PHYSICAL SEMANTIC GRAPH FOR FACE ANTI-SPOOFING VIA CROSS-ATTRIBUTE RELATIONAL REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent RGB face anti-spoofing (FAS) methods have achieved remarkable progress by learning discriminative representations from various visual cues. However, most existing approaches remain predominantly feature-driven, focusing on individual cues or their fusion while paying limited attention to the relational structure among heterogeneous physical evidence. In this paper, we introduce a new perspective that characterizes facial liveness through cross-attribute consistency among complementary physical semantics. Specifically, we propose a fixed Physical Semantic Graph (PSG), where each node represents observable or estimable semantic evidence associated with a physical property of facial liveness, including texture, frequency, geometry, reflectance, and physiology-related evidence, while graph edges encode prior relational hypotheses derived from physical dependencies and complementary observations. Based on PSG, we develop a cross-attribute reasoning mechanism that propagates complementary evidence along these predefined relations to learn a relational representation of facial liveness. Furthermore, we introduce Pairwise Physical Consistency Regularization (PPCR), which explicitly encourages the relational patterns of genuine faces to be more consistent than those of presentation attacks. Extensive cross-dataset experiments on multiple public FAS benchmarks demonstrate competitive generalization.

1 INTRODUCTION

Face anti-spoofing (FAS) distinguishes genuine faces from presentation attacks such as printed photos, replayed videos, and 3D masks (Yu et al., 2022), and is essential for reliable face recognition (Ding et al., 2021). Despite strong intra-dataset performance, RGB-based FAS still generalizes poorly across sensors, illumination, attack media, and acquisition conditions (Sun et al., 2023; Zhou et al., 2023a; Le & Woo, 2024). Existing methods mainly strengthen individual cues or fuse multiple cues (Wang et al., 2022b;a; Zhou et al., 2026), emphasizing what evidence is present rather than how heterogeneous evidence should relate.

A genuine face exhibits coherent appearance, image-formation, geometric, optical, and physiology-related properties (Wang et al., 2020; Ding et al., 2026). Texture is linked to frequency-domain statistics, while reflectance is constrained by surface material and geometry (Yu et al., 2020b; Cao & Ma, 2025). Although presentation attacks may imitate individual properties, reproduction can disrupt their cross-attribute relationships. Such relations may therefore provide a more transferable liveness cue than isolated, domain-sensitive appearance features.

We do not assume that these physical properties are directly measurable from RGB images. Instead, we represent them by observable or estimable semantic evidence, including texture statistics, frequency responses, monocular geometry, reflectance cues, and physiology-related proxies. Our hypothesis is that their relational patterns provide a useful inductive bias even when individual observations are imperfect or domain-sensitive. Accordingly, we construct a fixed Physical Semantic Graph (PSG) whose topology is defined by physically motivated relational priors rather than feature similarity or arbitrary learned connectivity. Cross-attribute reasoning then propagates complementary evidence over PSG, and Pairwise Physical Consistency Regularization (PPCR) encourages genuine faces to exhibit more stable relation representations than presentation attacks.

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.

The main contributions of this work are summarized as follows: (1) We introduce a relational perspective for RGB face anti-spoofing, characterizing facial liveness through cross-attribute consistency among observable or estimable physical semantic evidence rather than relying solely on individual cues or their feature-level fusion. (2) We propose a fixed Physical Semantic Graph (PSG) that encodes physics-derived relational priors among heterogeneous semantic evidence, providing a structured inductive bias for learning liveness representations. (3) We develop a cross-attribute reasoning mechanism over PSG together with Pairwise Physical Consistency Regularization (PPCR), which explicitly encourages genuine faces to exhibit stronger relational consistency than presentation attacks. (4) Extensive cross-domain experiments demonstrate competitive FAS performance. Controlled topology and relational analyses further demonstrate the effectiveness of physically structured relational modeling over feature-similarity-based or unstructured graph construction.

2 Related Work

2.1 Face Anti-Spoofing

Face anti-spoofing protects face recognition systems from print, replay, and 3D-mask attacks (Yu et al., 2022). Early methods used handcrafted texture, motion, image-quality, and frequency cues (Chingovska et al., 2012; Wen et al., 2015), whereas CNN-based methods learned representations directly from RGB images (Yu et al., 2020b; 2021; Wang et al., 2022a; Yu et al., 2020a)and auxiliary depth, reflectance, temporal, or physiological supervision improved robustness (Liu et al., 2018). Recent work emphasizes domain generalization and domain-invariant learning (Sun et al., 2023; Zhou et al., 2023a; Le & Woo, 2024; Zhou et al., 2023b). Nevertheless, most approaches remain feature-driven, strengthening individual cues or their fusion (Wang et al., 2023; Jiang et al., 2024; Liu et al., 2025), whereas we explicitly model relations among heterogeneous physical evidence.

2.2 Physical Prior-Based Representation

Physical priors have been studied through appearance, frequency, geometry, optics, and physiological cues (Bian et al., 2022), which provide complementary observations under different attacks. Prior methods typically use them as independent modalities, auxiliary supervision, or feature/score-level fusion (Chen et al., 2025). In contrast, we represent potentially imperfect physical factors through observable or estimable semantic evidence and focus on physically motivated relations among them as a relational inductive bias for cross-domain FAS.

2.3 Graph-Based Reasoning

Graph neural networks model relations through spatial adjacency, feature similarity, learned connectivity, or task-specific structures (He et al., 2026), and face-related studies have modeled landmarks, regions, or feature interactions (Chang et al., 2023). PSG instead defines nodes as physical semantic evidence and edges as predefined relations motivated by physical dependencies and complementary observations, allowing graph reasoning to model cross-attribute relations rather than generic neighboring-feature aggregation.

3 Method

3.1 Physical Semantic Graph

Existing graph-based FAS methods typically construct graphs over spatial regions or learned feature embeddings, where graph connectivity is determined by spatial adjacency, feature similarity, or dynamically learned correlations. While such graphs facilitate feature aggregation, they do not explicitly distinguish relations arising from dataset-dependent feature correlations from those motivated by prior knowledge of facial liveness. We instead consider a complementary perspective: although the underlying physical properties may only be imperfectly observed from RGB inputs, relations among their semantic evidence can provide a structured inductive bias for distinguishing genuine faces from presentation attacks. The overall motivation and construction of the proposed

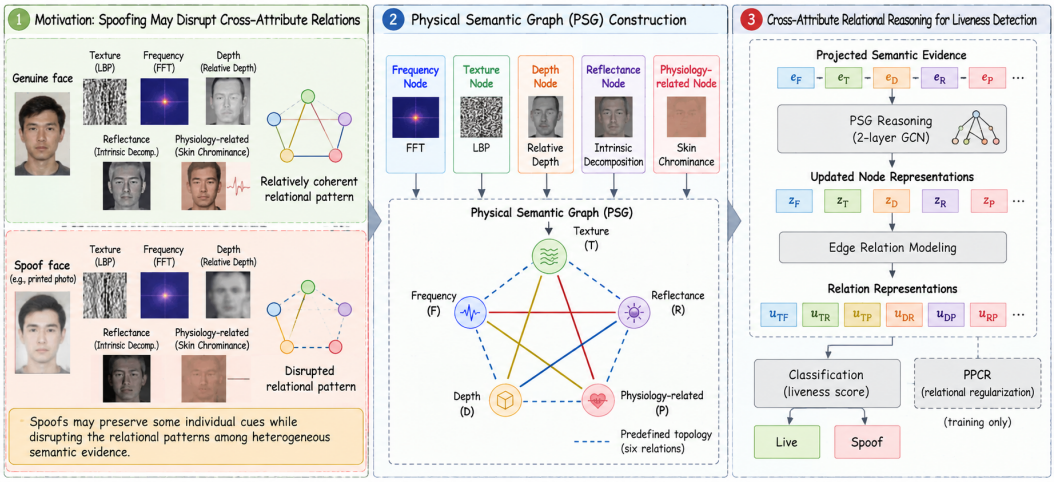

Figure 1: Overview of the proposed framework. Five physical semantics form a fixed PSG with predefined relations. Cross-attribute reasoning and relation modeling produce liveness representations, while PPCR regularizes edge relations during training.

PSG is illustrated in Fig. 1. The proposed PSG consists of five physical semantic nodes, corresponding to five complementary physical dimensions involved in facial liveness analysis.

3.1.1 Physical Semantic Evidence

We define a Physical Semantic Graph (PSG) as

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \tag{1}$$

where

$$\mathcal{V} = \{v_t, v_f, v_d, v_r, v_p\} \tag{2}$$

where V contains *texture*, *frequency*, *depth*, *reflectance*, *and physiology-related* semantics. Each node represents observable or estimable evidence rather than a directly measured physical variable. Given an aligned image x, a semantic-specific descriptor $h_i(x)$ is projected into a common d-dimensional space:

$$e_i = P_i\big(h_i(x)\big), \qquad i \in \{t, f, d, r, p\}, \tag{3}$$

where $P_i(\cdot)$ is a semantic-specific projection module and $e_i \in \mathbb{R}^d$ is the resulting semantic representation.

Texture Semantic (v_t). We compute uniform LBP with $P = 8$ and $R = 1$ for each aligned face and use the normalized $LBP_{8,1}$ histogram as $h_t(x)$ to characterize local appearance and micro-texture.

Frequency Semantic (v_f). The aligned RGB face is converted to grayscale and transformed by a 2-D FFT. The centered log-magnitude spectrum is computed as

$$F(x) = \log\left(1 + |\mathrm{fftshift}\left(\mathcal{F}_2(x_g)\right)|\right), \tag{4}$$

where $\mathcal{F}_2(\cdot)$ denotes the 2-D FFT. The standardized spectrum is processed by a frequency encoder to obtain $h_f(x)$.

Depth Semantic (v_d). A frozen pretrained MiDaS-Small model estimates the relative-depth map

$$D_x = \mathcal{D}_{\mathrm{pre}}(x), \tag{5}$$

which is normalized and processed by a depth encoder to obtain $h_d(x)$.

Reflectance Semantic (v_r). A frozen pretrained PIE-Net intrinsic decomposition model Das et al. (2022) extracts a reflectance/albedo map:

$$R_x = \mathcal{R}_{\mathrm{pre}}(x). \tag{6}$$

The resulting map is normalized and processed by a reflectance encoder to produce $h_r(x)$.

Physiology-Related Semantic (v_p). Because physiological dynamics cannot be directly measured from a single RGB frame, we use skin-region chrominance as a physiology-related appearance proxy. After excluding non-skin regions, normalized chrominance is computed as

$$r = \frac{R}{R+G+B+\epsilon}, \qquad g = \frac{G}{R+G+B+\epsilon}, \tag{7}$$

where ϵ ensures numerical stability. Statistics over the skin region form $h_p(x)$.

All descriptors are mapped to the same d-dimensional space. Frequency, depth, and reflectance use encoders before projection, whereas LBP and chrominance descriptors are directly projected by MLPs. We use $d = 128$ throughout.

3.1.2 Physics-Derived Graph Topology

We construct a sparse fixed topology from explicit relational hypotheses rather than feature similarity. The edge set is

$$\begin{aligned}\mathcal{E} = \{&(T,F),(T,R),(T,P),\\ &(D,R),(D,P),(R,P)\},\end{aligned} \tag{8}$$

with adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{9}$$

Self-connections are subsequently introduced during graph propagation by $\widetilde{A} = A + I$.

The six edges capture three types of priors. (T, F) links spatial texture and frequency-domain reproduction evidence. (T, R), (T, P), and (R, P) connect surface appearance, intrinsic reflectance, and skin chromatic observations. (D, R) and (D, P) cross-check geometry against material and skin-related appearance. These relations allow heterogeneous evidence to verify one another when an attack preserves some individual cues but disrupts their joint configuration.

We keep PSG sparse: (F, D), (F, R), (F, P), and (T, D) are excluded because we do not impose sufficiently direct relational hypotheses for them. Thus, each edge corresponds to an explicit testable prior rather than generic multi-cue aggregation. The topology is fixed across datasets and attack types and is evaluated against alternative constructions in Sec. 4.4.

3.2 Cross-Attribute Consistency Reasoning

PSG specifies which semantic evidence may interact, while the reasoning module models sample-dependent relations on this fixed structure. We hypothesize that genuine faces exhibit more stable cross-attribute patterns across domains, whereas presentation attacks perturb part of these relations. Consistency is therefore modeled at the relation level rather than by individual feature magnitude.

3.2.1 Relational Representation on PSG

For input x, the five projected semantics form

$$E^{(0)} = \begin{bmatrix} e_t^\top \\ e_f^\top \\ e_d^\top \\ e_r^\top \\ e_p^\top \end{bmatrix} \in \mathbb{R}^{5\times d}, \tag{10}$$

Using the adjacency matrix in Eq. (9), we add self-connections

$$\widetilde{A} = A + I, \tag{11}$$

and compute the symmetric normalization

$$\widehat{A} = \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}}, \tag{12}$$

where $\widetilde{D}$ is the degree matrix of $\widetilde{A}$.

Cross-attribute reasoning is then performed by graph convolution:

$$E^{(l+1)} = \sigma\left(\widehat{A} E^{(l)} W^{(l)}\right), \qquad l = 0, \ldots, L-1, \tag{13}$$

where $W^{(l)}$ is the learnable transformation and $\sigma(\cdot)$ the activation; we use $L = 2$ layers.

Equivalently, the update for node i is

$$e_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \widehat{A}_{ij} e_j^{(l)} W^{(l)}\right), \tag{14}$$

where $\mathcal{N}(i)$ is defined by the fixed PSG topology. Each node therefore aggregates only itself and semantically connected factors, preventing unrestricted all-to-all fusion.

3.2.2 From Semantic Features to Cross-Attribute Relations

The initial e_i describes an individual semantic factor, whereas consistency is represented by relations between connected factors. After graph reasoning, the contextualized node representation is

$$z_i = e_i^{(L)}. \tag{15}$$

For each edge, we define

$$q_{ij} = \phi\left(z_i, z_j\right), \qquad (i,j) \in \mathcal{E}, \tag{16}$$

and instantiate the relation operator as

$$q_{ij} = \left[z_i \,\|\, z_j \,\|\, |z_i - z_j| \,\|\, z_i \odot z_j\right], \tag{17}$$

combining the two node features, their absolute difference, and element-wise interaction. A shared projector then maps these interactions to a common relation space

$$u_{ij} = \rho(q_{ij}), \qquad (i,j) \in \mathcal{E}, \tag{18}$$

The resulting u_{ij} is used for both classification and relation-level regularization and represents how two heterogeneous semantic factors coexist, not whether their values are numerically identical.

3.2.3 Graph-Level Relational Representation

Edge relations are averaged to form the graph-level relational representation

$$c(x) = \frac{1}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} u_{ij}. \tag{19}$$

while node representations are averaged by

$$g(x) = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} z_i \tag{20}$$

Their concatenation is classified by

$$\hat{y} = f_{\text{cls}}\left(g(x) \| c(x)\right), \tag{21}$$

Thus, prediction uses both individual semantic evidence and the joint pattern of predefined cross-attribute relations.

3.3 Consistency-Guided Optimization

Classification alone does not explicitly encourage edge relations to encode liveness-related regularity. We therefore introduce Pairwise Physical Consistency Regularization (PPCR), which encourages genuine faces to exhibit more stable same-relation representations than presentation attacks.

3.3.1 CROSS-SAMPLE RELATIONAL SIMILARITY

For two observations a and b, the similarity between representations of the same PSG edge is measured by cosine similarity:

$$s_{ij}^{a,b} = \frac{(u_{ij}^a)^\top u_{ij}^b}{\|u_{ij}^a\|_2 \|u_{ij}^b\|_2}. \tag{22}$$

This similarity serves as a training surrogate for relational regularity rather than a direct physical measurement.

3.3.2 PAIRWISE PHYSICAL CONSISTENCY REGULARIZATION

PPCR uses genuine–genuine pairs as positives and genuine–spoof pairs as negatives. For each edge (i, j), a triplet (a, p, n) contains two genuine samples a, p and one spoof sample n. The edge-wise loss is

$$\mathcal{L}_{ij}^{\mathrm{PPCR}} = \left[m - s_{ij}^{a,p} + s_{ij}^{a,n} \right]_+, \tag{23}$$

where $[z]_+ = \max(0, z)$ and $m > 0$ is the margin, requiring

$$s_{ij}^{a,p} \geq s_{ij}^{a,n} + m. \tag{24}$$

Thus, the same relation is encouraged to be more stable across genuine observations. We normalize over all predefined edges:

$$\mathcal{L}_{\mathrm{PPCR}} = \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} \mathbb{E}_{(a,p,n)} \left[\mathcal{L}_{ij}^{\mathrm{PPCR}} \right]. \tag{25}$$

For PSG, $|E| = 6$, and the same normalization is used for all topology variants.

3.3.3 CLASSIFICATION AND OVERALL OBJECTIVE

Given $y \in \{0, 1\}$, the classification loss is

$$\mathcal{L}_{\mathrm{cls}} = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}). \tag{26}$$

The complete training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \lambda_{\mathrm{ppcr}} \mathcal{L}_{\mathrm{PPCR}}, \tag{27}$$

where λ_{ppcr} controls the regularization strength. PPCR is used only during training and introduces no cross-sample computation at inference.

4 EXPERIMENT

4.1 EXPERIMENTAL SETTINGS

Datasets. We use five public face anti-spoofing benchmarks: CASIA-FASD, Replay-Attack, MSU-MFSD, OULU-NPU, and SiW Liu et al. (2018). Replay-Attack contains print and replay attacks collected under controlled and adverse illumination conditions Chingovska et al. (2012). MSU-MFSD contains print and replay attacks captured using different mobile devices Wen et al. (2015), while OULU-NPU provides four official protocols covering variations in illumination, acquisition devices, and attack media Boulkenafet et al. (2017). The standard MCIO leave-one-domain-out evaluation uses CASIA-FASD, Replay-Attack, MSU-MFSD, and OULU-NPU; SiW is used for the consistency-score analysis in Fig. 2.

Evaluation Metrics. For all evaluations, we follow the standard domain generalization protocol in face anti-spoofing and report Half Total Error Rate (HTER) and Area Under the ROC Curve (AUC). Lower HTER and higher AUC indicate better cross-domain generalization performance.

Table 1: Cross-domain FAS comparison under the MCIO leave-one-out protocol. Entries report HTER(%)↓/AUC(%)↑; M, C, I, and O denote the unseen target domains.

Method	Venue	Type	M	C	I	O
CDCN† Yu et al. (2020b)	CVPR'20	CNN	12.5 / 95.1	20.4 / 90.3	15.7 / 93.4	18.5 / 91.8
ANRLLiu et al. (2021)	CVPR'21	DG	10.6 / 95.7	17.8 / 92.4	16.0 / 93.2	15.7 / 93.6
SSAN-RWang et al. (2022c)	CVPR'22	DG	7.7 / 94.8	10.0 / 96.7	8.9 / 96.8	13.7 / 93.6
TransFASWang et al. (2022b)	TIFS'23	Transformer	9.4 / 97.6	9.8 / 96.2	10.1 / 95.9	15.5 / 93.0
DiVT-MLiao et al. (2023)	ICCV'23	Transformer	8.9 / 96.1	10.7 / 96.9	**4.7 / 98.3**	13.1 / 94.0
SA-FASSun et al. (2023)	CVPR'23	DG	6.7 / 96.6	8.8 / 95.4	6.6 / 97.5	10.0 / 96.2
GAC-FASLe & Woo (2024)	CVPR'24	DG	7.1 / 97.6	9.2 / 95.2	4.3 / 98.9	8.6 / 97.2
CFPLLiu et al. (2024)	CVPR'24	Foundation	**5.3** / 97.2	8.6 / 95.1	8.4 / 97.7	8.50 / 96.3
I-FASZhang et al. (2025)	AAAI'25	Foundation	5.4 / 97.8	7.4 / 97.0	5.6 / 97.8	8.7 / 97.5
FGDA-APLYang et al. (2026)	TIFS'26	DA	5.8 / 97.6	9.2 / 97.2	11 / 96.3	7.9 / 97.6
Ours	–	**PSG**	5.6/**98.2**	**7.1 / 97.8**	5.2/97.8	**7.3/99.0**

4.2 Implementation Details

All face images are aligned, cropped, and resized to 256×256. We extract five types of semantic evidence for PSG construction: uniform LBP ($P = 8$, $R = 1$) for texture, FFT spectrum for frequency, a frozen pretrained MiDaS-Small model for relative depth, a frozen pretrained PIE-Net for reflectance, and normalized chrominance statistics from facial skin regions for physiology-related appearance. The extracted descriptors are mapped into a unified semantic space through semantic encoders.

The PSG topology is fixed for all datasets and contains the six predefined relations described before. Cross-attribute reasoning is performed using a two-layer GCN, followed by a MLP for relational representation and classification. PPCR is jointly optimized with the classification loss, with the margin m and regularization weight λ_{ppcr} set to 0.2 and 0.1, respectively.

We train the model using AdamW with an initial learning rate of 1×10^{-4} and a batch size of 64 for 50 epochs. All experiments are conducted on an NVIDIA RTX 4090 GPU. For cross-dataset evaluation, model training and hyperparameter selection are performed only on the source datasets without access to the target-domain labels. Each experiment is repeated with at least three random seeds, and the average results are reported.

4.3 Comparison with State-of-the-Art

Table 1 compares PSG with representative CNN-based, domain-generalization, Transformer-based, and foundation-model-based face anti-spoofing methods under the standard leave-one-out cross-dataset protocol. PSG achieves competitive or superior performance across the evaluated domains.

The results demonstrate that explicitly modeling relations among heterogeneous semantic evidence is effective for cross-domain face anti-spoofing. Unlike methods that primarily improve individual feature representations or learn domain-invariant features, our method organizes complementary semantic evidence according to predefined physical relations and performs reasoning over their interactions. This provides a fixed relational inductive bias shared across different target-domain evaluations and leads to consistently strong cross-dataset performance.

It is worth noting that the four target datasets exhibit different capture conditions and presentation attack characteristics. The consistent performance across these protocols suggests that the benefit of PSG is not restricted to a particular dataset. We further investigate whether this generalization originates from the proposed semantic relations, rather than merely from graph-based feature aggregation, in the following topology and relational analyses.

Table 2: Graph-topology comparison under I+C+M→O. Random and KNN use six edges; PSG+Shuffled preserves connectivity but permutes semantic identities.

Topology	#Edges	HTER(%)↓	AUC(%)↑
Fully Connected	10	12.7	92.0
Random	6	13.0	90.7
KNN	6	15.2	91.4
PSG + Shuffled	6	9.5	94.5
PSG (Ours)	6	**7.3**	**99.0**

Table 3: Contribution of individual physical semantic nodes under the $I + C + M \rightarrow O$ cross-dataset protocol. One semantic node is removed at a time together with its associated relations.

Configuration	**HTER(%)↓**	**AUC(%)↑**
w/o Texture	10.6	96.9
w/o Frequency	9.3	98.1
w/o Depth	9.7	97.3
w/o Reflectance	10.2	96.7
w/o Physiology-related	10.8	96.2
Complete PSG	**7.3**	**99.0**

Table 4: Component ablation under I+C+M→O. PSG, relation modeling (RM), and PPCR are progressively added to identical semantic descriptors.

Model	**PSG**	**RM**	**PPCR**	**HTER(%)↓**	**AUC(%)↑**
Baseline	✗	✗	✗	12.6	94.7
+ PSG	✓	✗	✗	9.1	96.9
+ RM	✓	✓	✗	8.2	97.7
+ PPCR	✓	✓	✓	**7.3**	**99.0**

4.4 Validation of Physical Semantic Graph Design

To isolate the effect of the physics-derived topology, we compare PSG with Fully Connected, edge-count-matched Random, and feature-similarity KNN graphs. Random and KNN use six undirected edges, with Random averaged over five sampled topologies. PSG+Shuffled retains the same six-edge connectivity but randomly permutes semantic node identities. All variants share identical semantic evidence, reasoning modules, and optimization.

As shown in Table 2, PSG outperforms all alternatives. Its advantage over edge-count-matched Random and KNN graphs shows that sparsity alone is insufficient, while the degradation of PSG+Shuffled indicates that performance depends on the correspondence between semantic attributes and predefined relations rather than connectivity alone.

4.5 Contribution of Physical Semantic Evidence

We remove each semantic node and its incident edges while keeping all other settings unchanged. As shown in Table 3, removing any node degrades performance, confirming the complementarity of the five semantics. Texture and physiology-related evidence cause relatively larger drops, while frequency, depth, and reflectance also provide consistent gains. Thus, the complete model does not rely on a single dominant cue.

4.6 Framework Ablation

We isolate the contributions of PSG, relation modeling (RM), and PPCR. The non-relational baseline directly averages the same five semantic embeddings:

$$g_{\text{base}}(x) = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} e_i, \qquad \hat{y} = f_{\text{cls}}(g_{\text{base}}(x)) \,. \tag{28}$$

Thus, all variants use identical semantic evidence. As shown in Table 4, PSG, RM, and PPCR progressively improve performance: PSG introduces structured semantic interactions, RM explicitly models edge-level relations, and PPCR regularizes their relational patterns. Their complementary gains show that the improvement does not arise from multi-cue aggregation alone.

4.7 Analysis of Physical Semantic Relations

To inspect the learned relation space, we compute on SiW the cosine similarity between the same edge representations from genuine–genuine (G–G) and genuine–spoof (G–S) pairs:

$$s_{ij}^{a,b} = \frac{(u_{ij}^{a})^{\top} u_{ij}^{b}}{\|u_{ij}^{a}\|_2 \|u_{ij}^{b}\|_2}. \tag{29}$$

The resulting similarities are aggregated over the predefined relations to analyze their overall distributions.

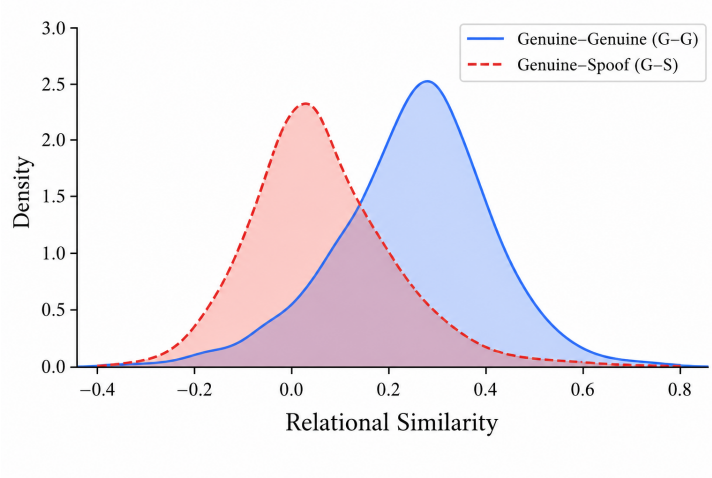

Figure 2: Cross-sample relational similarity on SiW for genuine–genuine (G–G) and genuine–spoof (G–S) pairs.

Similarities are aggregated over all predefined relations. As shown in Fig. 2, G–G pairs exhibit higher similarity than G–S pairs, indicating more stable same-relation representations among genuine observations. This separation is consistent with the intended effect of PPCR.

5 Conclusion

We presented a relational approach to cross-domain FAS that organizes texture, frequency, depth, reflectance, and physiology-related evidence into a fixed Physical Semantic Graph. Cross-attribute reasoning models predefined semantic interactions, while PPCR encourages stable genuine-face relation representations. Cross-domain, topology, node, and component analyses show that the gains arise from physically motivated relational structure rather than generic graph connectivity or a single cue. These results support relational modeling of heterogeneous physical evidence as a promising direction for generalizable face anti-spoofing.

AI USE STATEMENT

Generative AI tools were used solely to assist with language polishing and improving the clarity, conciseness, and readability of the manuscript. They were not used to formulate the research methodology, design experiments, generate or modify experimental data, or interpret experimental results. All technical content, experimental results, analyses, and conclusions were produced and verified by the authors. The authors have reviewed the AI-assisted edits and take full responsibility for the final content of the paper.

REFERENCES

Ying Bian, Peng Zhang, Jingjing Wang, Chunmao Wang, and Shiliang Pu. Learning multiple explainable and generalizable cues for face anti-spoofing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2310–2314. IEEE, 2022.

Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pp. 612–618. IEEE, 2017.

Junyi Cao and Chao Ma. Towards generalized face anti-spoofing from a frequency shortcut view. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1005–1015. IEEE, 2025.

Chih-Jung Chang, Yaw-Chern Lee, Shih-Hsuan Yao, Min-Hung Chen, Chien-Yi Wang, Shang-Hong Lai, and Trista Pei-Chun Chen. A closer look at geometric temporal dynamics for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1081–1091, 2023.

Geng Chen, Wuyuan Xie, Di Lin, Ye Liu, and Miaohui Wang. mmfas: Multimodal face anti-spoofing using multi-level alignment and switch-attention fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 58–66, 2025.

Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pp. 1–7. IEEE, 2012.

Partha Das, Sezer Karaoglu, and Theo Gevers. Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19758–19767. IEEE, 2022.

Feng Ding, Guopu Zhu, Yingcan Li, Xinpeng Zhang, Pradeep K Atrey, and Siwei Lyu. Anti-forensics for face swapping videos via adversarial training. *IEEE Transactions on Multimedia*, 24:3429–3441, 2021.

Feng Ding, Wenhui Yi, Yunpeng Zhou, Xinan He, Hong Rao, and Shu Hu. Decoupling bias, aligning distributions: Synergistic fairness optimization for deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 32704–32713, 2026.

Xinan He, Yue Zhou, Bing Fan, Bin Li, Guopu Zhu, and Feng Ding. Vlforgery face triad: Detection, localization and attribution via multimodal large language models. *Advances in Neural Information Processing Systems*, 38:163010–163044, 2026.

Fangling Jiang, Yunfan Liu, Haolin Si, Jingjing Meng, and Qi Li. Cross-scenario unknown-aware face anti-spoofing with evidential semantic consistency learning. *IEEE Transactions on Information Forensics and Security*, 19:3093–3108, 2024.

Binh M Le and Simon S Woo. Gradient alignment for cross-domain face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 188–199, 2024.

Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen. Domain invariant vision transformer learning for face anti-spoofing. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 6098–6107, 2023.

Ajian Liu, Shuai Xue, Jianwen Gan, Jun Wan, Yanyan Liang, Jiankang Deng, Sergio Escalera, and Zhen Lei. Cfpl-fas: Class free prompt learning for generalizable face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 222–232, 2024.

Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM international conference on multimedia*, pp. 1469–1477, 2021.

Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 389–398, 2018.

Yongluo Liu, Zun Li, and Lifang Wu. Dual consistency regularization for generalized face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 2025.

Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24563–24574, 2023.

Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20281–20290, 2022a.

Zezheng Wang, Zitong Yu, Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou, Feng Zhou, and Zhen Lei. Deep spatial gradient and temporal depth learning for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5042–5051, 2020.

Zezheng Wang, Zitong Yu, Xun Wang, Yunxiao Qin, Jiahong Li, Chenxu Zhao, Xin Liu, and Zhen Lei. Consistency regularization for deep face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 18:1127–1140, 2023.

Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guodong Guo. Face anti-spoofing using transformers with relation-aware mechanism. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):439–450, 2022b.

Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4123–4133, 2022c.

Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

Jing Yang, Xusheng Cui, Yuehai Chen, Shaoyi Du, Badong Chen, and Yuewen Liu. Fine-grained domain alignment for face anti-spoofing with asymmetric pseudo-labels. *IEEE Transactions on Information Forensics and Security*, 2026.

Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Face anti-spoofing with human material perception. In *European conference on computer vision*, pp. 557–575. Springer, 2020a.

Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao. Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5295–5305, 2020b.

Zitong Yu, Yunxiao Qin, Hengshuang Zhao, Xiaobai Li, and Guoying Zhao. Dual-cross central difference network for face anti-spoofing. *arXiv preprint arXiv:2105.01290*, 2021.

Zitong Yu, Yunxiao Qin, Xiaobai Li, Chenxu Zhao, Zhen Lei, and Guoying Zhao. Deep learning for face anti-spoofing: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45 (5):5609–5631, 2022.

Guosheng Zhang, Keyao Wang, Haixiao Yue, Ajian Liu, Gang Zhang, Kun Yao, Errui Ding, and Jingdong Wang. Interpretable face anti-spoofing: Enhancing generalization with multimodal large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 9896–9904, 2025.

Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20453–20463, 2023a.

Yue Zhou, Bing Fan, Pradeep K. Atrey, and Feng Ding. Exposing deepfakes using dual-channel network with multi-axis attention and frequency analysis. In *Proceedings of the 2023 ACM Workshop on Information Hiding and Multimedia Security*, pp. 169–174, 2023b.

Yue Zhou, Xinan He, KaiQing Lin, Bing Fan, Feng Ding, and Bin Li. Breaking latent prior bias in detectors for generalizable aigc image detection. *Advances in Neural Information Processing Systems*, 38:30649–30679, 2026.