this approach was that the original EEGEyeNet [3] paper stated that Task 2 strives to serve as a transitional step for the development of a fully EEG-based ET.

3.3 Phase 1: Pretraining on Angle/Amplitude Prediction (Task 2)

We first pretrained the *EEGViT-pretrained* model on Task 2: Angle/Amplitude of the EEGEyeNet Dataset, which was adapted to follow the same output format as Task 3 to facilitate transferability. Task 2 consists of EEG segments centered around saccade onsets, ensuring that each sample includes at least one saccade event. This provides more informative signal patterns for learning eye movement dynamics. We converted Angle and Amplitude output predictions from task 2's dataset to x, y coordinates using the Polar to Cartesian coordinate conversion formula. The model was trained for 15 epochs, with performance monitored using Root Mean Square Error (RMSE) as the evaluation metric. This RMSE was not further divided by 2 (as is the case for task 3 for interpretation) this was just the pretraining phase. RMSE with the best validation loss was selected. RMSE of 73.2686 was observed.

3.4 Phase 2: Fine-tuning on Absolute Position Prediction (Task 3)

Following pretraining, the EEGViT model was fine-tuned on Task 3, which comprises fixation samples derived using a different data preparation strategy. Unlike Task 2, which anchors saccades centrally within the EEG segments, Task 3 includes EEG windows corresponding to stable fixation events, making the dataset distributionally distinct despite sharing the same underlying data recordings. This setup tests the model's generalization ability across these two different data preparation constraints. The model was again trained for 15 epochs with all the same settings except that the ViT encoder weights were loaded from the previously pretrained model, however, they were not freezed during the training. The results were underwhelming – an RMSE of 56.65 (printed as 113.3103 in the code, but we divide by 2 for better interpretation) was the best observed RMSE in this training(Observed on Epoch 6). Furthermore, the RMSE kept increasing for the next epochs. Hence, we can confirm that pretraining on task 2 by posing it as a task 3 problem did not generalize to task 3.

3.5 Data Source & Split

The EEG data used for model training were drawn from the "Large Grid Paradigm" section of the EEGEyeNet dataset, in which participants fixated on 25 distinct screen positions [3]. The minimally pre-processed version of the dataset was utilized for all experiments. This dataset includes recordings from 27 participants, amounting to 21,464 samples. The EEGEyeNet dataset [3] was split by participant ID to prevent data leakage, ensuring that samples from the same individual did not appear across training, validation, and testing sets. The data were then divided into 70% for training, 15% for validation, and 15% for testing – a split ratio followed as per the original paper [3].

3.6 Training Configuration

Each model except the pretraining on Task 2 was trained on a single A100 40 GB GPU in Google Colab environment, with each epoch taking approximately one to two minutes (only on the A100 also depending on the selected model configuration) to process all 64 sample batches. The pretraining on task 2 was trained on an RTX 3070. Each model was trained for 15 Epochs. The initial learning rate of 1e-4 was reduced by a factor of 0.9 every 6 epochs. Adam Optimizer was used.

Chapter 4: Results

4.1 Evaluation Metrics

Similar to the approach taken in [3] and other related studies, all models in this work were trained using the Mean Squared Error (MSE) loss function to guide learning, while performance was evaluated using the Root Mean Squared Error (RMSE) metric.

$$MSE = {}^{1}\!/_{n} \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

$$RMSE = \sqrt{MSE}$$

The original error values, measured in pixels, were converted to millimeters using a scale of 2 pixels per millimeter for improved interpretability. Hence, the Absolute Position RMSE in mm is calculated as RMSE divided by 2.

4.2 Benchmarks

Lower RMSE values correspond to more accurate predictions, reflecting closer alignment with the ground truth.

Model	Abs. Position RMSE (mm)
Naïve Baseline	123.3
KNN	119.7 ± 0
RBF SVC/SVR	123 ± 0
Linear Regression	118.3 ± 0
Ridge Regression	118.2 ± 0
Lasso Regression	118 ± 0
Elastic Net	118.1 ± 0
Random Forest	116.7 ± 0.1
Gradient Boost	117 ± 0.1
AdaBoost	119.4 ± 0.1
XGBoost	118 ± 0

Table 4.1: Results of Machine Learning models from the EEGEyeNet Benchmark [3]

Deep Learning models from [3] outperformed the naive baseline, but still left room for improvement, with the best – CNN, achieving an average RMSE of 70.4 mm.

EEGViT model architecture was a breakthrough and drastically reduced the RMSE from previous 70.4 mm to 55.4. The Two-Step convolution representations introduced in this model likely enhanced the effectiveness of the attention mechanism within the transformer architecture. Many subsequent architectures and techniques, including our own, are derived from the pretrained EEGViT model.

Model	Abs. Position RMSE (mm)	Study
Naïve Baseline	123.3	[3]
CNN	70.4 ±1.1	[3]
PyramidalCNN	73.9 ±1.9	[3]
EEGNet	81.3 ±1.0	[3]
InceptionTime	70.7 ±0.8	[3]
Xception	78.7 ±1.6	[3]
EEGViT Pretrained	55.4 ±0.2	[125]
MTL-Transformer	54.1 ±0.2	[9]
EEGMobile	53.6 ±0.6	[10]
Qiu, C. et al (2024)	53.06 ±	[11]
EEGViT-TCNet	51.8 ±2	[12]
EEG-DCViT	51.6 ±	[13]
This Study	**50.77 ±0.41**	-
Wu, J et al (2025)	48.9 ±	[7]

Table 4.2: Deep Learning Architectures benchmarked on the Absolute Position Task

Our study made further modifications to the EEGViT-TCNet architecture to achieve an RMSE of 50.77 mm. As stated before, we were not able to replicate the study results by [7], however, we have added that in to this list and to the best of our knowledge, believe this is the SOTA reported performance.

4.3 Limitations

While the proposed method demonstrates gains in accuracy, it still results in an RMSE of around 5.07 cm. This performance remains inferior to that of commercial video-based eye-tracking systems, which typically deliver higher accuracy and faster response times. Moreover, the EEGEyeNet dataset was acquired under controlled laboratory conditions, with participants instructed to stay still and focus on fixed screen points—an environment that does not align with real-world usage. Additionally, EEG setups tend to be more intricate and less user-friendly compared to widely available video-based solutions.

The model's performance may not generalize well to other EEG datasets due to the inherently complex and variable nature of EEG data collection. Differences in hardware setups, electrode placements, recording protocols, and environmental conditions can significantly impact signal quality and characteristics. These variations make it challenging to transfer models trained on one dataset—such as EEGEyeNet—to another without substantial performance degradation or the need for additional fine-tuning and preprocessing adjustments.

Chapter 5: Discussion

EEG data is inherently noisy, which presents a significant challenge when developing models aimed at accurately predicting positional information. This noise stems from various sources, including muscle activity, eye movements, and external electrical interference, making it difficult for models to extract meaningful patterns directly from raw signals. As a result, robust preprocessing and noise removal techniques are crucial to enhance signal quality and improve model performance. However, implementing effective preprocessing steps often requires a level of domain expertise to identify and filter out relevant artifacts without losing critical neural information. This adds complexity to the pipeline, as both signal processing and neuroscientific understanding must be integrated to ensure the model can focus on true cognitive signals rather than background noise.

There is strong potential to adapt and train different vision transformer architectures, (one such example is to test the Swin Transformer [8]), for EEG-based tasks by leveraging their success in image-based applications. Since EEG data can be represented as time-frequency images or topographical maps, these image-like formats make it feasible to apply pretrained vision transformers. Using models like Swin (which have surpassed many benchmarks across various datasets), which are already trained on large-scale visual datasets, can provide a powerful starting point for feature extraction, enabling the model to capture spatial and temporal patterns in EEG data more effectively. Fine-tuning these pretrained architectures on EEG-specific tasks could significantly boost performance, especially when labeled data is limited.

Chapter 6: Conclusion

This work contributes toward improving EEG-based gaze estimation. Notably, by optimizing the kernel size in the EEGViT-TCNet architecture, the model achieved an RMSE of 50.77 mm—representing a 2% improvement over the original model and nearing state-of-the-art performance. Additionally, pretraining across multiple cognitive tasks was explored as a strategy to enhance accuracy. However, we can conclude that pretraining across tasks doesn’t translate well for EEG Data due to the inherent nature of data capture and other complexities, even if underlying data is captured under same settings for the two different tasks.

These findings suggest that while Transformer-based architectures and their variants hold strong potential for improving EEG-based gaze prediction, effective data preprocessing and noise removal are equally critical. By enhancing signal quality and reducing artifacts for minimally processed data, these steps can significantly boost model performance, when combined with advanced deep learning techniques.

Bibliography

[1] Michal Teplan. 2002. Fundamentals of EEG measurement. Measurement science review 2, 2 (2002), 1–11.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. Advances in neural information processing systems 30 (2017)

[3] Ard Kastrati, Martyna Beata Płomecka, Damián Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, and Nicolas Langer. 2021. EEGEyeNet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. arXiv preprint arXiv:2111.05100 (2021).

[4] Yang, R., & Modesitt, E. (2023). *ViT2EEG: Leveraging Hybrid Pretrained Vision Transformers for EEG Data*. In *Proceedings of the KDD Undergraduate Consortium (KDD-UC '23)*.

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. IEEE, 248–255.

[7] Wu, J., Dou, J., & Utoft, S. (2025). Refining Human-Data Interaction: Advanced Techniques for EEGEyeNet Dataset Precision. In *International Conference on Human-Computer Interaction* (pp. 407-419). Springer, Cham.

[8] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).

[9] Li, W., Zhou, N., & Qu, X. (2024, June). Enhancing eye-tracking performance through multi-task learning transformer. In *International Conference on Human-Computer Interaction* (pp. 31-46). Cham: Springer Nature Switzerland.

[10] Liang, T., & Damoah, A. (2025). EEGMobile: Enhancing Speed and Accuracy in EEG-Based Gaze Prediction with Advanced Mobile Architectures. In *International Conference on Human-Computer Interaction* (pp. 341-355). Springer, Cham.

[11] Qiu, C., Liang, B., & Key, M. L. (2024, June). Effect of Kernel Size on CNN-Vision-Transformer-Based Gaze Prediction Using Electroencephalography Data. In *International*

Conference on Human-Computer Interaction (pp. 60-71). Cham: Springer Nature Switzerland.

[12] Modesitt, E., Yin, H., Huang Wang, W., & Lu, B. (2024, June). Fusing Pretrained ViTs with TCNet for Enhanced EEG Regression. In *International Conference on Human-Computer Interaction* (pp. 47-59). Cham: Springer Nature Switzerland.

[13] Key, M. L., Mehtiyev, T., & Qu, X. Advancing EEG-Based Gaze Prediction: Depthwise Separable Convolution and Pre- Processing Enhancements in EEGViT.

[14] Sun, R., Chan, C., Hsiao, J., & Tang, A. (2020). Eeg artifact to signal: predicting horizontal gaze position from sobi-dans identified ocular artifact components.. https://doi.org/10.1101/2020.08.29.272187

[15] Bagheri, I., Alizadeh, S., khorasgani, M., & Asgharighajari, M. (2024). A systematic investigation based on bci and eeg implemented using machine learning algorithms. IJSETPUB, 1(4), 55-60. https://doi.org/10.63053/ijset.45