SIFR: DOES SESSION-INVARIANT FACTORIZED REPRESENTATION IMPROVE CROSS-SESSION EEG IDENTITY RECOGNITION?

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-session EEG identity recognition faces a direct tension: a model has to keep the stable cues that distinguish one person from another, while still adapting to changes caused by headset re-wearing, acquisition conditions, and task state. We propose SIFR, a Session-Invariant Factorized Representation framework that turns this problem into controlled residual compensation inside the normalization layer. Its core module, FactorNorm, adds an input-conditioned session branch on top of shared affine parameters. A low-rank gate adjusts the compensation dimension by dimension, while zero-residual initialization keeps the starting point exactly aligned with LayerNorm. Session classification supervision gives the compensation branch an information target, and the Orthogonal Prototype Constraint (OPC) limits directional overlap with the current identity prototype. Under explicit assumptions, we derive a linear response bound for this geometric constraint and show how it limits perturbations along a specified direction. Across session-wise records on PhysioNet, MCD, and SEED-IV, SIFR reaches mean identification accuracies of 99.28%, 82.14%, and 72.34%, improving over the strongest external mean-ACC baseline on each dataset by 0.21, 4.85, and 3.06 percentage points, respectively. Ablation and identity-parameter freezing experiments further examine the roles of gating, geometric regularization, and restricted parameter updates. The results show that organizing cross-session representation learning around explicit compensation can improve identification accuracy while also making genuine and impostor identity claims more separable in score space; the component analysis further exposes the roles played by the gate and the geometric constraint.

Keywords: EEG identity recognition; cross-session generalization; factorized normalization; adaptive gating; orthogonal prototype constraint

1 INTRODUCTION

When the same person is recorded again, EEG signals can shift because of electrode contact, repositioning, task state, and recording time, even though identity remains unchanged (Del Pozo-Banos et al., 2014; Das et al., 2016; Maiorana & Campisi, 2017; Arnau-González et al., 2021; Özdenizci et al., 2019; Debie et al., 2021; Miao et al., 2023; Jin et al., 2024b; Liu et al., 2025). For interactive EEG systems, such variation can increase re-enrollment burden, false rejection, and false acceptance. The central goal is therefore to learn an identity representation that remains reliable across recording sessions.

Existing EEG models improve discrimination through convolution, attention, and graph structure (Lawhern et al., 2018; Mane et al., 2021), while domain alignment and invariant learning reduce distribution differences across recordings (Arjovsky et al., 2019; Sun & Saenko, 2016; Ganin & Lempitsky, 2015; Long et al., 2015; Özdenizci et al., 2019). However, when identity cues and session variation are entangled in the same features, these approaches do not explicitly determine where compensation should occur or how strongly it should modify the representation.

We use normalization as the intervention point. LayerNorm standardizes each token and applies globally shared affine parameters (Ba et al., 2016), but provides no explicit input-dependent com-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

pensation path. Prior normalization-based transfer methods suggest that this location can be effective under distribution shift (Chang et al., 2019; Apicella et al., 2023; Pilault et al., 2020). We therefore hypothesize that cross-session robustness can be improved by preserving a shared affine reference while confining input-conditioned changes to a controlled residual.

SIFR implements this idea through FactorNorm, which decomposes the effective affine transformation into global identity parameters and a gated session residual. A low-rank gate controls the compensation magnitude, session supervision encourages the residual to encode session-related variation, and the Orthogonal Prototype Constraint (OPC) limits its directional overlap with the current identity prototype. Together, these components connect statistical compensation with an explicit geometric constraint.

Our contributions are threefold. First, we introduce a zero-residual factorized affine module that separates a shared identity reference from input-conditioned compensation in the later Transformer blocks. Second, we combine session supervision with a prototype-direction constraint and derive a conditional response bound that clarifies the scope of OPC. Third, across three datasets with distinct acquisition protocols, we provide session-wise comparisons, component ablations, and identity-branch freezing experiments to evaluate recognition performance, component roles, and restricted updating.

2 Related Work

2.1 EEG Identity Representation and Cross-Session Evaluation

CNN-, Transformer-, and graph-based models such as EEGNet, TSCeption, FBCNet, ATCNet, DGCNN, and ConFormer provide strong EEG representation baselines (Lawhern et al., 2018; Mane et al., 2021; Altaheri et al., 2022; Ding et al., 2022; Song et al., 2022; 2018). However, discriminative feature extraction does not by itself guarantee stability across sessions, However, discriminative feature extraction does not by itself guarantee stability across sessions, where identity information must be preserved under recording variation (Del Pozo-Banos et al., 2014; Das et al., 2016; Maiorana & Campisi, 2017; Arnau-González et al., 2021; Özdenizci et al., 2019; Debie et al., 2021; Miao et al., 2023; Jin et al., 2024b; Liu et al., 2025).

Multi-session EEG authentication further distinguishes identification from verification (Xu et al., 2021). Identification selects the correct enrolled identity, whereas verification accepts or rejects an identity claim from a score. We therefore use closed-set cross-session identification as the main task and retain EER as a complementary verification metric.

2.2 Conditional Normalization and Domain Compensation

Conditional normalization methods such as conditional batch normalization, FiLM, and AdaIN generate affine transformations from conditioning signals (Huang & Belongie, 2017; Perez et al., 2018; De Vries et al., 2017). Here, CLN denotes conditional LayerNorm in a general sense, where an input descriptor modulates normalization parameters; comparisons must therefore specify the conditioning input, generator, insertion position, and update range (Pilault et al., 2020).

Normalization has also been used for domain adaptation through DSBN, SPD domain-specific normalization, and dynamically transformed instance normalization (Chang et al., 2019; Kobler et al., 2022; Jiao et al., 2022). These studies motivate normalization as an intervention point under distribution shift (Apicella et al., 2023; Kobler et al., 2022; Jiao et al., 2022). SIFR differs by jointly organizing a shared affine reference, gated residual, session supervision, prototype-direction constraint, and restricted-update scheme.

2.3 Distribution Alignment and Prototype Constraints

DANN, CORAL, and MMD reduce cross-domain discrepancy through adversarial, second-order, and kernel-based alignment (Sun & Saenko, 2016; Ganin & Lempitsky, 2015; Long et al., 2015). When domain variation is entangled with class information, alignment may interfere with discrimination, motivating more controlled compensation (Özdenizci et al., 2019; Bousmalis et al., 2016).

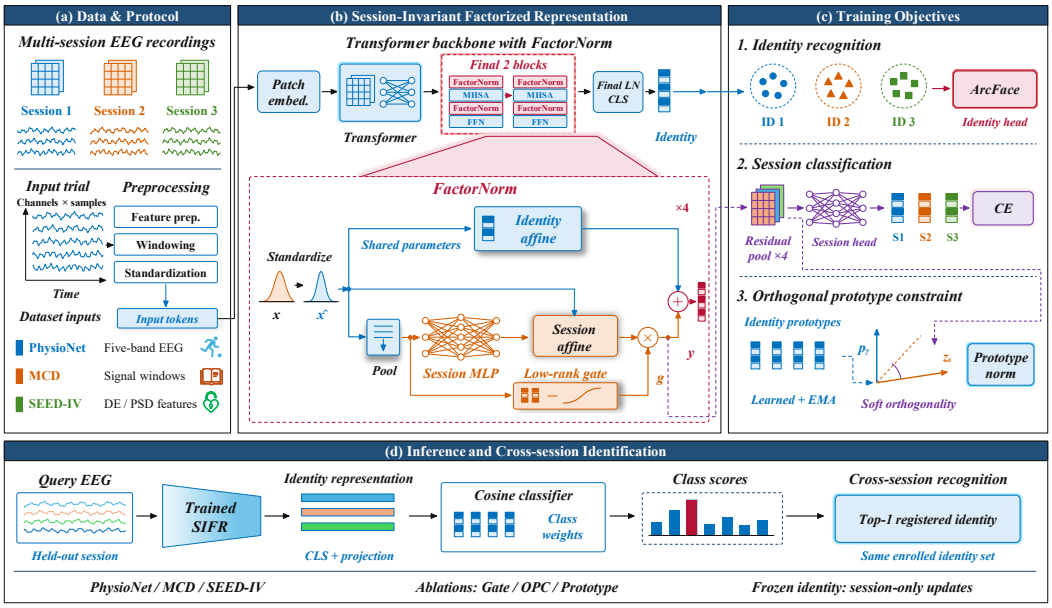


Figure 1: Overview of SIFR. FactorNorm combines a shared identity affine reference with a gated input-conditioned session residual in the later Transformer blocks. Session classification and OPC regularize the residual during training, while inference uses only the identity path for enrolled-identity recognition. Waveforms, clusters, and score bars are schematic illustrations.

SIFR therefore constrains the organization of the input-conditioned residual rather than directly aligning complete feature distributions.

Prototype learning represents discriminative structure through class references (Snell et al., 2017). SIFR uses identity prototypes to constrain the direction of session compensation, rather than replacing the classifier with prototype distance. This converts reduced interference along the identity direction into an explicit training objective while keeping its geometric interpretation separate from the final prediction behavior (Ranasinghe et al., 2021).

Figure 1 summarizes the data flow, factorized normalization, training objectives, and cross-session inference procedure.

3 Methodology

3.1 Task Definition and Overall Pipeline

Each EEG sample has an identity label a and a session label m, with the same K enrolled identities shared across training and test sessions. Dataset-specific EEG signals or frequency-domain features are tokenized and processed by a Transformer (Vaswani et al., 2017). FactorNorm replaces the normalization modules in the last two Transformer blocks, with four instances in total. After the final LayerNorm (Ba et al., 2016), the CLS token is used for identity classification.

The backbone learns identity-discriminative features, while FactorNorm provides input-dependent affine compensation. Session supervision and OPC act only during training on the compensation representation; inference requires neither identity nor session labels for generating the compensation.

3.2 From Standardization to a Conditional Descriptor

For an input with N tokens $x_n \in \mathbb{R}^D$, we standardize each token along the feature dimension and average the results:

$$\hat{x}_n = \frac{x_n - \mu_n}{\sqrt{v_n + \varepsilon}}, \qquad s = \frac{1}{N}\sum_{n=1}^{N} \hat{x}_n. \tag{1}$$

The descriptor s reduces direct offset and scale effects on the conditional generator. We refer to it as a session descriptor, although it may still retain identity information.

3.3 Identity Reference and Gated Session Residual

The identity branch contains global affine parameters, whereas the session branch generates input-dependent scale and bias residuals from s:

$$\begin{bmatrix} \gamma_{\text{sess}} \\ \beta_{\text{sess}} \end{bmatrix} = W_2\,\mathrm{GELU}(W_1 s + b_1) + b_2. \tag{2}$$

A second low-rank path generates a dimension-wise gate:

$$g = \sigma(VU^\top s + b_g), \qquad U, V \in \mathbb{R}^{D\times q}, \quad g \in (0,1)^D. \tag{3}$$

The gate scales only the session residual, which is added to the shared affine reference:

$$\begin{aligned} \gamma_{\text{eff}} &= \gamma_{\text{id}} + g \odot \gamma_{\text{sess}}, \\ \beta_{\text{eff}} &= \beta_{\text{id}} + g \odot \beta_{\text{sess}}. \end{aligned} \tag{4}$$

The module output and compensation term are

$$\begin{aligned} y_n &= \gamma_{\text{eff}} \odot \hat{x}_n + \beta_{\text{eff}}, \\ d_n &= g \odot (\gamma_{\text{sess}} \odot \hat{x}_n + \beta_{\text{sess}}). \end{aligned} \tag{5}$$

Thus, FactorNorm consists of a shared affine path plus a gated residual d_n, rather than a mixture of two branches. Zero-residual initialization exactly recovers the corresponding LayerNorm at initialization (Ba et al., 2016); details are given in Appendix B.

3.4 Identity and Session Representations

The session representation averages the compensation terms across the four FactorNorm modules, while the final CLS token defines the identity representation:

$$\begin{aligned} z_{\text{sess}} &= \frac{1}{L_f}\sum_{l=1}^{L_f}\frac{1}{N_l}\sum_{n=1}^{N_l} d_{ln}, \qquad L_f = 4, \\ z_{\text{id}} &= \big(\mathrm{LN}_{\text{final}}(H)\big)_{\text{CLS}}. \end{aligned} \tag{6}$$

Here, z_{sess} is an aggregate compensation descriptor rather than the final classifier displacement, whereas z_{id} has already been affected by both affine branches. This distinction determines the scope of the geometric analysis below.

3.5 Geometric Constraint of OPC and Its Scope

For each identity k, we maintain a prototype $p_k \in \mathbb{R}^D$ (Snell et al., 2017). OPC penalizes the squared cosine between the session representation and the corresponding identity prototype (Ranasinghe et al., 2021):

$$\mathcal{L}_{\text{OPC}} = \frac{1}{B}\sum_{i=1}^{B} \cos^2(z_{\text{sess},i}, p_{a_i})\,. \tag{7}$$

Table 1: PhysioNet Cross-Session Identity Recognition Results (%)

Method	S1			S2			S3		
	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓
ArjunViT	90.07 ± 1.69	89.78 ± 1.41	2.09 ± 0.41	96.28 ± 1.37	96.26 ± 1.37	0.44 ± 0.08	93.90 ± 1.62	93.71 ± 1.13	0.91 ± 0.08
TSCeption	98.78 ± 1.26	98.65 ± 0.82	0.67 ± 0.08	99.36 ± 1.35	99.35 ± 0.83	0.09 ± 0.02	99.08 ± 1.34	99.06 ± 0.71	0.21 ± 0.04
DGCNN	98.67 ± 1.91	98.55 ± 1.50	0.49 ± 0.09	99.36 ± 1.85	99.36 ± 1.85	**0.05 ± 0.02**	98.58 ± 1.85	98.55 ± 1.54	0.26 ± 0.03
CSPNet	97.61 ± 0.87	97.47 ± 1.79	0.58 ± 0.12	99.04 ± 0.84	99.04 ± 1.45	0.06 ± 0.02	98.48 ± 0.70	98.39 ± 1.67	0.30 ± 0.04
ConFormer	97.99 ± 1.46	97.86 ± 0.61	0.53 ± 0.06	99.18 ± 1.45	99.17 ± 0.66	0.08 ± 0.02	97.96 ± 1.68	97.84 ± 0.57	0.56 ± 0.05
FBCNet	62.81 ± 0.94	61.70 ± 1.22	6.88 ± 0.81	71.41 ± 0.92	70.45 ± 1.37	6.63 ± 0.68	64.07 ± 1.04	63.11 ± 1.33	8.16 ± 1.62
ATCNet	98.09 ± 0.69	97.99 ± 0.94	0.53 ± 0.08	99.26 ± 0.68	99.26 ± 0.94	0.12 ± 0.02	98.78 ± 0.82	98.74 ± 0.96	0.35 ± 0.05
EEGNet	88.19 ± 1.81	87.71 ± 0.57	2.33 ± 0.38	94.19 ± 1.93	94.19 ± 0.56	1.11 ± 0.12	91.28 ± 1.71	91.19 ± 0.63	1.71 ± 0.18
W-ResNet	98.38 ± 0.96	98.22 ± 1.34	0.81 ± 0.10	99.08 ± 0.99	99.08 ± 1.56	0.08 ± 0.02	97.90 ± 0.81	97.77 ± 1.51	0.38 ± 0.07
iCaRL	74.02 ± 0.85	73.72 ± 1.74	4.09 ± 0.69	81.98 ± 0.94	81.42 ± 1.75	3.45 ± 0.44	79.94 ± 0.93	79.28 ± 1.63	4.05 ± 0.40
GCNN	92.40 ± 1.43	92.24 ± 1.61	1.13 ± 0.22	98.09 ± 1.50	98.09 ± 1.82	0.27 ± 0.04	95.03 ± 1.32	94.73 ± 1.68	1.07 ± 0.11
GVAE	95.03 ± 1.20	94.79 ± 1.17	1.08 ± 0.15	98.78 ± 1.13	98.78 ± 0.93	0.20 ± 0.04	97.42 ± 1.07	97.34 ± 1.05	0.63 ± 0.12
CNN-Attention	98.14 ± 0.87	97.99 ± 1.74	0.59 ± 0.09	99.18 ± 0.82	99.18 ± 1.97	0.12 ± 0.02	98.69 ± 0.67	98.61 ± 1.68	0.38 ± 0.07
LRMD	98.64 ± 0.76	98.52 ± 1.09	0.40 ± 0.04	99.01 ± 0.81	99.01 ± 1.10	**0.05 ± 0.02**	98.75 ± 0.68	98.72 ± 0.97	0.23 ± 0.04
PGCN	95.55 ± 2.15	95.40 ± 1.19	0.69 ± 0.11	98.66 ± 1.83	98.65 ± 1.37	0.30 ± 0.05	97.15 ± 1.94	96.98 ± 1.16	0.67 ± 0.10
SIFR	**98.80 ± 0.31**	**98.77 ± 0.32**	**0.14 ± 0.05**	**99.71 ± 0.24**	**99.71 ± 0.25**	**0.05 ± 0.03**	**99.34 ± 0.29**	**99.33 ± 0.30**	**0.12 ± 0.04**

Mean ± SD over three random seeds. F1 denotes macro-F1; best values in each column are bolded, including ties.

OPC is a soft directional constraint rather than a hard projection. To make its effect explicit, consider a unit prototype p, perturbation d, and linear classification weight w. If $w = ap + e$ with $e^\top p = 0$ and $|p^\top d| \le \eta\|d\|$, then

$$\begin{aligned} w = ap + e, \qquad & e^\top p = 0, \\ |p^\top d| \le \eta\|d\| \quad \Longrightarrow \quad & |w^\top d| \le \|d\|\big(|a|\eta + \|e\|\big). \end{aligned} \tag{8}$$

The bound separates residual prototype alignment from the classifier component orthogonal to the prototype. Its proof and an incorrect-class counterexample are given in Appendix A. Because the actual identity head also contains learned projection and normalization, the bound explains OPC's directional preference rather than guaranteeing invariant classification logits.

3.6 Joint Objective and Prototype Update

Identity supervision uses ArcFace (Deng et al., 2019), while the session head applies cross-entropy to z_{sess}. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\text{id}} + \lambda_s \mathcal{L}_{\text{sess}} + \lambda_o \mathcal{L}_{\text{OPC}} + \lambda_p \frac{1}{K} \sum_{k=1}^{K} \|p_k\|^2. \tag{9}$$

Session supervision encourages the compensation representation to encode session information, whereas OPC limits its overlap with identity directions. Prototypes are initialized orthogonally, jointly optimized, and additionally updated by EMA with $\rho = 0.99$. We use $\lambda_s = 0.05$, $\lambda_o = 0.05$, and $\lambda_p = 0.0005$. Inference uses only the identity prediction path; detailed loss and prototype-update definitions are provided in Appendix B.

4 Experimental Design

4.1 Research Questions and Data Records

We evaluate SIFR from three perspectives: overall cross-session recognition,component contribution, and performance under restricted updating. All ACC,macro-F1, and EER values are reported as mean ± standard deviation over three random seeds, while session means are used only to summarize trends. The freezing experiment follows the same three-seed setting before and after freezing.

4.2 Datasets and Cross-Session Protocols

PhysioNet contains 109 identities with 64-channel EEG (Goldberger et al., 2000). We group different runs into three pseudo-sessions and use five-band 320×512 inputs; thus, this setting mainly

Table 2: MCD Cross-Session Identity Recognition Results (%)

Method	IQ			Memory			Reading		
	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓
ArjunViT	30.17 ± 1.67	27.31 ± 1.24	24.37 ± 4.86	79.86 ± 1.62	79.36 ± 1.13	4.42 ± 0.66	77.85 ± 1.46	77.48 ± 1.09	5.24 ± 0.66
TSCeption	49.06 ± 1.38	46.19 ± 0.85	15.38 ± 2.63	90.97 ± 1.41	90.77 ± 0.75	2.40 ± 0.38	88.33 ± 1.21	88.23 ± 0.82	2.81 ± 0.42
DGCNN	41.11 ± 1.66	38.28 ± 1.49	19.17 ± 3.72	86.04 ± 1.80	85.88 ± 1.83	3.44 ± 0.51	83.40 ± 1.79	83.27 ± 1.49	4.13 ± 0.43
CSPNet	46.53 ± 0.70	43.71 ± 1.59	15.35 ± 2.84	90.52 ± 0.81	90.42 ± 1.86	2.33 ± 0.36	88.19 ± 0.82	88.03 ± 1.60	3.05 ± 0.27
ConFormer	45.28 ± 1.61	42.03 ± 0.58	26.67 ± 4.08	94.65 ± 1.51	94.60 ± 0.68	1.98 ± 0.16	91.94 ± 1.37	91.92 ± 0.55	2.67 ± 0.36
FBCNet	40.45 ± 1.14	38.91 ± 1.19	16.70 ± 2.00	88.40 ± 0.94	88.54 ± 1.41	2.50 ± 0.39	76.42 ± 0.97	76.37 ± 1.34	6.35 ± 0.73
ATCNet	46.04 ± 0.75	42.60 ± 0.99	16.73 ± 1.99	90.24 ± 0.70	90.11 ± 0.97	2.64 ± 0.43	87.67 ± 0.74	87.43 ± 0.99	3.30 ± 0.30
EEGNet	24.65 ± 1.76	21.43 ± 0.61	26.94 ± 5.27	95.14 ± 1.81	94.92 ± 0.57	1.72 ± 0.29	93.20 ± 1.66	93.18 ± 0.61	2.11 ± 0.35
W-ResNet	42.85 ± 0.83	38.81 ± 1.51	17.88 ± 2.55	93.23 ± 0.86	93.22 ± 1.45	1.18 ± 0.18	91.56 ± 0.78	91.47 ± 1.54	1.78 ± 0.33
iCaRL	41.67 ± 0.98	37.69 ± 1.79	18.54 ± 1.67	90.94 ± 0.88	90.76 ± 1.84	2.19 ± 0.41	87.64 ± 0.94	87.48 ± 1.53	2.81 ± 0.43
GCNN	42.53 ± 1.59	40.66 ± 1.49	18.44 ± 3.12	84.20 ± 1.59	84.00 ± 1.49	3.46 ± 0.45	83.58 ± 1.54	83.42 ± 1.89	4.54 ± 0.57
GVAE	42.40 ± 1.16	38.78 ± 1.01	19.51 ± 2.80	86.67 ± 0.99	86.59 ± 1.08	2.85 ± 0.41	84.03 ± 1.07	84.00 ± 1.07	3.80 ± 0.66
CNN-Attention	37.88 ± 0.80	34.77 ± 1.69	20.51 ± 2.34	87.26 ± 0.81	86.99 ± 1.77	2.69 ± 0.39	83.92 ± 0.65	83.77 ± 1.70	3.47 ± 0.59
LRMD	45.62 ± 0.75	42.06 ± 1.11	16.70 ± 1.51	85.59 ± 0.76	85.55 ± 1.14	3.27 ± 0.37	84.94 ± 0.80	84.61 ± 0.97	5.72 ± 1.02
PGCN	43.19 ± 1.79	39.93 ± 1.16	17.85 ± 2.13	83.40 ± 2.02	83.28 ± 1.11	3.74 ± 0.30	81.18 ± 1.79	80.99 ± 1.11	4.27 ± 0.64
SIFR	**55.73 ± 0.48**	**53.31 ± 0.50**	**11.49 ± 0.60**	**96.22 ± 0.33**	**96.19 ± 0.35**	**0.76 ± 0.12**	**94.48 ± 0.36**	**94.44 ± 0.38**	**1.42 ± 0.15**

Mean ± SD over three random seeds. F1 denotes macro-F1; best values in each column are bolded, including ties.

reflects run-group variation rather than confirmed cross-day or electrode re-wearing effects. MCD contains 45 participants with 60-channel inputs of length 200 under IQ, Memory, and Reading conditions. IQ involves electrode reconnection, whereas Memory and Reading share the same physical recording, so the three conditions are not treated as independent re-wearing sessions. SEED-IV contains 15 identities, 62 channels, and three recording sessions (Zheng et al., 2019), using 620-dimensional DE/PSD features with windows of length 12. Together, the datasets cover run-group, task/reconnection, and cross-recording variation without implying a strict ordering of drift strength.

PhysioNet and SEED-IV reserve 10% of target-session samples per identity for validation, whereas MCD uses source-session validation. Training uses source-session data, but test accuracy also enters early stopping. We therefore interpret the reported results under this evaluation flow rather than as strict source-only model-selection estimates. Further splitting and implementation details are provided in Appendix B.

4.3 Compared Methods and Evaluation Metrics

We compare SIFR with fifteen external baselines spanning convolutional, Transformer-based, graph-based, attention-based, and continual-learning approaches (Lawhern et al., 2018; Mane et al., 2021; Seyfizadeh et al., 2024; Wang et al., 2020; Altaheri et al., 2022; Ding et al., 2022; Song et al., 2022; Arjun et al., 2021; Song et al., 2018; Jin et al., 2024a; Kipf & Welling, 2016; Xu et al., 2021; Rebuffi et al., 2017; Kong et al., 2018; Behrouzi & Hatzinakos, 2022). A same-backbone LayerNorm baseline and component ablations isolate the structural contribution of SIFR, while CLN (Pilault et al., 2020) is retained only as a supplementary reference for restricted updating.

ACC and macro-F1 are higher-better and EER is lower-better. Best values are bolded, including ties; reported differences are in percentage points.

4.4 Main Results and Session-Wise Analysis

PhysioNet is close to saturation: SIFR reaches a mean ACC of 99.28%, only 0.21 percentage points above TSCeption at 99.07%. The small margin is therefore better interpreted as maintained performance under an easier protocol than as evidence of statistical significance.

On MCD, SIFR achieves a mean ACC of 82.14%, improving over ConFormer by 4.85 percentage points, with the largest difficulty in the IQ condition (55.73% ACC). Lower EERs across all three conditions further indicate improved genuine–impostor score separability in addition to closed-set identification.

On SEED-IV, SIFR reaches a mean ACC of 72.34%, exceeding ConFormer by 3.06 percentage points, together with a mean EER of 10.46%. The consistent EER reduction across sessions suggests that the ACC gains are accompanied by better identity-claim separability under the current evaluation protocol.

Table 3: SEED-IV Cross-Session Identity Recognition Results (%)

Method	S1			S2			S3		
	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓
ArjunViT	60.05 ± 1.69	56.35 ± 1.41	15.05 ± 2.96	76.93 ± 1.37	72.56 ± 1.37	9.89 ± 1.83	65.61 ± 1.62	59.23 ± 1.13	21.00 ± 1.76
TSCeption	58.97 ± 1.26	56.69 ± 0.82	10.54 ± 1.20	74.34 ± 1.35	69.11 ± 0.83	11.17 ± 1.57	65.15 ± 1.34	56.37 ± 0.71	18.95 ± 3.49
DGCNN	52.10 ± 1.91	48.64 ± 1.50	15.51 ± 2.89	60.60 ± 1.85	53.39 ± 1.85	18.20 ± 2.32	53.38 ± 1.85	50.76 ± 1.54	24.36 ± 2.60
CSPNet	58.99 ± 0.87	54.95 ± 1.79	13.20 ± 2.62	74.04 ± 0.84	70.01 ± 1.45	9.09 ± 1.69	67.74 ± 0.70	62.03 ± 1.67	15.58 ± 2.16
ConFormer	61.70 ± 1.46	56.02 ± 0.61	13.35 ± 1.61	78.65 ± 1.45	75.19 ± 0.66	11.73 ± 1.47	67.49 ± 1.68	62.22 ± 0.57	15.32 ± 1.36
FBCNet	57.62 ± 0.94	53.41 ± 1.22	14.46 ± 1.70	77.80 ± 0.92	75.41 ± 1.37	8.74 ± 0.90	60.51 ± 1.04	56.12 ± 1.33	20.66 ± 4.10
ATCNet	49.07 ± 0.69	40.72 ± 0.94	29.22 ± 4.37	69.03 ± 0.68	63.70 ± 0.94	17.41 ± 1.54	55.06 ± 0.82	51.74 ± 0.96	24.44 ± 3.24
EEGNet	56.61 ± 1.81	53.30 ± 0.57	13.92 ± 2.29	70.08 ± 1.93	63.39 ± 0.56	14.37 ± 1.50	63.37 ± 1.71	55.89 ± 0.63	21.92 ± 2.27
W-ResNet	60.97 ± 0.96	56.18 ± 1.34	20.33 ± 2.52	70.93 ± 0.99	66.23 ± 1.56	16.15 ± 2.34	60.94 ± 0.81	53.75 ± 1.51	20.87 ± 3.66
iCaRL	61.38 ± 0.85	57.19 ± 1.74	11.79 ± 2.00	76.24 ± 0.94	72.08 ± 1.75	8.89 ± 1.14	65.01 ± 0.93	58.31 ± 1.63	17.65 ± 1.75
GCNN	59.08 ± 1.43	55.76 ± 1.61	12.01 ± 2.39	72.51 ± 1.50	67.87 ± 1.82	10.37 ± 1.40	64.19 ± 1.32	57.50 ± 1.68	19.85 ± 1.97
GVAE	60.16 ± 1.20	58.18 ± 1.17	9.57 ± 1.37	75.12 ± 1.13	70.78 ± 0.93	9.22 ± 1.71	66.71 ± 1.07	63.36 ± 1.05	16.23 ± 3.15
CNN-Attention	58.06 ± 0.87	55.43 ± 1.74	10.52 ± 1.58	77.78 ± 0.82	74.08 ± 1.97	10.22 ± 1.15	64.94 ± 0.67	58.09 ± 1.68	15.30 ± 2.85
LRMD	60.96 ± 0.76	58.51 ± 1.09	10.47 ± 0.98	76.47 ± 0.81	72.19 ± 1.10	10.07 ± 0.94	58.78 ± 0.68	53.15 ± 0.97	18.37 ± 3.42
PGCN	56.43 ± 2.15	51.57 ± 1.19	13.54 ± 2.22	73.49 ± 1.83	68.65 ± 1.37	8.88 ± 1.44	64.56 ± 1.94	59.49 ± 1.16	17.66 ± 2.54
SIFR	**62.12 ± 0.45**	**59.18 ± 0.47**	**9.14 ± 0.55**	**83.36 ± 0.38**	**80.33 ± 0.40**	**8.55 ± 0.48**	**71.55 ± 0.42**	**66.57 ± 0.44**	**13.69 ± 0.60**

Mean ± SD over three random seeds. F1 denotes macro-F1; best values in each column are bolded, including ties.

4.5 Ablation Study

The ablation removes the gate, OPC, or prototype bank from the full SIFR model, with a standard LayerNorm baseline included for comparison. Removing the gate lets the session residual enter affine fusion directly; removing OPC retains the prototype bank, whereas removing the prototype bank also disables prototype-dependent OPC and therefore measures their joint effect.

Relative to LayerNorm, full SIFR improves mean ACC by 0.82, 4.10, and 4.94 percentage points on PhysioNet, MCD, and SEED-IV, respectively. Removing the gate reduces mean ACC by 0.96, 3.35, and 3.92 points, while removing OPC reduces it by 0.61, 2.53, and 3.19 points. OPC removal also increases MCD-IQ EER from 11.49% to 13.33% and mean SEED-IV EER from 10.46% to 21.91%. The full configuration achieves the highest mean ACC on all three datasets, supporting complementary roles for gated compensation and prototype-related constraints. Since prototype removal also disables OPC, its effect cannot be attributed to the prototype regularizer alone.

4.6 Identity-Branch Freezing Experiment

The freezing experiment tests how much recognition ability remains when identity-related parameters are fixed. After full training, Stage 2 freezes the embedding, attention, feed-forward layers, shared identity affine parameters, identity head, and prototypes, while updating only the session MLP, low-rank gate, and session classifier on source-session data; prototype EMA is disabled. Thus, the experiment measures restricted continued training rather than target-session adaptation.

Figure 2 organizes the visual analysis from three angles: component contribution, performance distribution, and representation structure. Figure 2(a–b) shows multi-metric comparisons among ablation variants and accuracy distributions across methods. Figure 2(c,e) presents representation patterns over channel space and over frequency-band/time dimensions. Figure 2(d,f) illustrates the design intuition behind factorized affine modulation and the prototype-direction constraint, while Fig. 2(g) compares baseline and SIFR identity distributions in a two-dimensional projection. The 2D clusters and auxiliary decision regions are used only for qualitative inspection; by themselves, they do not prove high-dimensional orthogonality or statistical independence.

The mean ACC drops are 1.72, 3.27, and 4.28 percentage points on PhysioNet, MCD, and SEED-IV, respectively, compared with existing CLN records of 8.14, 8.37, and 8.59 points (Pilault et al., 2020). Because the CLN results provide only dataset-level means under a different implementation and model-selection setting, this comparison is interpreted as directional evidence for restricted updating.

SIFR adds only 1.40–2.04% parameters, with session-related parameters accounting for 2.10–2.33% of the full model; detailed counts are given in Appendix **??**.

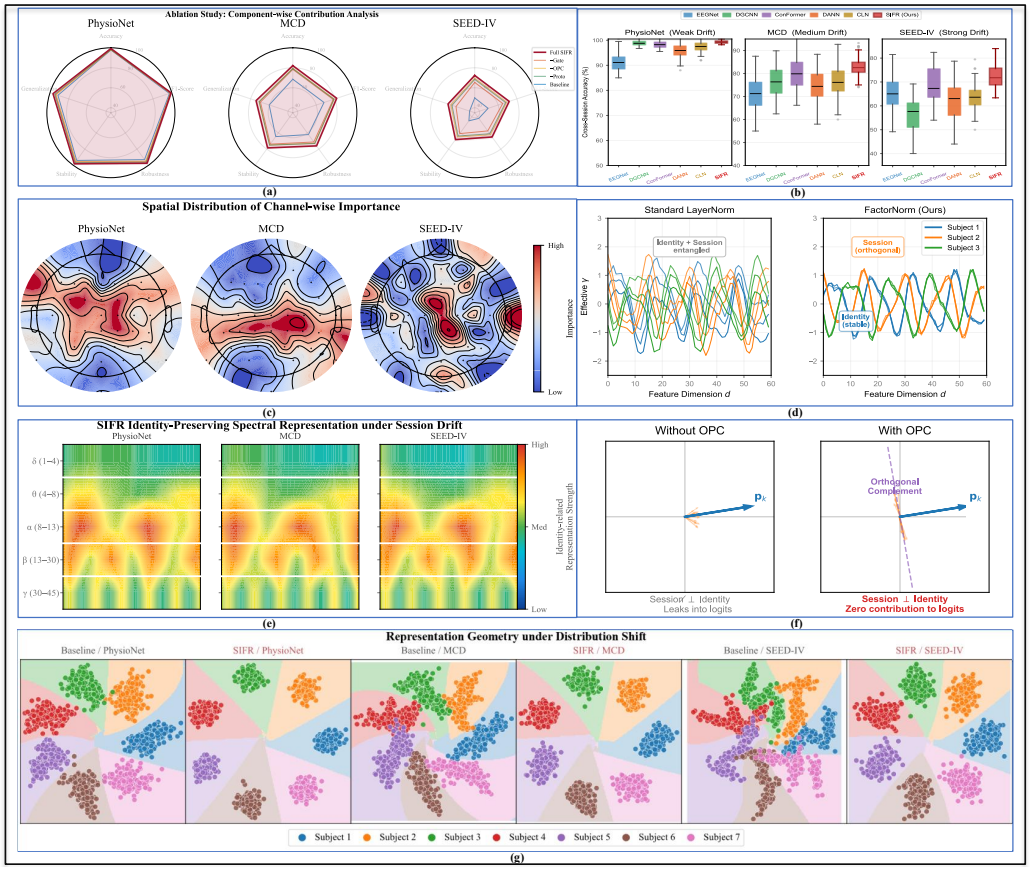


Figure 2: Component and representation analyses of SIFR. (a) Radar plots of the full model and ablations. (b) Cross-session ACC distributions of six methods. (c) Channel-importance maps in scalp space. (d) LayerNorm–FactorNorm affine modulation. (e) Identity-related representation strength over time and frequency. (f) OPC geometry between session compensation and identity prototypes. (g) Two-dimensional identity representations of the baseline and SIFR. In (g), colors denote seven identity classes and the background indicates decision regions from a two-dimensional auxiliary classifier.

Table 4: Session-Wise Ablation Results (%)

Method	S1			S2			S3		
	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓	ACC ↑	F1 ↑	EER ↓
SIFR (PhysioNet)	**98.80 ± 0.31**	**98.77 ± 0.32**	**0.14 ± 0.05**	**99.71 ± 0.24**	**99.71 ± 0.25**	**0.05 ± 0.03**	**99.34 ± 0.29**	**99.33 ± 0.30**	**0.12 ± 0.04**
w/o Gate	97.56 ± 0.97	97.53 ± 0.47	0.51 ± 0.06	98.78 ± 1.18	98.75 ± 0.60	0.56 ± 0.09	98.64 ± 1.04	98.61 ± 0.49	0.66 ± 0.10
w/o OPC	98.08 ± 0.72	98.07 ± 0.47	0.44 ± 0.08	99.31 ± 0.66	99.30 ± 0.58	0.22 ± 0.02	98.64 ± 0.64	98.58 ± 0.58	0.25 ± 0.05
w/o Prototypes	98.61 ± 1.17	98.57 ± 0.53	0.39 ± 0.05	99.47 ± 1.03	99.47 ± 0.45	**0.05 ± 0.02**	99.25 ± 1.06	99.25 ± 0.48	0.24 ± 0.05
LayerNorm	97.34 ± 1.64	97.30 ± 1.49	0.68 ± 0.11	99.25 ± 1.44	99.25 ± 1.40	0.07 ± 0.02	98.81 ± 1.43	98.80 ± 1.20	0.41 ± 0.08
SIFR (MCD)	**55.73 ± 0.48**	**53.31 ± 0.50**	**11.49 ± 0.60**	**96.22 ± 0.33**	**96.19 ± 0.35**	**0.76 ± 0.12**	**94.48 ± 0.36**	**94.44 ± 0.38**	**1.42 ± 0.15**
w/o Gate	49.93 ± 1.30	47.43 ± 0.59	20.07 ± 1.99	93.89 ± 1.17	93.81 ± 0.49	2.40 ± 0.30	92.57 ± 1.10	92.50 ± 0.49	3.12 ± 0.27
w/o OPC	50.17 ± 0.57	47.65 ± 0.49	13.33 ± 1.59	95.49 ± 0.66	95.48 ± 0.49	2.01 ± 0.37	93.19 ± 0.59	93.13 ± 0.52	2.85 ± 0.42
w/o Prototypes	50.90 ± 0.97	47.35 ± 0.53	18.12 ± 3.33	94.20 ± 1.19	94.18 ± 0.46	2.19 ± 0.24	92.22 ± 0.95	92.13 ± 0.45	2.92 ± 0.25
LayerNorm	48.96 ± 1.62	45.73 ± 1.54	22.36 ± 2.37	93.58 ± 1.53	93.53 ± 1.29	2.47 ± 0.23	91.60 ± 1.46	91.50 ± 1.59	2.92 ± 0.38
SIFR (SEED-IV)	**62.12 ± 0.45**	**59.18 ± 0.47**	**9.14 ± 0.55**	**83.36 ± 0.38**	**80.33 ± 0.40**	**8.55 ± 0.48**	**71.55 ± 0.42**	**66.57 ± 0.44**	**13.69 ± 0.60**
w/o Gate	56.18 ± 0.97	49.27 ± 0.47	23.49 ± 2.73	79.44 ± 1.18	75.15 ± 0.60	16.91 ± 2.64	69.64 ± 1.04	63.63 ± 0.49	14.52 ± 2.20
w/o OPC	58.23 ± 0.72	49.83 ± 0.47	34.12 ± 6.17	79.20 ± 0.66	73.44 ± 0.58	17.85 ± 1.71	70.04 ± 0.64	60.62 ± 0.58	13.75 ± 2.59
w/o Prototypes	60.68 ± 1.17	53.87 ± 0.53	14.56 ± 1.75	78.28 ± 1.03	72.53 ± 0.45	16.25 ± 2.19	67.04 ± 1.06	59.90 ± 0.48	23.14 ± 4.60
LayerNorm	55.04 ± 1.64	47.29 ± 1.49	14.71 ± 2.44	80.56 ± 1.44	75.59 ± 1.40	13.98 ± 2.25	66.61 ± 1.43	59.23 ± 1.20	16.61 ± 3.26

Mean ± SD over three random seeds. w/o Prototypes also disables the prototype-dependent OPC.

Table 5: Session-wise ACC before and after identity-branch freezing.

Dataset	Session	Before (%)	After (%)	Drop (pp)
PhysioNet	S1	98.80 ± 0.31	96.59 ± 0.58	2.21 ± 0.12
PhysioNet	S2	99.71 ± 0.24	98.27 ± 0.55	1.44 ± 0.22
PhysioNet	S3	99.34 ± 0.29	97.82 ± 0.51	1.52 ± 0.19
MCD	IQ	55.73 ± 0.48	52.01 ± 0.50	3.72 ± 0.23
MCD	Memory	96.22 ± 0.33	93.08 ± 0.60	3.14 ± 0.20
MCD	Reading	94.48 ± 0.36	91.52 ± 0.52	2.96 ± 0.15
SEED-IV	S1	62.12 ± 0.45	57.93 ± 0.49	4.19 ± 0.17
SEED-IV	S2	83.36 ± 0.38	78.82 ± 0.43	4.54 ± 0.14
SEED-IV	S3	71.55 ± 0.42	67.45 ± 0.53	4.10 ± 0.16

Mean $\pm$ SD over three random seeds. Drop is computed per seed as pre-freezing minus post-freezing ACC.

5 Discussion and Limitations

The main value of SIFR is to make session compensation explicit in its location, scale, and supervision target. Shared affine parameters provide a reference, the conditional residual captures input-dependent variation, the gate controls the compensation magnitude, and OPC imposes a directional preference on the update. This organization makes the compensation process directly inspectable through component removal and interventions on the update range, rather than relying on a larger conditional generator alone. A general conditional affine module may still represent a similar effective transformation, so the contribution of SIFR lies in the joint design of parameter organization, training objectives, and update scheme. Identification and verification further provide complementary evidence: ACC measures whether the correct identity ranks highest among candidate classes, whereas EER measures the separability of genuine and impostor claims under a decision threshold. SIFR achieves higher mean ACC and lower mean EER across all three datasets, suggesting that the improvement is not limited to closed-set ranking. However, OPC should be interpreted only as a directional constraint on the compensation process; it does not require every identity logit to remain unchanged, nor does the current EER analysis establish open-set authentication performance or deployment-level false-acceptance rates.

The present evidence should also be interpreted within the current validation protocol. Target-session validation, test-driven early stopping, and overlapping-window splits can affect the estimated degree of cross-session generalization and reduce the independence between model selection and final evaluation. These factors therefore limit how strongly the reported results can be interpreted, even though they do not alter the numerical records themselves. More importantly, the task studied here is enrolled-identity cross-session recognition rather than unseen-subject recognition: identities observed during evaluation are already represented during training or enrollment, while the session changes. Extending SIFR to unseen identities would require a different evaluation design, such as a new enrollment procedure, transferable classification weights, or a metric-learning formulation, rather than simply replacing the conditioning signal with subject information. The current results should therefore be understood as evidence for session-robust recognition of enrolled identities, not as evidence for cross-subject or unseen-identity generalization.conditions should follow the study documentation.

6 Conclusion

We presented SIFR, which organizes cross-session compensation through a shared identity affine reference and a gated session residual. Zero-residual initialization preserves the LayerNorm starting point, while session supervision and OPC constrain the content and direction of compensation. Across three datasets, SIFR improves mean ACC and EER relative to the compared baselines, while ablation and freezing experiments support the roles of its main components and restricted-update design. These results establish SIFR as a lightweight and interpretable framework for enrolled-identity cross-session EEG recognition.

AI USE STATEMENT

Generative AI tools were used to assist with English translation and language editing, mathematical typesetting, and manuscript organization. They were not used to formulate research hypotheses or mathematical claims, develop the proposed methodology, design or conduct experiments, generate or process experimental data, implement the proposed method, interpret experimental results, or determine the scientific conclusions of this work. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

REFERENCES

H. Altaheri, G. Muhammad, and M. Alsulaiman. Physics-informed attention temporal convolutional network for eeg-based motor imagery classification. *IEEE Transactions on Industrial Informatics*, 19(2):2249–2258, 2022.

A. Apicella, F. Isgrò, A. Pollastro, et al. On the effects of data normalization for domain adaptation on eeg data. *Engineering Applications of Artificial Intelligence*, 123:106205, 2023.

M. Arjovsky, L. Bottou, I. Gulrajani, et al. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

A. Arjun, A. S. Rajpoot, and M. R. Panicker. Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5723–5726. IEEE, 2021.

P. Arnau-González, S. Katsigiannis, M. Arevalillo-Herráez, et al. BED: A new data set for eeg-based biometrics. *IEEE Internet of Things Journal*, 8(15):12219–12230, 2021.

J. L. Ba, J. R. Kiros, and G. E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

T. Behrouzi and D. Hatzinakos. Graph variational auto-encoder for deriving eeg based graph embedding. *Pattern Recognition*, 121:108202, 2022.

K. Bousmalis, G. Trigeorgis, N. Silberman, et al. Domain separation networks. *Advances in Neural Information Processing Systems*, 29, 2016.

W. G. Chang, T. You, S. Seo, et al. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7354–7362, 2019.

R. Das, E. Maiorana, and P. Campisi. Eeg biometrics using visual stimuli: A longitudinal study. *IEEE Signal Processing Letters*, 23(3):341–345, 2016.

H. De Vries, F. Strub, J. Mary, et al. Modulating early visual processing by language. *arXiv preprint arXiv:1707.00683*, 2017.

E. Debie, N. Moustafa, and A. Vasilakos. Session invariant eeg signatures using elicitation protocol fusion and convolutional neural network. *IEEE Transactions on Dependable and Secure Computing*, 19(4):2488–2500, 2021.

M. Del Pozo-Banos, J. B. Alonso, J. R. Ticay-Rivas, et al. Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41(15):6537–6554, 2014.

J. Deng, J. Guo, N. Xue, et al. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

Y. Ding, N. Robinson, S. Zhang, et al. TSception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2022.

Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pp. 1180–1189. PMLR, 2015.

A. L. Goldberger, L. A. N. Amaral, L. Glass, et al. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23): e215–e220, 2000.

X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1501–1510, 2017.

B. Jiao, L. Liu, L. Gao, et al. Dynamically transformed instance normalization network for generalizable person re-identification. In *European Conference on Computer Vision*, pp. 285–301. Springer, 2022.

M. Jin, C. Du, H. He, et al. PGCN: Pyramidal graph convolutional network for eeg emotion recognition. *IEEE Transactions on Multimedia*, 26:9070–9082, 2024a.

X. Jin, X. Yang, W. Kong, et al. TSFAN: Tensorized spatial-frequency attention network with domain adaptation for cross-session eeg-based biometric recognition. *Journal of Neural Engineering*, 21(4):046005, 2024b.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

R. J. Kobler, J. Hirayama, Q. Zhao, et al. SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg. In *Advances in Neural Information Processing Systems*, volume 35, pp. 6219–6235, 2022.

X. Kong, W. Kong, Q. Fan, et al. Task-independent eeg identification via low rank matrix decomposition. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 412–419. IEEE, 2018.

V. J. Lawhern, A. J. Solon, N. R. Waytowich, et al. EEGNet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

H. Liu, X. Jin, D. Liu, et al. Joint disentangled representation and domain adversarial training for eeg-based cross-session biometric recognition in single-task protocols. *Cognitive Neurodynamics*, 19(1):31, 2025.

M. Long, Y. Cao, J. Wang, et al. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, pp. 97–105. PMLR, 2015.

E. Maiorana and P. Campisi. Longitudinal evaluation of eeg-based biometric recognition. *IEEE Transactions on Information Forensics and Security*, 13(5):1123–1138, 2017.

R. Mane, E. Chew, K. Chua, et al. FBCNet: A multi-view convolutional neural network for brain-computer interface. *arXiv preprint arXiv:2104.01233*, 2021.

Y. Miao, W. Jiang, N. Su, et al. MLDA: Multi-loss domain adaptor for cross-session and cross-emotion eeg-based individual identification. *IEEE Journal of Biomedical and Health Informatics*, 27(12):5767–5778, 2023.

O. Özdenizci, Y. Wang, T. Koike-Akino, et al. Adversarial deep learning in eeg biometrics. *IEEE Signal Processing Letters*, 26(5):710–714, 2019.

E. Perez, F. Strub, H. De Vries, et al. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

J. Pilault, A. Elhattami, and C. Pal. Conditionally adaptive multi-task learning: Improving transfer learning in nlp using fewer parameters & less data. *arXiv preprint arXiv:2009.09139*, 2020.

K. Ranasinghe, M. Naseer, M. Hayat, et al. Orthogonal projection loss. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12313–12323. IEEE, 2021.

S.-A. Rebuffi, A. Kolesnikov, G. Sperl, et al. iCaRL: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

A. Seyfizadeh, R. L. Peach, P. Tovote, et al. Enhancing security in brain–computer interface applications with deep learning: Electroencephalogram-based user identification. *Expert Systems with Applications*, 253:124218, 2024.

J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30, 2017.

T. Song, W. Zheng, P. Song, et al. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2018.

Y. Song, Q. Zheng, B. Liu, et al. EEG Conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 710–719, 2022.

B. Sun and K. Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision*, pp. 443–450. Springer International Publishing, 2016.

A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, et al. CSPNet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390–391, 2020.

T. Xu, H. Wang, G. Lu, et al. E-Key: An eeg-based biometric authentication and driving fatigue detection system. *IEEE Transactions on Affective Computing*, 14(2):864–877, 2021.

W. L. Zheng, W. Liu, Y. Lu, et al. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, 49(3):1110–1122, 2019.

A Initialization Equivalence and the OPC Response Bound

At initialization, W_2 and b_2 in Eq. (2) are both zero, so the session scale residual and bias residual are zero. From Eqs. (4) and (5), we have

$$W_2 = 0,\ b_2 = 0 \quad \Longrightarrow \quad d_n = 0, \qquad y_n = \gamma_{\mathrm{id}} \odot \hat{x}_n + \beta_{\mathrm{id}}. \tag{10}$$

Therefore, as long as the normalization dimension, ε, and shared affine parameters are the same, the initialized module is equivalent to the corresponding LayerNorm. This result does not require the gate value itself to be zero.

For Eq. (8), expand the inner product using w=ap+e, then apply the triangle inequality and the Cauchy–Schwarz inequality:

$$\begin{aligned} |w^\top d| &\le |a|\,|p^\top d| + |e^\top d| \\ &\le |a|\eta\|d\| + \|e\|\|d\|. \end{aligned} \tag{11}$$

OPC constrains only the direction of the sample's own identity prototype and does not simultaneously constrain other classes. Take the correct-class prototype and classifier weight as (1,0), and let the perturbation and an incorrect-class weight both be (0,1). Then OPC is zero and the correct-class linear response is unchanged, while the incorrect-class logit increases by 1. This counterexample shows that zero OPC does not imply invariance of all class scores or classification margins.

The full network also contains cross-layer propagation, feature projection, and normalization. If the response of the final classification score to a perturbation is to be analyzed, one should use the perturbation after its actual propagation together with the local Jacobian of the corresponding classifier. The aggregated compensation descriptor in Eq. (6) cannot be substituted directly for that quantity. Another point is the degenerate case of zero residual or cancellation across modules: both can make OPC small, but neither means that the data contain no session drift.

B Extended Methodology and Implementation Details

B.1 Restricted-Update Protocol

In Stage 2 of the freezing experiment, gradients are disabled for the embedding, attention, feed-forward layers, shared identity affine parameters, identity head, and identity prototypes. Only the session MLP, low-rank gate, and session classifier remain trainable, and the additional prototype EMA update is disabled.

Stage 2 continues to use source-session samples with their identity and session labels. Therefore, this experiment measures restricted continued training after freezing identity-related parameters rather than adaptation using target-session samples. Freezing refers to the trainable parameter set; updates in the compensation path may still change the final identity representation.

B.2 Implementation Settings and Data Usage

Table 6 reports the training configurations used for the experiments summarized in the main text. The same Stage 1 configuration is used for the corresponding main-result and freezing experiments, so the pre-freezing accuracies in Table 5 coincide with the trained SIFR results reported in the main tables.

Table 6: Training configurations used in the reported experiments.

Setting	PhysioNet	MCD	SEED-IV
Feature Dimension D	256	256	256
Transformer Layers	6	6	4
Attention Heads	8	8	8
MLP Bottleneck r	16	16	8
Gate Rank q	16	16	8
Patch Length	8	8	1
Batch Size	16	64	128
Learning Rate	0.001	0.001	0.0005
Weight Decay	0.01	0.01	0.1
Dropout	0.2	0.1	0.3
ArcFace Scale	64.0	30.0	20.0
ArcFace Margin	0.3	0.5	0.4
Maximum Epochs	150	200	200
Warm-up Epochs	10	10	10
Early-Stopping Patience	40	60	60

The reported settings correspond to the configurations used for the results in the main text. They are listed for reproducibility and are not intended as a cross-dataset hyperparameter comparison.

SIFR increases the parameter count by 1.40%–2.04%, while session-related parameters account for 2.10%–2.33% of the full model.

The table lists settings visible in the current training entrypoints rather than back-inferring the run configuration behind every historical result. Different hyperparameters across datasets are not the same thing as a sensitivity sweep.

The PhysioNet preprocessing file groups runs 03–06, 07–10, and 11–14, but the correspondence between preprocessing output locations and training input locations still needs confirmation from

Table 7: Parameter Counts under the Same Construction Configuration

Dataset	Full Model Parameters	Baseline Parameters	Increase (%)	Session-Parameter Share (%)
PhysioNet	5,646,915	5,533,955	2.04	2.10
MCD	5,071,683	4,975,107	1.94	2.33
SEED-IV	3,469,859	3,421,955	1.40	2.23

Source: parameter counts from the existing implementation. This table measures model size, not runtime.

processing records. SEED-IV uses a sliding window of length 12 and stride 1, followed by a random validation/test split inside the target session; adjacent windows can share 11 time points. For a strictly independent evaluation, the split should first be made by trial or complete record, and windows should then be constructed, with model selection and stopping based only on data permitted by the protocol.

The freezing stage limits updating by disabling gradients for identity-related parameters and passing only session parameters to the optimizer. Stage 2 still calls the full loss, while identity prototypes no longer receive the extra EMA update. The training/evaluation mode and normalization running state in code should also be fixed explicitly in the final reproduction, so "freezing identity parameters" is not equivalent to the whole identity computation function remaining unchanged step by step.

FactorNorm initialization. The weights and bias of the final layer in the session MLP are initialized to zero, making the initial session residual exactly zero. When the shared affine parameters are initialized from the corresponding LayerNorm parameters (Ba et al., 2016), FactorNorm therefore reproduces LayerNorm exactly at initialization. The gate bias is initialized to -2, yielding an initial gate value of approximately 0.12. This value only controls the initial residual scale; the exact LayerNorm starting point is guaranteed by the zero session residual, not by treating the gate value as zero.