Improving Cross-Session Generalization in ECoG-Based Motor Decoding Using Self-Supervised Learning

Hugo Demule
Under the supervision of Yuhan Xie and Prof. Shoaran.
Integrated Neurotechnologies Laboratory, EPFL – Campus Biotech, Geneva, Switzerland

***Abstract*—Decoding motor behavior from electrocorticography (ECoG) signals is difficult due to strong variability across recording sessions and the limited availability of labeled data. The goal of this semester project was to improve a supervised baseline model by exploring self-supervised learning (SSL) approaches and alternative training losses. The baseline model is a transformer-based architecture that takes wavelet-transformed ECoG signals as input and predicts the wrist position of a monkey.**

TS-TCC [1] is applied as a self-supervised pretraining method, using contextual and temporal contrastive losses without relying on labels. After pretraining, only the encoder is retained and used for downstream regression tasks. The results show that the self-supervised pretrained model generalizes better to future recording sessions compared to the fully supervised baseline, achieving higher mean R^2 scores across recordings. Label fraction experiments further demonstrate that SSL allows the model to reach reasonable performance with as little as 1% of labeled data.

Additional losses based on soft temporal and instance-level contrastive learning were also evaluated [2]. While these losses showed promising behavior on smaller datasets, they tended to degrade performance when applied to large-scale pretraining, often smoothing predictions and reducing performance on well-performing sessions. Overall, this project highlights the potential of self-supervised learning to improve generalization and label efficiency for ECoG-based motor decoding.

I. Introduction

Electrocorticography (ECoG) signal decoding for motor behavior faces significant challenges: strong variability across recording sessions, limited labeled data availability, and poor generalization to future time periods. This work addresses wrist position regression from 64-channel ECoG recordings using self-supervised learning (SSL) to improve cross-session generalization. A self-supervised pre-training method named TS-TCC [1] is leveraged on five sessions, followed by supervised fine-tuning on a sixth session; generalization is then evaluated on temporally distant sessions. The goal is to demonstrate that self-supervised pretraining achieves superior generalization compared to purely supervised approaches.

II. Related Work

ECoG-based neural decoding faces significant generalization challenges across recording sessions [3]. Recent clinical demonstrations [4] highlight both promise and challenges of real-world neural decoding. Self-supervised learning approaches like BrainBERT [5] and TS-TCC [1] have shown promise for learning representations from unlabeled neural data. TS-TCC uses temporal and contextual contrasting: temporal contrasting predicts future representations from past context, while contextual contrasting maximizes agreement between augmented views of the same window. Contrastive learning methods from computer vision [6] have been successfully adapted for time-series [7]. Extensions like soft contrastive learning [2] introduce continuous similarity measures for more nuanced temporal relationships.

III. Methods

A. Dataset and Preprocessing

The dataset consists of 64-channel ECoG recordings from non-human primates performing reach-and-grasp tasks. The electrode array covers primary motor (M1) and somatosensory (S1) cortices. The regression target is the z-axis wrist position during reaching motions.

Recordings span multiple sessions across different days: D1–D5 (sessions from December 2-5 and 9, 2024) are used for self-supervised pretraining, D6 (December 16, 2024) for supervised downstream training, and 12 test sessions (December 2024 to February 2025) for generalization evaluation. Test sessions are never used during model training or validation.

Preprocessing transforms raw signals into time-frequency representations: signals are segmented into 500-sample (1 second) windows at 500 Hz, normalized and quantized to [0, 1023]. Continuous wavelet transform (CWT) extracts features across alpha (10–30 Hz), beta (30–60 Hz), and gamma (80–100 Hz) bands, yielding 5 frequency channels per electrode (320 features total). Sessions are split into 60% training, 20% validation, and 20% test.

B. Baseline Model

The supervised baseline developed at the Integrated Neurotechnologies Laboratory (INL) uses a transformer architecture: CWT features (500×320) are downsampled to 10 time steps, processed through linear embedding (320→32), positional encoding, 2-layer transformer with linear attention (2 attention heads, 128-dim feedforward layer), global average pooling, and linear regression head. Training uses MSE

loss, Adam optimizer (lr=3e-4), batch size 128, evaluated with R^2 score. The model suffers from poor cross-session generalization due to overfitting session-specific patterns and aggressive temporal downsampling that could potentially discard fine-grained dynamics.

C. Self-Supervised Learning Framework

TS-TCC follows a two-stage pipeline: self-supervised pretraining followed by downstream regression. The CNN encoder contains three 1D convolutional blocks (channels: $320 \to 32 \to 64 \to 128$) applied to full-length sequences $x \in \mathbb{R}^{B\times 320\times 500}$, preserving temporal structure while gradually reducing length through max pooling. The resulting representation is $z \in \mathbb{R}^{B\times 128\times 65}$, which is transposed to $z' \in \mathbb{R}^{B\times 65\times 128}$ and processed by a transformer encoder (hidden dimension 100). This yields contextualized timestep features $h \in \mathbb{R}^{B\times 65\times 100}$ and a global temporal embedding $c_t \in \mathbb{R}^{B\times 100}$. During SSL, linear predictors map c_t to future latent targets in $\mathbb{R}^{128}$ across 22 timesteps, enforcing temporal predictive consistency between two augmented views. For downstream tasks, the transformer and projection head are discarded, keeping only the pretrained encoder that produces z, which is flattened and fed to a linear head to predict the target value.

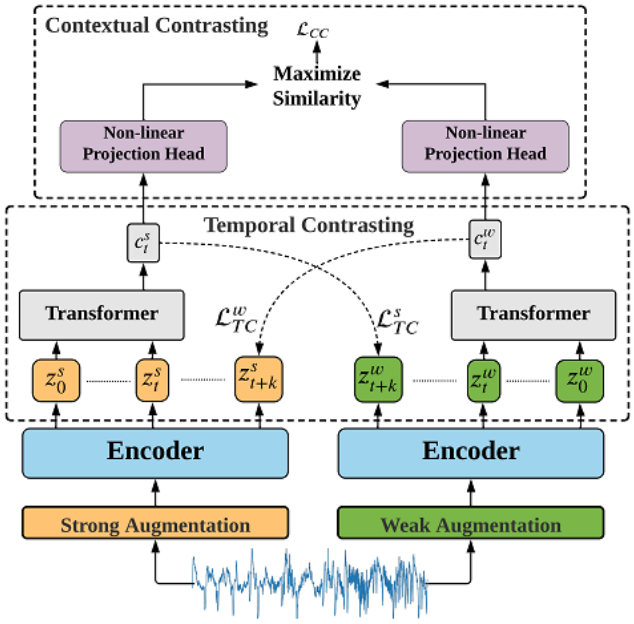

Figure 1: Illustration of TS-TCC extracted from *Figure 1* in [1]

Pretraining optimizes two contrastive objectives:

Contextual contrasting maximizes agreement between augmented views of the same window using scaling and Gaussian noise. Given $2N$ projected context vectors $\{c_i\}_{i=1}^{2N}$, the loss is

$$\mathcal{L}_{\text{CC}} = -\sum_{i=1}^{N} \log \frac{\exp\big(\text{sim}(c_i, c_i^+)/\tau\big)}{\sum_{m=1}^{2N} \mathbb{1}[m \neq i]\, \exp(\text{sim}(c_i, c_m)/\tau)},$$

where $\text{sim}(u, v) = \frac{u^\top v}{\|u\|\|v\|}$ is cosine similarity, c_i^+ is the paired view from the same sample, and $\tau = 0.8$ is the temperature parameter controlling softmax sharpness. The contribution of this term to the total objective is weighted by $\lambda_{\text{CC}} = 0.7$.

Temporal contrasting predicts future representations from context. Let c_t^s denote the strong-augmentation context vector and z_{t+k}^w the weak-augmentation latent at a future step $t+k$. The loss

$$\mathcal{L}_{\text{TC}}^s = -\frac{1}{K}\sum_{k=1}^{K} \log \frac{\exp\big((W_k c_t^s)^\top z_{t+k}^w\big)}{\sum_{n\in\mathcal{N}_{t,k}} \exp((W_k c_t^s)^\top z_n^w)},$$

is computed symmetrically for weak→strong as $\mathcal{L}_{\text{TC}}^w$. The final SSL objective is

$$\mathcal{L}_{\text{SSL}} = \lambda_{\text{TC}}\,(\mathcal{L}_{\text{TC}}^s + \mathcal{L}_{\text{TC}}^w) + \lambda_{\text{CC}}\,\mathcal{L}_{\text{CC}},$$

with $\lambda_{\text{TC}} = 1.0$.

D. Additional Losses and Extensions

Soft temporal and instance-level contrastive losses [2] were evaluated as extensions to TS-TCC. While showing promise on smaller datasets, they exhibited inconsistent performance when applied to large-scale pretraining on D1–D5, improving some sessions while degrading others.

IV. Experiments

A. Experimental Setup

Self-supervised pretraining uses sessions D1–D5 (40 epochs, batch size 128, Adam lr=3e-4) with contextual contrasting (τ=0.8, weight=0.7) and temporal contrasting (22 future timesteps, weight=1.0), implemented within the TS-TCC framework. Two backbone architectures are considered: the standard TS-TCC encoder and the baseline transformer model INL.

Downstream training on D6 follows two strategies: Linear Probing (LP), where the pretrained encoder is frozen and only a regression head is trained, and Fine-Tuning (FT), where the entire encoder is updated jointly with the regressor. As supervised references, a model trained only on D6 (Sup) and another one trained with labels from D1–D5 used for pretraining (P-Sup) are included. A frozen random encoder (Rand + LP) provides a baseline to assess the effect of self-supevised pretraining compared to untrained representations.

Evaluation measures cross-session generalization using R^2 scores on 12 temporally distant test sessions spanning days to months after training. Additional experiments include instance-level (Inst) and temporal (Temp) contrastive losses, applied independently during pretraining.

Dataset	SSL + LP	+ temp 1	+ inst 1	SSL + FT	+ temp 2	+ inst 2	Rand + LP	Sup INL	Sup	P-Sup INL	P-Sup
20241218	+0,5946	+0,5423	+0,6190	+0,7109	+0,6921	+0,6568	+0,3403	**+0,7168**	+0,6964	+0,5875	+0,6189
20241219	+0,5550	+0,4619	+0,5649	**+0,6740**	+0,5443	+0,6133	+0,2672	+0,6326	+0,5762	+0,6582	+0,6529
20241230	+0,5256	+0,3330	+0,4682	+0,6311	+0,5655	+0,5740	+0,1460	+0,2800	+0,5201	+0,4431	**+0,6500**
20250107	+0,1044	+0,0368	+0,3158	+0,3920	+0,3487	**+0,4080**	-0,0582	+0,0588	+0,1653	+0,1825	+0,3020
20250108	+0,1483	+0,0702	+0,2464	**+0,5012**	+0,4002	+0,4384	-0,0529	-0,2165	+0,1565	+0,3623	+0,4963
20250109	**+0,5682**	+0,3743	+0,3571	+0,4292	+0,2545	+0,3285	-0,2606	-0,4897	-0,2919	+0,4587	+0,5238
20250110	**+0,5633**	+0,4657	+0,4226	+0,4543	+0,3483	+0,3412	-0,4053	-0,9399	-0,0892	+0,4639	+0,5242
20250205	-0,0305	+0,1665	+0,1373	+0,2838	+0,2262	+0,1814	-0,0103	-0,1864	+0,1194	-0,3722	**+0,2968**
20250206	+0,0356	**+0,1459**	+0,0344	-0,1092	+0,1274	-0,0924	-0,0028	-0,9344	-0,4594	-0,3229	-0,4111
20250211	**+0,4352**	+0,2812	+0,2180	+0,2479	+0,2560	+0,1889	+0,0869	-0,5900	-0,0481	-0,4582	+0,1138
20250212	+0,2320	+0,1463	+0,2258	+0,2501	+0,2101	+0,2491	-0,0952	-0,0531	+0,1697	+0,1876	**+0,3137**
20250213	**+0,3352**	+0,1482	+0,1353	+0,1786	+0,1808	+0,0703	+0,0058	+0,1250	+0,1192	+0,0403	+0,1355
mean	+0,3389	+0,2643	+0,3121	**+0,3870**	+0,3462	+0,3298	-0,0033	-0,1331	+0,1362	+0,1859	+0,3514
median	+0,3852	+0,2239	+0,2811	**+0,4106**	+0,3021	+0,3348	-0,0065	-0,1198	+0,1379	+0,2750	+0,4050
std	+0,2317	+0,1691	+0,1800	+0,2345	+0,1744	+0,2254	+0,2053	+0,5417	+0,3412	+0,3861	+0,3050

Figure 2: Comparison of R^2 scores across models and training strategies described in IV-A. LP and FT denote Linear Probing (frozen encoder) and Fine-Tuning (full update), respectively; Sup and P-Sup indicate supervised training on D6 only and on D1–D5, and Rand + LP a frozen random encoder. *inst* and *temp* refer to instance-level and temporal contrastive losses. Unless marked as INL (baseline transformer), all models use TS-TCC.

B. Results

Self-supervised pretraining significantly improves cross-session generalization (Figure 2). Compared to the supervised D6 baseline, SSL + LP and SSL + FT increase mean test-session R^2 by +0.2027 and +0.2508, respectively. For the INL model, the gains are even larger: +0.4720 (LP) and +0.5201 (FT).

Some sessions show especially strong improvements with LP. For example, sessions 20250109 / 20250110 improve by +0.8601 / +0.6525 over the supervised D6 TS-TCC model, and by +1.0579 / +1.5032 over the supervised D6 INL model. Finally, SSL + FT also outperforms supervised training on D1–D5 by +0.0356 on average, despite using no labels.

1) Label Fractioning: Label fraction experiments (see Figure 3) confirm previous results and show remarkable efficiency: SSL + FT achieves $R^2 = 0.4359$ with only 1% labeled data vs. $R^2 = 0.2428$ for supervised baseline (+0.1931). SSL enables reasonable performance ($R^2 > 0.4$) with just 1–3% labeled data. In addition, SSL + LP even outperforms the supervised baseline when having less than 20% of labeled data.

2) Embedding Analysis: To further understand the representations learned during self-supervised pretraining, D6's embeddings are analysed from a random and pretrained CNN encoder using the UMAP (n_neighbors=10, target_metric=l_2) manifold learning technique (see Figure 4). For the random encoder, local consistency (using K=10 nearest neighbors), which quantifies how similar the target values of neighboring points are in the embedding space, and LP R^2 scores reach 0.772 and 0.415, respectively, whereas the pretrained encoder on D1–D5 achieves 0.997 local consistency and 0.727 R^2. Better clustering indicates that SSL effectively organizes input data according to underlying motor patterns without using position labels during pretraining, supporting improved downstream regression performance.

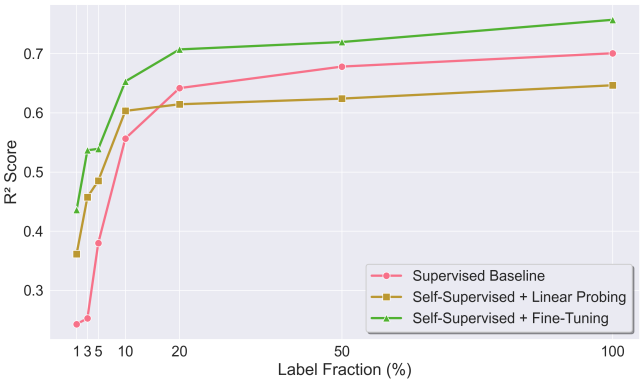

Figure 3: Performance (R^2 score) as a function of the fraction of D6 used for training. Only the training split of D6 is fractioned, while the test split remains unchanged. Self-supervised pretraining is performed on D1–D5.

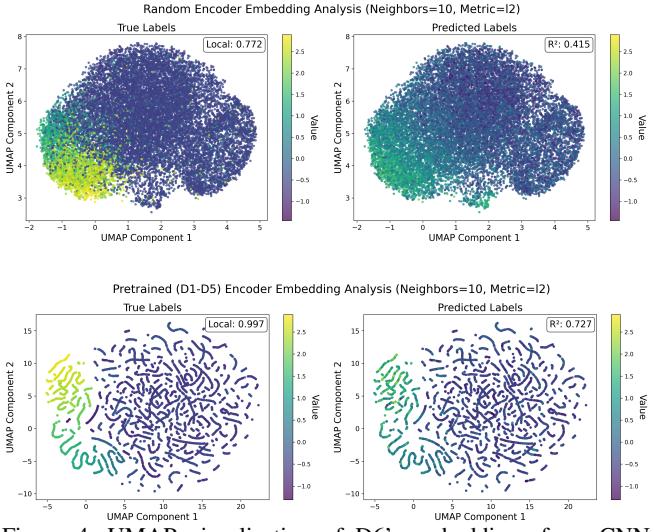

Figure 4: UMAP visualization of D6's embeddings from CNN encoders. Top row shows a random encoder while bottom row shows a pretrained encoder on D1–D5. Left plots are colored by true wrist positions with local consistency scores; right plots show predicted wrist positions via linear probing on D6 with R^2 scores.

3) Ablation Study: The contribution of each loss component in the TS-TCC framework is investigated via ablation. The Contextual Contrastive (CC) loss, as well as the Temporal Contrastive losses from weak (TC^W) and strong (TC^S) augmentations, are removed individually. Removing both temporal contrastive losses simultaneously is denoted as TC. All ablations are evaluated using LP and FT, reporting average and median R^2 on the same cross-session test set as Figure 2.

Model Config	TS-TCC	– CC	– TC^W	– TC^S	– TC
LP (avg R^2)	**0.3389**	0.2849	0.2904	0.3115	0.1228
LP (med R^2)	**0.3852**	0.3201	0.2525	0.3025	0.1126
FT (avg R^2)	0.3870	**0.3902**	0.3464	0.1588	0.2505
FT (med R^2)	**0.4106**	0.3634	0.2883	0.1676	0.2517

Table I: Ablation study reporting average (avg) and median (med) R^2. The TS-TCC model is ablated by removing individual losses: Contextual Contrastive (CC), Weak and Strong Temporal Contrastive (TC^W, TC^S), or both simultaneously (TC).

V. Discussion

SSL improves generalization by learning robust representations from data structure rather than potentially biased label supervision. The temporal and contextual contrastive objectives capture stable motor-related neural dynamics across sessions, evidenced by SSL + FT outperforming supervised baselines even without position labels during pretraining.

A critical limitation is sensitivity to data scaling and normalization across recording sessions. To achieve the reported results, adaptive normalization was applied by scaling each test session using statistics computed from the first 12% of that session's data. Without this adaptive scaling, all models performed poorly when using normalization parameters from training sessions, making meaningful cross-session comparison difficult. It reveals the fundamental challenge of neural signal non-stationarity over time, highlighting that even SSL approaches remain sensitive to distributional shifts in neural recordings.

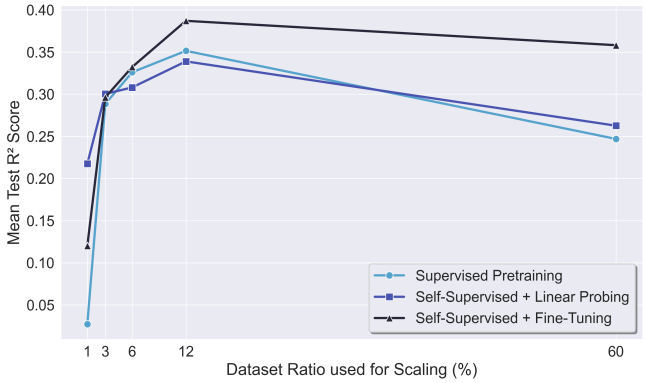

Figure 5: Different mean decoding performances (R^2 score) based on the data's ratio used for scaling the entire session's data.

The demonstrated label efficiency (reasonable performance with 1–3% labeled data) addresses neural decoder development bottlenecks. For hardware deployment, self-supervised pretrained models can serve as teacher networks for knowledge distillation into efficient architectures suitable for real-time BMI applications.

VI. Conclusion

This work demonstrates that TS-TCC self-supervised pretraining significantly improves ECoG-based motor decoding cross-session generalization (+0.20–0.25 R^2 improvement) compared to supervised (Sup) approaches. Key findings: (1) SSL achieves superior future-session performance while supervised baselines often fail catastrophically, (2) remarkable label efficiency enables reasonable performance with only 1–3% labeled data, and (3) SSL can outperform supervised models trained on multiple sessions (P-Sup) without using labels. These results establish SSL as valuable for neural signal processing, particularly for long-term BMI applications requiring cross-session stability.

References

[1] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. *arXiv preprint arXiv:2106.14112*, 2021.

[2] Seunghan Lee, Taeyoung Park, and Kibok Lee. Soft contrastive learning for time series. *arXiv preprint arXiv:2312.16424*, 2023.

[3] Steven M Peterson, Zoe Steine-Hanson, Nathan Davis, Rajesh PN Rao, and Bingni W Brunton. Generalized neural decoders for transfer learning across participants and recording modalities. *Journal of Neural Engineering*, 18(2):026014, 2021.

[4] Henri Lorach, Andrea Galvez, Valeria Spagnolo, Felix Martel, Serpil Karakas, Nadine Intering, Molywan Vat, Olivier Faivre, Cathal Harte, Salif Komi, et al. Walking naturally after spinal cord injury using a brain–spine interface. *Nature*, 618(7963):126–133, 2023.

[5] Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. Brainbert: Self-supervised representation learning for intracranial recordings. *arXiv preprint arXiv:2302.14367*, 2023.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

[7] Kexin Zhang, Qingsong Wen, Chaoli Zhang, Rongyao Cai, Ming Jin, Yong Liu, James Y Zhang, Yuxuan Liang, Guansong Pang, Dongjin Song, et al. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *IEEE transactions on pattern analysis and machine intelligence*, 46(10):6775–6794, 2024.