ADAPTIVE CONNECTIVITY IS NOT CORRESPONDENCE: A CROSS-POSE STUDY OF VISION GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision graph networks dynamically connect image regions through input-dependent edges, yet it remains unclear whether these learned connections capture true cross-view correspondence. We investigate this distinction using cross-pose face recognition as a controlled testbed. While achieving high frontal accuracy, the tested vision-graph models underperform a convolutional baseline on frontal-profile verification. Through controlled interventions, we find that grid anchoring emerges at an intermediate stage, while features at supplied correspondences retain a measurable advantage over identity-substituted controls. A geometric bound on cell-constrained sampling reach, however, does not imply a bound on recognition capacity. Under matched data and training budgets, pose augmentation alone outperforms adding the tested explicit correspondence alignment. Under a full 20-epoch schedule, pose augmentation improves extreme-pose genuine acceptance from 86.05% to 92.02 $\pm$ 0.99%. These recognition gains replicate across training datasets, whereas connectivity patterns are dataset-dependent, and the augmented models become less robust to pixel-noise corruptions. Together, our results show that adaptive connectivity, cross-view correspondence, and downstream recognition require distinct evidence and should be evaluated separately.

1 INTRODUCTION

Learning representations that are invariant to extreme viewpoint changes remains a fundamental challenge in face recognition, primarily due to non-rigid deformations and self-occlusions. To investigate how vision graph architectures handle these spatial misalignments, we introduce *GraphFace*, a model built upon the Pyramid Vision Graph Neural Network (PViG-S) (Han et al., 2022). Unlike rigid grid-based convolutional operations, GraphFace formulates the image as a graph where nodes exchange information dynamically via feature-dependent edges before being aggregated into a global identity embedding.

A natural reading of vision graph networks is that adaptive connectivity—the learned mechanism determining which regions exchange information—also captures cross-view correspondence, associating the same physical surface point across views. Whether that reading holds is not established by within-image content linking, and we test it here. However, dynamically updating graph edges does not inherently guarantee geometric correspondence, nor is explicit surface matching strictly required for successful identity recognition. Disentangling these mechanisms is critical for understanding what vision graphs actually learn, and they are not always distinguished.

We utilize aligned faces to provide a highly controlled spatial testbed: while preprocessing anchors landmarks to standard image coordinates, out-of-plane rotations systematically displace visible content and introduce occlusion. Under a shared training protocol, GraphFace operates in a highly efficient regime (1.49 GMACs compared to 6.33 GMACs for a standard face-specific ResNet-50 baseline). Although GraphFace trails the computationally heavier baseline by a marginal 0.70 percentage points in mean accuracy across seven benchmarks, it exhibits a more pronounced deficit in extreme cross-pose scenarios.

In this work, we systematically decouple network connectivity, geometric correspondence, and downstream recognition. Our contributions are threefold: (i) Controlled Interventional Analyses:

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.

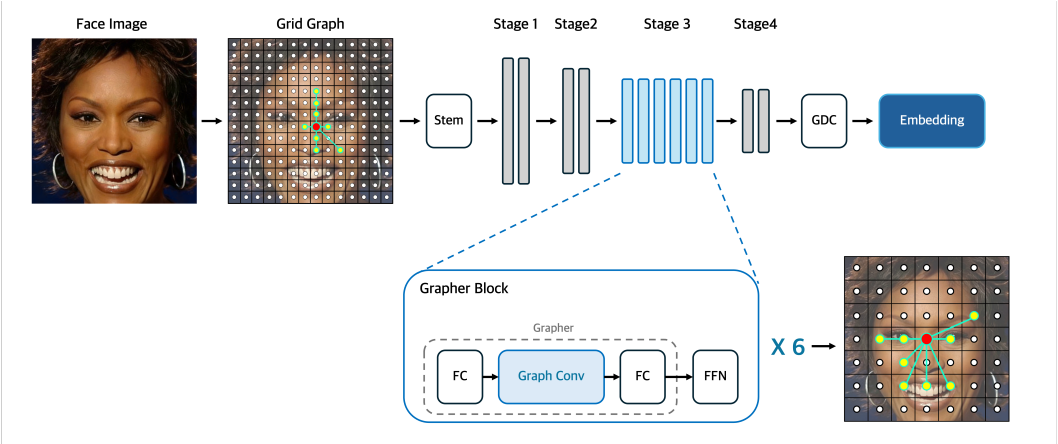

Figure 1: **GraphFace architecture.** An aligned face image passes through a convolutional stem, four graph-processing stages, and a global depthwise convolution (GDC) readout. The output is an identity embedding. The inset illustrates feature-based neighbor selection within an image; cross-view correspondence is evaluated separately.

We isolate spatially anchored connectivity from true feature correspondence, demonstrating how grid preferences emerge at an intermediate stage; (ii) Rigorous Budget-Matched Comparisons: We compare pose augmentation against an explicit correspondence-alignment objective under matched data and training budgets; in the tested setting, pose augmentation alone accounts for the extreme-yaw gains; (iii) Geometric and Evaluation Controls: We separate cell-constrained sampling reach from actual recognition capacity, delimiting the claims that can be made about adaptive structures.

Crucially, while the recognition gains replicate across training datasets, the connectivity patterns are dataset-dependent: grid anchoring at extreme yaw replicates, but the yaw-averaged pattern does not.

2 GraphFace: Architecture and Representation

Graph construction and feature aggregation. GraphFace augments PViG-S with parametric rectified linear units (PReLU) and a global depthwise convolution (GDC) readout (Han et al., 2022; He et al., 2015; Chen et al., 2018). The convolutional stem maps a 112×112 image to a 28×28 feature grid. Subsequent stages contain 28^2, 14^2, 7^2, and 4^2 nodes, respectively; we identify stages by these resolutions. For node i, let r_i denote its grid position and x_i its feature vector. Feature and spatial distances, d_f and d_s, define the neighbor-selection rule:

$$X^{(0)} = \mathrm{Stem}(I) + P_{\mathrm{abs}}, \qquad d_f(i,j) = \|x_i - x_j\|_2, \quad d_s(i,j) = \|r_i - r_j\|_2, \tag{1}$$

$$d_\alpha(i,j) = \begin{cases} d_f(i,j) + e^{\mathrm{rel}}_{ij}, & \alpha = 0, \\ (1-\alpha)\widetilde{d}_f(i,j) + \alpha \widetilde{d}_s(i,j) + e^{\mathrm{rel}}_{ij}, & \alpha > 0, \end{cases} \tag{2}$$

$$\mathcal{N}_\alpha(i) = \arg\operatorname*{topk}_j \{-d_\alpha(i,j)\}, \qquad k = 9. \tag{3}$$

The self term has zero feature distance and the smallest relative-position term, so it is always selected and each node has eight distinct neighbors. Here P_{abs} and e^{rel}_{ij} encode absolute and relative position, respectively, and a tilde denotes instance-wise min–max normalization. Equation (3) selects the nine nearest neighbors. GraphFace-V uses $\alpha = 0$, while GraphFace-H combines feature and spatial distances with $\alpha = 0.5$. Both variants retain positional information. Because the two branches also differ in distance normalization, varying α does not isolate the effect of positional information alone.

A Grapher block aggregates the node feature and its differences from neighboring features. In the simplified expression below, the maximum is channel-wise, $\|$ denotes concatenation, the W matrices are learned projections, and σ is the activation:

$$x'_i = x_i + W_{\mathrm{out}}\, \sigma\left(W_u \left[W_{\mathrm{in}} x_i \,\|\, \max_{j \in \mathcal{N}_\alpha(i)} W_{\mathrm{in}}(x_j - x_i) \right] \right). \tag{4}$$

Table 1: **Frontal and cross-pose verification.** LFW and CFP-FP report ten-fold accuracy (%). Higher d' indicates greater separation between genuine and impostor similarity scores. The tested graph models underperform R50 on frontal–profile verification. For ViHGNN, accuracy is from a single fixed-batch evaluation and d' from three evaluation seeds ($\pm$ is their SD), not independent training runs. ViT-S is an unadapted patch transformer trained under the same recipe without face-specific patch or positional priors; its cross-pose figure should be read accordingly (Table 11).

Backbone	Params (M)	GMACs	LFW	CFP-FP	CFP-FP d'
IResNet-50	43.59	6.334	99.78	97.67	4.045
GraphFace-V	31.25	1.493	99.72	95.77	3.558
GraphFace-H	31.25	1.493	99.75	95.74	3.576
ViHGNN-S	25.27	1.175	99.80	95.64	3.321 ± 0.006
GreedyViG-S	10.35	0.433	99.55	93.76	3.124
ViT-S	29.03	1.088	99.45	90.11	2.524

Each block also includes a feed-forward network (FFN). The final GDC readout pools spatial features using channel-specific weights $w_{uv}^{(c)}$, and a projection yields the normalized 512-dimensional identity embedding z:

$$y_c = \sum_{u,v} w_{uv}^{(c)} X_{uv}^{(c)}, \qquad z = \frac{W_e y}{\|W_e y\|_2} \in \mathbb{R}^{512}. \tag{5}$$

ArcFace supervises identity discrimination by imposing an angular margin between classes (Deng et al., 2019).

Spatial dependence of the readout. GDC assigns distinct weights to spatial positions. Consequently, relocating a feature can change the readout even when graph connections transform consistently. Appendix E derives the exact invariance condition. This result motivates a controlled comparison with average pooling, but does not identify GDC as the cause of the observed recognition deficit.

3 Cross-pose recognition and correspondence analysis

Training and evaluation. The main models share the MS1MV2 training dataset, ArcFace loss ($m = 0.5, s = 64$), stochastic gradient descent, batch size 512, twenty epochs, and seed 2026. Evaluation uses batch size 128, 32-bit floating point, and a common implementation. Table 1 compares three graph architectures with a convolutional neural network (CNN) and a vision transformer (ViT) (Han et al., 2022; 2023; Munir et al., 2024). This protocol controls training conditions rather than model capacity or architecture-specific optimization. Appendix B reports the complete benchmarks and timing measurements.

We complement accuracy with two measures. The first, d', quantifies separation between the similarity-score distributions of genuine pairs (same identity, $+$) and impostor pairs (different identities, $-$). For score means μ and variances σ^2, it is

$$d' = \frac{\mu_+ - \mu_-}{\sqrt{(\sigma_+^2 + \sigma_-^2)/2}}. \tag{6}$$

This statistic is independent of a decision threshold. The second measure is the genuine acceptance rate (GAR), the fraction of same-identity pairs accepted, stratified by *yaw*, or horizontal head-rotation angle. Each model uses its own full-set equal-error-rate (EER) threshold, at which false acceptance and false rejection rates coincide. These GAR values are neither within-bin accuracies nor comparisons at a shared false-acceptance rate. On CFP-FP (Sengupta et al., 2016), V trails R50 by 0.39, 1.10, and 7.53 pp for 30–50°, 50–70°, and $> 70°$, respectively. The final group contains 810 genuine pairs. The results locate the largest deficit at extreme yaw without establishing a sharp transition at 70°.

Distinct levels of analysis. We distinguish graph connectivity, local feature correspondence, sampling reach, and identity discrimination (Table 2).

Table 2: **Scope of the four measurements.** Connectivity, feature correspondence, and recognition are assessed empirically; sampling reach is characterized under an explicit geometric constraint.

Object	Measurement	Interpretation
Connections	Graph-overlap difference Δ_{anchor}	Fixed-position versus correspondence-based edge agreement
Local features	Controlled feature advantage D_{corr}	Feature advantage at supplied correspondences over an identity-substituted control
Sampling location	Minimum target error e_i^*	Target reachability under the original-cell constraint
Identity vector	Accuracy, d', yaw-group GAR	Identity discrimination across viewpoints

Dependence on learned connectivity. Varying α over $\{0, 0.5, 1\}$ changes the earlier five-benchmark mean by only 0.22 pp. In contrast, replacing learned neighbors with random neighbors at evaluation reduces CFP-FP and LFW accuracy by 27.74 and 1.63 pp, respectively. This intervention demonstrates reliance on learned connectivity but does not explain the original cross-pose deficit. Despite different edge patterns, V and H have similar intermediate representations: matched-layer centered kernel alignment (CKA), a measure of representational similarity, is 0.95–0.98 (Kornblith et al., 2019). Appendix C retains the original qualitative comparison.

3.1 Grid anchoring versus correspondence-based edge transport

We fit a three-dimensional face model to 298 frontal CFP images (300 sampled from the frontal-side slots with detection success and $|\text{yaw}| \leq 10^\circ$; two failed face detection) and render each face at yaw angles $\{0, 15, 30, 45, 60, 75\}^\circ$ (Blanz & Vetter, 1999; Guo et al., 2020). The renderer supplies correspondence maps and visibility information. Each rotated view is compared with a 0° render, controlling for the rendering process. For each node i visible at 0°, let $N_0(i)$ and $N_\theta(\cdot)$ denote the k-nearest-neighbor sets ($k = 9$, including the node itself) at 0° and at yaw θ, and let m_θ be the dense correspondence. The grid hypothesis predicts that node i keeps its frontal neighbors, $\hat{N}_{\text{grid}}(i) = N_0(i)$; the content hypothesis predicts that the correspondent $m_\theta(i)$ inherits the transported neighbors, $\hat{N}_{\text{content}}(m_\theta(i)) = m_\theta(N_0(i))$. Agreement with the observed graph is measured per node by Jaccard overlap of predicted and observed neighbor sets,

$$J(S,T) = \frac{|S \cap T|}{|S \cup T|}, \qquad \Delta_{\text{anchor}} = J_{\text{content}} - J_{\text{grid}}, \tag{7}$$

and J_{grid} and J_{content} are the means of these per-node overlaps over visible nodes, over all grapher layers of a stage, and over identities. Both predicted and observed sets are restricted to a common visible domain; a node whose union is empty under either hypothesis is excluded from both. Negative Δ_{anchor} indicates agreement with fixed grid positions. Construction rules, the many-to-one quantization, and the separately constructed identity contrast $B - D$ (Table 16) are in Appendix D; J_{grid} and J_{content} are reported as means without intervals.

At $14^2/75^\circ$, grid/content overlaps are $0.527/0.438$ for V and $0.671/0.543$ for H, indicating grid anchoring at this intermediate stage. V instead shows a small content preference at 28^2 and 7^2 (Figure 2a). A complementary analysis compares graph overlap for same-identity and different-identity pairs. At 14^2, this advantage decreases from 0.0839 at 15° to 0.0050 at 75° (Figure 2b). Thus this overlap statistic becomes less discriminative with yaw; it does not quantify all identity information encoded by the graph.

The pattern depends on the training dataset. For V trained on CASIA-WebFace (Yi et al., 2014), the yaw-averaged 14^2 gap is $+0.002 \pm 0.003$ across three training seeds, with a reported 95% t-interval of $[-0.004, +0.009]$. This differs from the negative MS1MV2 value of -0.061. At 75°, however, the CASIA gap remains negative (-0.035 ± 0.002). Grid anchoring at extreme yaw therefore replicates, while the full yaw-averaged pattern does not.

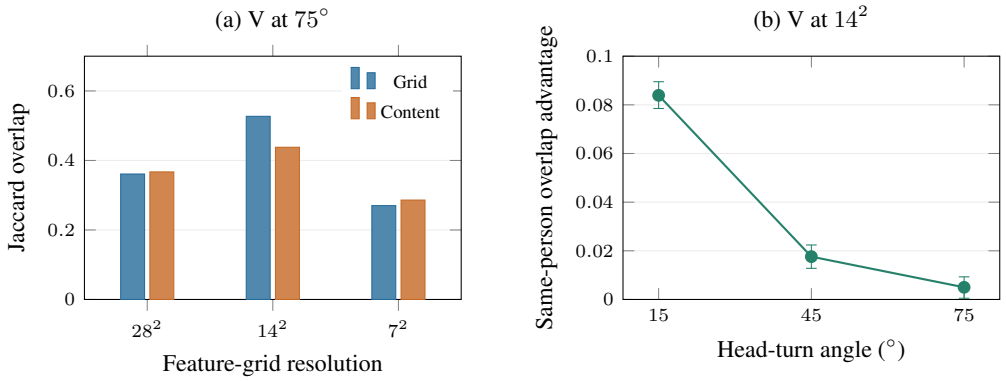

Figure 2: **Layer-specific grid anchoring.** (a) Grid and content predictions at 75° should be compared within each resolution; absolute overlaps are not directly comparable across resolutions. (b) At 14^2, the same-identity overlap advantage decreases with yaw but remains positive. Error bars are reported 95% bootstrap intervals, shown only at the three recorded angles.

3.2 Spatial sensitivity and correspondence in local features

An affine-warp experiment separately probes sensitivity to spatial layout. After a two-dimensional image warp, realigning the extracted feature map before reconstructing the graph increases graph overlap and embedding similarity from $0.639/0.729$ to $0.863/0.984$. This recovery supports a contribution from spatial layout. Interpolation, padding, and nonlinear feature extraction nevertheless preclude exact recovery, and a 2D warp does not reproduce viewpoint-induced self-occlusion. Appearance changes remain a possible additional source of error.

We next evaluate features at supplied corresponding locations, without requiring the network to predict those locations. For visible points $i \in V$, the matching-location advantage compares feature similarity at the mapped position $m_\theta(i)$ with similarity at the unchanged grid position i:

$$M_{\text{feat}} = \mathbb{E}_{i \in V}[\cos(f_0[i], f_\theta[m_\theta(i)]) - \cos(f_0[i], f_\theta[i])]\,, \tag{8}$$

$$D_{\text{corr}} = M_{\text{true}} - M_{\text{shuffled}}, \tag{9}$$

Here f denotes the feature map and cos is cosine similarity. The identity-substituted control substitutes another person's image at the same pose while retaining the maps. Subtracting this control yields D_{corr}, the correspondence advantage beyond that control. Without correspondence supervision (five-epoch checkpoint), 14^2 features give $M_{\text{true}} = 0.1075$, $M_{\text{shuffled}} = -0.0125$, and $D_{\text{corr}} = 0.1200$ (bootstrap percentile interval $[0.1169, 0.1231]$). A positive D_{corr} therefore shows identity-conditioned feature preservation at supplied locations; it does not show that the network localizes those correspondences.

4 Geometric limits of cell-constrained sampling

Connectivity and feature analyses do not directly determine whether a sampling operation can reach a displaced surface point. We therefore consider a *query*, the image-plane coordinate at which a feature is sampled, constrained to its original square cell. For a cell of side length h centered at r_i, the admissible region is

$$C_i = r_i + [-h/2, h/2]^2, \qquad p_i(\theta) \in C_i. \tag{10}$$

Equation (10) explicitly constrains each query to its original cell. Figure 3 distinguishes the projected surface point from the admissible sampling location.

Theorem 1 (Exact fixed-cell tracking error)**.** *Let $q_i(\theta) \in \mathbb{R}^2$ denote the projected location of a visible surface point initially at r_i. Under Equation (10), the maximum query displacement and*

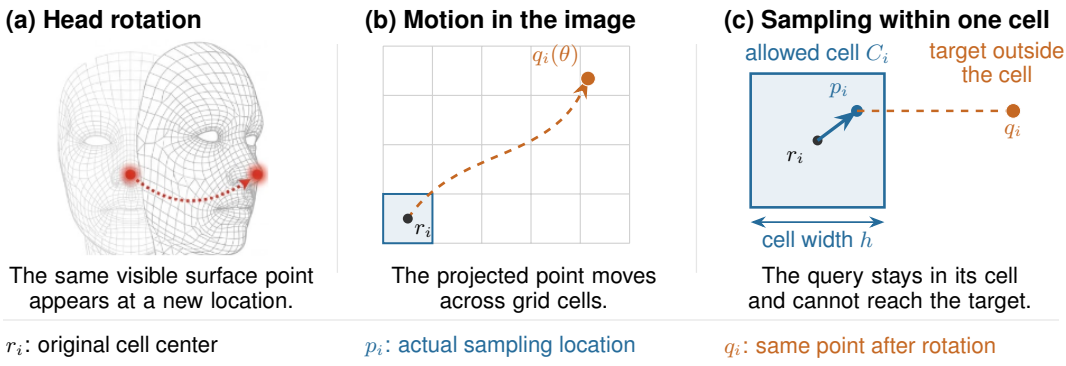

Figure 3: **Projected surface motion and cell-constrained sampling.** (a) Head rotation changes the image location of a visible surface point. (b) The projected point can move continuously across grid cells. (c) A query $p_i \in C_i$ cannot reach a target q_i outside its original cell. The dashed segment denotes the error of the illustrated query; the minimum achievable error is the distance from q_i to C_i. This constraint concerns sampling coordinates, not the full information represented by the sampled features.

minimum tracking error satisfy

$$\|p_i - r_i\|_2 \le h/\sqrt{2}, \tag{11}$$

$$e_i^*(\theta) := \min_{p \in C_i} \|p - q_i(\theta)\|_2 = \left\| \left(|q_i(\theta) - r_i| - \tfrac{h}{2}\mathbf{1} \right)_+ \right\|_2 . \tag{12}$$

Absolute value and $(\cdot)_+$ act coordinate-wise, with $(a)_+ = \max(a, 0)$. Exact tracking is feasible if and only if $q_i(\theta) \in C_i$; tracking within tolerance ε is feasible if and only if $e_i^*(\theta) \le \varepsilon$.

Proof. Each coordinate displacement is bounded by $h/2$, giving the Euclidean bound $h/\sqrt{2}$. The nearest admissible point is the coordinate-wise projection of the target onto the square. Its residual coordinates are $(|q_{i,a} - r_{i,a}| - h/2)_+$, whose Euclidean norm gives Equation (12). □

Continuous and quantized target motion. A target $q_i = r_i + (0.51h, 0)$ lies outside the cell after a displacement of $0.51h$, with minimum error $0.01h$. If the target is quantized to a distinct grid center, the stronger bound $e_i^* \ge h/2$ applies. A minimum displacement of one cell is therefore a property of distinct grid centers, not continuous surface motion.

Displacement statistics and tracking coverage. At 14^2, mean distance to the quantized target increases from 0.44 to 2.65 cells between 15° and 75°, while mean query displacement decreases from 0.189 to 0.099 cells. At 75°, these distances are 21.2 and 0.79 pixels, respectively. Their ratio, $0.099/2.65 = 3.7\%$, compares mean distances rather than the fraction of tracked points. Substituting the maximum in-cell reach, $h/\sqrt{2} \approx 0.707$ cells, for the mean reach gives $0.707/2.65 \approx 26.7\%$, which likewise does not bound tracking coverage. Coverage requires individual target locations, displacement directions, and an error tolerance (Appendix E).

Scope of the geometric result. Sampled features may encode a broad receptive field, and graph aggregation can transfer information across cells. In particular, spatially repeated identity embeddings can support discrimination without pointwise tracking. The theorem therefore constrains the sampling rule rather than recognition capacity. Table 3 distinguishes interventions on sampling, connectivity, and representation learning.

Table 3: **Scope of the analyzed interventions.** Theorem 1 applies only to queries constrained to their original cells. Recognition effects require separate empirical evaluation.

Intervention	Modified component	Evidence and scope
Cell-local landmark / graph-branch resampling	Sampling restricted to C_i	The reach bound applies. A missing training input invalidated the original landmark experiment.
Positional warp / landmark distance bias	Positional signals / neighbor selection	Empirical outcomes are not implied by the sampling theorem.
Unrestricted offsets	Sampling beyond the original cell	One seed, five epochs: 83.757% CFP-FP versus 84.957% for the control. Other training settings remain untested.
Edge rewiring	Connections	A gain on the restricted graph reversed on the intact global graph.
Pose augmentation / alignment loss	Learned features	Requires matched added images and training budgets.

Table 4: **Matched-budget intervention results.** All six conditions use one training seed and the same number of steps; the final three also share the rendered images. Augmentation alone achieves the highest CFP-FP accuracy. Dashes indicate unreported D_{corr}. GAR uses model-specific EER thresholds; training-seed variability is unavailable. The map-permuted arm keeps the rendered images and permutes their correspondence maps across identities; this is a different operation from the identity-substituted control of Section 3.2, which replaces the image and keeps the maps.

Condition	CFP-FP accuracy (%)	$\text{GAR}_{>70^\circ}$ (%)	D_{corr}
No renders, V-equivalent	84.957	62.716	—
Landmark distance bias	83.914	62.469	—
14^2 windowed graph	82.929	61.481	—
Pose renders only	88.000	75.062	0.156
Pose renders + map-permuted alignment	86.557	75.309	0.029
Pose renders + true alignment	87.014	72.716	0.034

5 Pose augmentation and correspondence supervision

5.1 Alignment under matched data and training budgets

Feature alignment encourages similar representations at corresponding locations. Isolating its contribution requires controlling for the additional training images. We compare three conditions using the same 39,249 rendered pairs: pose augmentation alone, augmentation with map-permuted alignment, and augmentation with true alignment. The true-alignment objective adds a cosine-similarity term to the recognition loss:

$$\mathcal{L} = \mathcal{L}_{\text{ArcFace}} + \lambda \mathbb{E}_{i \in V} \left[1 - \cos\left(f_0[i], f_\theta[m_\theta(i)] \right) \right]. \tag{13}$$

The alignment term is applied at the 14^2 and 7^2 stages (Appendix F). The main setting is $\lambda = 1$, selected after short calibration runs. A benefit attributable to correspondence supervision requires improvement over both the augmentation-only and map-permuted controls.

True alignment reduces accuracy by 0.986 pp relative to augmentation alone and improves it by 0.457 pp relative to map-permuted alignment (Table 4). Both are evaluated at the same pinned batch, so batch-size sensitivity (Section 6) is not a confound for this paired difference, which nonetheless remains a small single-seed effect. This five-epoch negative result applies to the tested attractive-only cosine alignment; contrastive objectives, other loss weights, and regularization strategies were not examined.

Feature similarity versus spatial specificity. The controlled feature advantage D_{corr} is 0.156 for augmentation alone, compared with 0.029 and 0.034 for map-permuted and true alignment. A small alignment loss is compatible with either globally uniform features or spatially uniform features that remain distinct across identities. The reduction in D_{corr} therefore does not establish complete loss of identity information.

Table 5: **Augmentation across training datasets.** MS1MV2 entries are percentages; the CASIA row reports baseline-to-augmentation gains in percentage points, paired by seed. Where available, $\pm$ denotes SD across three training seeds.

Training corpus	Model / comparison	CFP-FP	$\text{GAR}_{>70°}$
MS1MV2	R50, fixed reference	97.67	93.58
MS1MV2	V, fixed reference	95.77	86.05
MS1MV2	V + pose renders	96.82 ± 0.07	92.02 ± 0.99
CASIA-WebFace	Paired augmentation gain	$+1.41 \pm 0.50$	$+6.21 \pm 1.03$

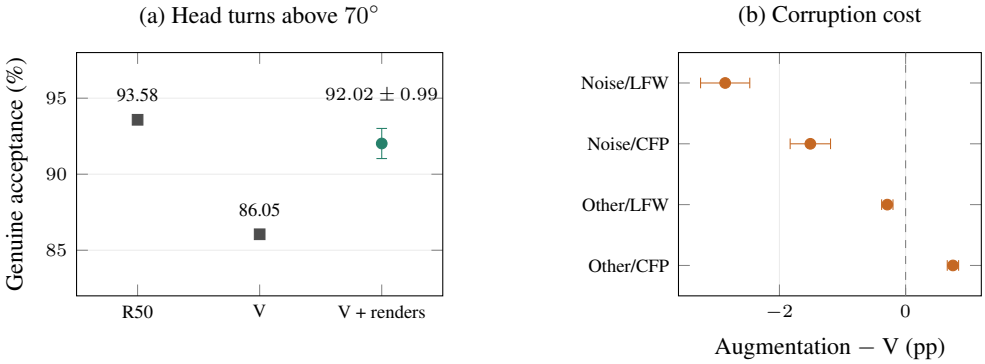

Figure 4: **Pose recognition and corruption robustness after augmentation.** (a) Extreme-pose GAR, with SD across training seeds; both references are fixed models. (b) Accuracy changes under severity-5 corruptions, with reported standard errors (SEs) combining training-seed and corruption-draw variation. These SEs differ from the SD in panel (a); Appendix G specifies the available uncertainty information.

5.2 Full-schedule evaluation and cross-dataset replication

After twenty epochs, pose augmentation achieves $96.82 \pm 0.07\%$ CFP-FP accuracy and $92.02 \pm 0.99\%$ extreme-pose GAR across seeds 2026/2027/2028 (Table 5). The GAR deficit relative to the fixed R50 reference of 93.58% decreases from 7.53 to 1.56 ± 0.99 pp. The SD excludes uncertainty from retraining the CNN, which did not receive the additional rendered images. This narrowing therefore shows augmentation compensating for the graph backbone's cross-pose deficit; whether the residual gap persists once the CNN is trained on the same renders is an open question (Table 7).

On CASIA-WebFace, augmentation improves recognition for all three seeds, although the yaw-averaged grid-anchoring pattern does not replicate (Section 3). The training dataset changes while test identities remain fixed. Recognition gains thus replicate across training datasets more consistently than the proposed connectivity explanation. Pose augmentation is established prior work (Masi et al., 2016; Zhao et al., 2017); here it serves as a necessary control for gains that might otherwise be attributed to correspondence supervision. These gains occur without changing the fixed image grid.

5.3 Robustness under image corruption

We evaluate eighteen severity-5 corruption types from the OODFace/ImageNet-C framework (Kang et al., 2024; Hendrycks & Dietterich, 2019). Three training seeds are assessed under three draws of each stochastic corruption, with identical corrupted images across models. Relative to V, augmentation reduces mean accuracy across five pixel-noise types by 2.86 pp on LFW and 1.51 pp on CFP-FP. Across the other thirteen types, the changes are -0.29 and $+0.75$ pp (Figure 4). The degradation is concentrated in pixel noise, although CASIA AgeDB-30 accuracy also decreases by 0.68 ± 0.55 pp.

Under pixel noise, the augmented model trails R50 by 1.57 pp on LFW but retains a 1.20 pp advantage on CFP-FP. The mechanism underlying this degradation remains unresolved (Appendix G).

Table 6: **Controls for interpreting apparent gains.** Each row concerns a distinct experiment and unit of measurement; the values do not constitute a cross-method ranking.

Control	Reported comparison	Potential confound
Identity-shuffled input	True landmarks: $+0.0489$; another person's landmarks: $+0.0486$ on the same score	Identity-independent response attributed to correspondence
Intact baseline	Rewiring changes CFP-FP by $+0.129$ pp after windowing, but -0.128 pp on the intact graph	Recovery of removed edges attributed to new capability
Same added images	True alignment: 87.014%; images only: 88.000% CFP-FP	Additional-data effects attributed to the loss
Fixed evaluation batch	Batch 64 versus 128: up to 0.467 pp accuracy change for tested top-k models	Numerical variation attributed to the method

6 Discussion and Limitations

Related architectures. ViG, hypergraph models, and axial graphs adapt information exchange. Specific windowed-graph and adaptive-gating architectures such as WiGNet and AdaptViG (Spadaro et al., 2025; Munir et al., 2026) address related questions but are not evaluated as external baselines here; windowed partitioning of the 14^2 graph is used internally as a diagnostic control (Table 4, Appendix F). Deformable sampling (Dai et al., 2017) and part-based face transformers (Sun & Tzimiropoulos, 2022) instead modify sampling locations.

Attribution controls. Table 6 distinguishes identity-specific effects, recovery of removed connections, additional-data effects, and numerical variation (Appendix H).

Numerical sensitivity of graph selection. Changing evaluation batch size from 64 to 128 alters accuracy on seven of seven benchmarks for V and on six of seven for H and the posedeform variant, by up to 0.467 pp. R50, ViT-S, and threshold-based GreedyViG show unchanged accuracy, without implying identical tensors. Floating-point results can depend on batching (PyTorch Contributors, 2026), and top-k selection may amplify perturbations near its boundary. A gap γ between the last selected and first excluded distances (Lemma 2) guarantees stability under distance perturbations bounded by ϵ when $\gamma > 2\epsilon$ (Appendix E). This is a possible mechanism, not a verified cause; threshold selection can also be boundary-sensitive.

Limitations. The study covers aligned 112×112 faces, one main graph-model scale, and $k = 9$. Most architecture comparisons use one seed; alignment and unrestricted-offset pilots use five epochs. Needed controls include a CNN with the same rendered images, independently trained GDC and average-pooling variants, sampling at known correspondences without cell constraints, and within-identity spatial shuffling. Connectivity patterns depend on the training dataset; one-to-many identification and demographic fairness remain unevaluated. The augmented arms process 32 additional rendered pairs per batch; a control with equal image budget and no added pose diversity was not run.

Conclusion. Adaptive connectivity, cross-view correspondence, and identity discrimination require distinct evidence. Grid-anchored connectivity coexists with a feature advantage at supplied correspondences. Pose augmentation improves recognition, whereas the tested alignment loss reduces both that advantage and recognition relative to augmentation alone under matched data and training budgets. Recognition gains replicate more consistently than the connectivity pattern, and a correspondence-based explanation must account for these differences across training conditions and datasets. The geometric result characterizes cell-constrained sampling without imposing a general recognition limit. Claims of content tracking should therefore assess connectivity, feature correspondence, and recognition separately under controlled evaluation.

AI USE STATEMENT

This work was produced with substantial use of generative AI assistants, and we describe that use in full. The research question, the choice of testbed, and the conceptual framing of this manuscript, including the distinction between adaptive connectivity, cross-view correspondence, and recognition, originated with the authors.

Of the uses ICLR requires us to disclose, the following apply. AI assistants formulated and refined the hypotheses tested by individual experiments, designed the experimental protocols and control conditions, implemented essentially all of the code, executed the experiments, and contributed to interpreting the results. An assistant proposed the initial form of the geometric bound on cell-constrained sampling reach, which the authors restated and extended into Theorem 1 and Corollary 1 as they appear here; an assistant also proposed a generalization of that bound, which was refuted under a pre-registered check and is not reported. Assistants also translated drafts and reformatted derived data for analysis. We did not use generative AI to generate synthetic datasets: the rendered multi-yaw face images come from a 3D morphable model fitting and rendering pipeline, not from a generative model. Formulating survey or interview questions and transcribing research recordings are not applicable.

Of the uses ICLR recommends disclosing, the following also apply. The authors wrote the text of Sections 1 to 6 and produced the figures; AI assistants drafted earlier versions of the manuscript that informed its content, and after external review contributed corrections and clarifications amounting to roughly one sixth of the main text and portions of the appendix. Assistants also searched for and formatted references, summarized related literature, suggested experimental parameters, and assisted with general brainstorming.

We have reviewed the claims, the numbers that support them, and the code that produces those numbers. Every quantitative claim in this paper traces to a stored analysis record, and the accompanying scripts regenerate the summary records from the stored ones; the manuscript carries an automated guard that fails if a value appears without a matching record. Training code, checkpoints, and raw pair scores are not included (Appendix H.2). Every confirmatory experiment was pre-registered: its prediction and decision rule were committed to version control, or for three early experiments recorded in the script header (Appendix H.1), before measurement, so that an AI-proposed hypothesis could not be revised after seeing the outcome. This discipline repeatedly overturned AI-proposed claims. We maintain a ledger of every number retired or corrected during the project, which accompanies the code. Analyses that could not be verified were dropped rather than reported.

We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Face recognition involves sensitive biometric data and may enable surveillance or exhibit unequal error rates across demographic groups. Aggregate benchmark performance establishes neither fairness nor suitability for deployment. Any release must comply with the original dataset conditions. The accompanying package includes no training data or checkpoints. Figure 5 (Appendix C) reproduces an example from an earlier version of this study, and the geometric illustration (Figure 3) is schematic.

REPRODUCIBILITY STATEMENT

Sections 2–5 specify the model, diagnostic analyses, and principal comparisons. The appendices provide proofs, complete reported results, uncertainty definitions, and outstanding reproducibility requirements. The accompanying source rebuilds the manuscript and summary figures. Reproducing the empirical results additionally requires the configurations, checkpoints, pair scores, and rendering records listed in Appendix H.2.

REFERENCES

Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH*, pp. 187–194, 1999.

Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 67–74, 2018.

Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. In *Chinese Conference on Biometric Recognition*, 2018.

Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 764–773, 2017.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z. Li. Towards fast, accurate and stable 3D dense face alignment. In *European Conference on Computer Vision*, pp. 152–168, 2020.

Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, 2016.

Kai Han, Yunhe Wang, Jianyuan Guo, Yehui Tang, and Enhua Wu. Vision GNN: An image is worth graph of nodes. In *Advances in Neural Information Processing Systems*, volume 35, pp. 8291–8303, 2022.

Yan Han, Peihao Wang, Souvik Kundu, Ying Ding, and Zhangyang Wang. Vision HGNN: An image is more than a graph of nodes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019.

Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life' Images*, 2008.

Caixin Kang, Yubo Chen, Shouwei Ruan, Shiji Zhao, Ruochen Zhang, Jiayi Wang, Shan Fu, and Xingxing Wei. OODFace: Benchmarking robustness of face recognition under common corruptions and appearance variations. *arXiv preprint arXiv:2412.02479*, 2024.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International Conference on Machine Learning*, volume 97, pp. 3519–3529, 2019.

Iacopo Masi, Anh Tuan Tran, Tal Hassner, Jatuporn Toy Leksut, and Gérard Medioni. Do we really need to collect millions of faces for effective face recognition? In *European Conference on Computer Vision*, pp. 579–596, 2016.

Brianna Maze, Jocelyn Adams, James A. Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K. Jain, W. Tyler Niggel, Janet Anderson, Jordan Cheney, and Patrick Grother. IARPA Janus Benchmark-C: Face dataset and protocol. In *International Conference on Biometrics*, pp. 158–165, 2018.

Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: The first manually collected, in-the-wild age database. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

Mustafa Munir, William Avery, Md Mostafijur Rahman, and Radu Marculescu. GreedyViG: Dynamic axial graph construction for efficient vision GNNs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6118–6127, 2024.

Mustafa Munir, Md Mostafijur Rahman, and Radu Marculescu. AdaptViG: Adaptive vision GNN with exponential decay gating. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2026.

PyTorch Contributors. Numerical accuracy. PyTorch documentation, 2026. URL https://docs.pytorch.org/docs/main/notes/numerical_accuracy.html. Accessed September 16, 2026.

Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M. Patel, Rama Chellappa, and David W. Jacobs. Frontal to profile face verification in the wild. In *IEEE Winter Conference on Applications of Computer Vision*, 2016.

Gabriele Spadaro, Marco Grangetto, Attilio Fiandrotti, Enzo Tartaglione, and Jhony H. Giraldo. WiGNet: Windowed vision graph neural network. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 859–868, 2025.

Zhonglin Sun and Georgios Tzimiropoulos. Part-based face recognition with vision transformers. In *British Machine Vision Conference*, 2022.

Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K. Jain, James A. Duncan, Kristen Allen, Jordan Cheney, and Patrick Grother. IARPA Janus Benchmark-B face dataset. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 90–98, 2017.

Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

Jian Zhao, Lin Xiong, Jayashree Karlekar, Jianshu Li, Fang Zhao, Zhecan Wang, Sugiri Pranata, Shengmei Shen, Shuicheng Yan, and Jiashi Feng. Dual-agent GANs for photorealistic and identity preserving profile face synthesis. In *Advances in Neural Information Processing Systems*, pp. 66–76, 2017.

Tianyue Zheng and Weihong Deng. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. Technical report, Beijing University of Posts and Telecommunications, 2018.

Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

A Model specification and comparison protocols

Implementation details. GraphFace combines the PViG-S stem and pyramid with PReLU, a GDC readout, and a 512-dimensional identity embedding. Equation (4) is a simplified single-head formulation. Complete specification additionally requires the original configuration for batch normalization, biases, head grouping, dilation, self-edges, and squared versus unsquared feature distance. Executable training code and checkpoint hashes are absent from the supplied manuscripts; these implementation choices cannot be inferred reliably from the equation.

ArcFace supervises identity discrimination. For normalized embeddings and class weights, the reported objective is

$$\mathcal{L}_{\text{ArcFace}} = -\frac{1}{B}\sum_{b=1}^{B}\log\frac{e^{s\cos(\phi_{b,y_b}+m)}}{e^{s\cos(\phi_{b,y_b}+m)}+\sum_{j\neq y_b}e^{s\cos\phi_{b,j}}}, \qquad m=0.5,\ s=64. \tag{14}$$

Here $\phi_{b,j}$ is the angle between the embedding of image b and the weight vector of class j; y_b is the ground-truth class, and B is the batch size. The objective constrains identity discrimination without uniquely specifying the spatial organization of intermediate features. Additional alignment supervision may therefore alter local correspondence information while preserving some identity discrimination.

Table 7: Controlled factors and residual differences across comparisons. A shared seed index does not make different architectures equivalent random experiments.

Comparison	Controlled factors	Residual differences or limitations
Main models	Training data, loss, optimizer, schedule, seed number, evaluation code	Capacity, computation, speed, and architecture-specific tuning
Five-epoch loss tests	Added images, steps, seed number, evaluation code	Short schedule; training-seed variability unmeasured
Twenty-epoch augmentation	Graph-model family, schedule; three training seeds	Fixed CNN without pose augmentation
CASIA replication	Paired baseline and augmentation seeds, evaluation procedure	Different training corpus; shared test identities
Corruption tests	Identical corrupted images for compared models	Distinct training-seed and corruption-draw variation
Batch-size test	Model checkpoint, image pairs, precision, evaluation code	Responsible numerical operation not isolated

B Recognition benchmarks and computational performance

MS1MV2 contains approximately 5.8M images of 85K identities. Table 8 uses evaluation batch size 128. The reported mean spans frontal, pose-related, and age-related benchmarks rather than frontal verification alone. Earlier component ablations use batch size 64 and are reported separately.

Table 8: Reported face-verification accuracy (%) at evaluation batch size 128. Static resampling denotes a separate model configuration; its recognition score alone does not establish surface-point tracking.

Model	LFW	CFP-FF	CFP-FP	AgeDB	CALFW	CPLFW	VGG2-FP	Mean
R50	99.78	99.90	97.67	98.00	96.13	92.30	95.38	97.02
ViT-S	99.45	99.59	90.11	95.22	95.22	88.62	91.66	94.27
GreedyViG-S	99.55	99.69	93.76	96.52	95.57	89.47	93.36	95.41
ViHGNN	99.80	99.71	95.64	97.55	96.00	90.52	94.06	96.18
V	99.72	99.79	95.77	97.55	95.87	90.78	94.78	96.32
H	99.75	99.81	95.74	97.22	95.82	91.00	94.28	96.23
Static resampling	99.77	99.83	95.93	97.55	95.93	91.03	94.54	96.37

Training-seed variability for V (three seeds) is 95.59 ± 0.24 on CFP-FP and 3.578 ± 0.029 in d'; a second H seed gives d' 3.621. The d' distance to the CNN (0.47) is an order of magnitude larger than this dispersion.

LFW and CFP-FF primarily evaluate near-frontal verification. CFP-FP, CPLFW, and VGGFace2-FP emphasize pose variation, whereas AgeDB-30 and CALFW emphasize age variation (Huang et al., 2008; Sengupta et al., 2016; Zheng & Deng, 2018; Cao et al., 2018; Moschoglou et al., 2017; Zheng et al., 2017). All are established benchmarks.

Table 9: Genuine acceptance rates (%) on CFP-FP at model-specific full-set EER thresholds. The reported yaw groups contain 3,483 genuine pairs; of 3,500 genuine pairs, 9 lack a yaw estimate and 8 fall below 30° and are not shown.

Estimated yaw	Number of genuine pairs	R50	H	V
30°–50°	760	99.47	99.21	99.08
50°–70°	1,913	98.64	98.01	97.54
$> 70°$	810	93.58	85.43	86.05

The test-pair yaw estimator differs from the rendering model. Reproduction requires its version, angle convention, pair-exclusion criteria, and threshold procedure. Full-set EER results provide descriptive comparisons; deployment-oriented evaluation requires independent calibration and a common false-acceptance rate.

Template-based verification. IJB-B/C aggregate images and video frames into identity templates (Whitelam et al., 2017; Maze et al., 2018). Evaluation uses shared aligned images and media-weighted features. These benchmarks extend pairwise verification but neither isolate pose variation nor evaluate one-to-many identification.

Table 10: Reported true-acceptance rate (TAR, %) at fixed false-acceptance rates (FAR). The IJB-B result at FAR 10^{-6} is omitted because that operating point is poorly resolved.

Model	IJB-C 10^{-4}	IJB-C 10^{-6}	IJB-B 10^{-4}
R50	95.97	87.23	94.58
R50 + GDC	95.88	87.66	94.40
ViHGNN	95.05	84.90	93.55
Static resampling	94.94	86.16	93.14
V	94.80	86.59	92.96
H	94.74	85.16	92.78
GreedyViG-S	92.53	78.73	90.09
ViT-S	91.79	81.30	89.67

Historical component ablations. The raw PViG-S → hybrid → GDC → PReLU sequence yields five-benchmark means of 94.67, 94.18, 94.52, and 95.90% at evaluation batch size 64; V yields 96.02%. Corresponding noise means are 70.93, 69.75, 69.96, and 76.57%. Adding PReLU increases the noise mean by 6.61 pp within this sequence, compared with a total raw-to-final increase of 5.64 pp. The sequence does not constitute a factorial ablation of all components. Adding the same GDC readout and PReLU to ViT-S (two seeds) changes the five-benchmark mean by $+0.01$ pp and CFP-FP by -0.66 pp; the components do not close its cross-pose gap (Table 11).

Table 11: GDC readout and PReLU added to ViT-S. Means over two training seeds (2026 and 2027), evaluated in same-seed pairs at batch size 64 as in the historical ablations. The five-benchmark mean spans LFW, CFP-FP, AgeDB-30, CALFW, and CPLFW; the severe-noise mean is over the five severity-5 pixel-noise types on LFW.

Model	Five-benchmark mean (%)	CFP-FP (%)	Severe-noise mean (%)
ViT-S	93.69	90.23	82.12
ViT-S + GDC + PReLU	93.70	89.57	81.65
Δ	$+0.01$	-0.66	-0.47

Arithmetic cost and measured latency. Timing measurements use an RTX 3090 Ti, 32-bit floating point, 50 warm-up iterations, 100 timed iterations, and three runs. MAC counts omit some neighbor-selection and feature-gathering costs. R50 requires $6.334/1.493 \approx 4.2$ times V's MACs and approximately 1.4 times its parameters. Counting two operations per MAC preserves this ratio.

Table 12: Hardware-specific inference measurements. Lower arithmetic cost does not imply lower single-image latency. Throughput is measured in images per second.

Model	GMACs	Batch-one latency (ms)	Batch-32 throughput (images/s)
R50	6.334	1.89	1,580
ViT-S	1.088	2.53	7,888
V	1.493	4.38	1,627
H	1.493	6.03	1,229

C HISTORICAL QUALITATIVE AND GRAPH ANALYSES

Table 13: Historical GraphFace-H Jaccard overlaps from 50 real-image pairs per group. These results are distinct from the later 500-pair analysis and rendered-image experiments.

Group	28^2	14^2	7^2	4^2
Same identity, frontal–frontal (A)	0.646	0.749	0.350	0.921
Same identity, frontal–profile (B)	0.563	0.663	0.297	0.850
Different identity, frontal–frontal (C)	0.582	0.692	0.320	0.901
Different identity, frontal–profile (D)	0.559	0.659	0.295	0.848

For 500 pairs per group at 7^2, the same-identity overlap advantage is 0.0183 for matched poses and 0.0028 across poses, with reported interval $[0.0017, 0.0039]$ for the latter. The cross-pose advantage is approximately 6.5 times smaller but remains positive. This reduction in mean Jaccard overlap does not imply complete loss of identity information from the graph.

The earlier pose-sensitivity statistic is $1-r$, where r compares mean frontal and profile maps. Across $28^2, 14^2, 7^2, 4^2$, H yields $0.30, 0.19, 0.07, 0.03$ and V yields $0.37, 0.33, 1.26, 0.34$; corresponding R50 values are $0.92, 1.09, 0.56, 0.18$. V has its largest response at 7^2, whereas H and R50 peak at other stages. Map-change statistics, graph overlap, and Grad-CAM attribution quantify distinct properties.

D CONNECTIVITY AND FEATURE-ANALYSIS PROTOCOLS

All compared neighbor sets require a common visible domain, with valid correspondences for both endpoints before transport. Grid quantization can merge points, and occlusion can remove them; the transported graph is therefore not necessarily a permutation of the frontal graph. Locations at which grid and content predictions coincide cannot distinguish the hypotheses. The 4.9% in Table 17 is the demand-zero fraction from the re-runnable displacement records; the 4.9% unchanged-map fraction at $14^2/75°$ quoted for the probe counts the same nodes by construction (a node whose correspondent stays in its cell has zero demand) but is recorded separately in the probe's verdict note, without rerunnable evidence of its own, and is excluded from the principal findings.

Construction of J_{grid} and J_{content}. Graphs are extracted on the full 196-node render; the visible restriction is applied afterwards to the neighbor sets. Neighbor sets are the top-k ($k = 9$) by the model's selection distance and include the node itself, since the self-distance is minimal. For each grapher layer of a stage and each visible source node i, the grid prediction $S_g = N_0(i) \cap \text{Vis}$ is compared with the observed $T_g = N_\theta(i) \cap \text{Vis}$, where visibility is applied by node index; the content prediction $S_c = m_\theta(N_0(i) \cap \text{Vis})$ is compared with the observed $T_c = N_\theta(m_\theta(i)) \cap m_\theta(\text{Vis})$. Sets are unordered and directed from the source node; duplicates arising from many-to-one quantization

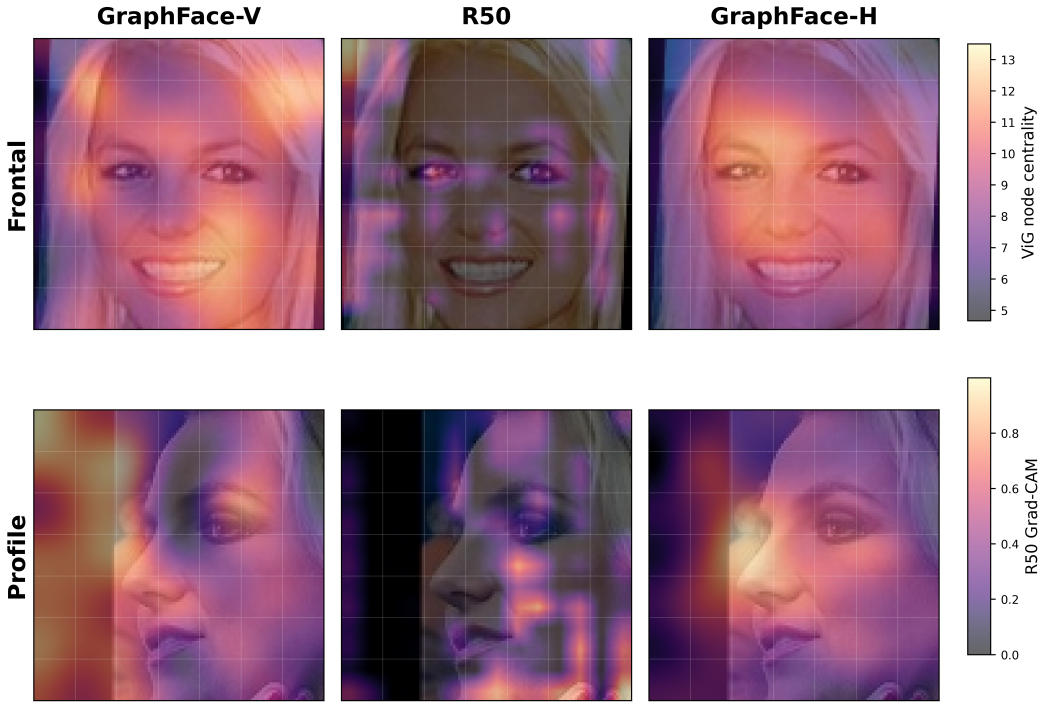

Figure 5: **Original frontal–profile example from the earlier GraphFace manuscript.** The images are unchanged. GraphFace maps show node centrality, a connectivity statistic, whereas R50 maps show Grad-CAM attribution. Their values therefore represent different quantities. The graph maps use 7^2 nodes, while the main controlled analysis focuses on 14^2. This example illustrates pose-dependent responses without identifying a common failure mechanism or a shared most-sensitive layer.

are merged, and a transported neighbor that coincides with $m_\theta(i)$ counts as a hit. Per-node Jaccard values are summed over nodes and layers (two layers at 28^2 and 14^2, six at 7^2), divided by the count of nodes with non-empty unions under both hypotheses, and averaged over identities. No interval is computed for these means. Filtering occluded neighbors only from the content prediction would bias this comparison against content transport.

Construction of $B - D$. Same-identity overlap B compares $N_0(i)$ with $N_\theta(i)$ at the same node index on a fixed 49-node subsample, without transport or visibility restriction; D does the same across identities paired in a fixed cyclic order. Intervals are BCa over 10,000 resamples of identities, each draw applied to both the B and the D value of the resampled identities so that the two values of one identity stay paired (equivalently, a bootstrap of the per-identity differences); an earlier version resampled the two arrays independently, which gave wider intervals in eleven of the twelve cells of Table 16 (and a 3% narrower one for V at $7^2/15°$) with the same signs. Because D pairs each identity with the next in the cyclic order, D_a also involves identity $a + 1$; identity-level resampling pairs B_a with D_a but does not model that second link. This construction differs from that of J_{grid} and J_{content} (Figure 2a), and the two are not directly comparable.

Table 14: Graph overlap at 75° on a common visible domain. Grid and content predictions should be compared within each resolution because graph density and coordinate quantization vary across resolutions.

Resolution	V grid	V content	H grid	H content
28^2	0.361	0.367	0.564	0.516
14^2	0.527	0.438	0.671	0.543
7^2	0.270	0.286	0.329	0.336

The static-resampling configuration of Table 8 was run through the same probe at three training seeds (Table 15). Both overlaps at $14^2/75°$ rise relative to V, the grid overlap by more than the content overlap, so the gap becomes more negative; at 7^2 its identity contrast is below that of V at 75° (0.0014 versus 0.0021) and above it at lower yaws, most clearly at 15°. Its recognition gain is therefore not accompanied by a reduced grid preference in this probe.

Table 15: Static resampling under the rendered-image probe: three training seeds (mean ± SD) against V (one seed). Overlaps at $14^2/75°$ on the common visible domain; $B - D$ at 7^2.

Quantity	V	Static resampling	Δ
J_{grid}, $14^2/75°$	0.527	0.616 ± 0.010	$+0.089$
J_{content}, $14^2/75°$	0.438	0.505 ± 0.010	$+0.067$
$J_{\text{content}} - J_{\text{grid}}$, $14^2/75°$	-0.090	-0.111 ± 0.004	-0.021
$B - D$, $7^2/75°$	0.0021	0.0014 ± 0.0002	-0.0007
$B - D$, $7^2/15°$	0.0320	0.0433 ± 0.0010	$+0.0113$

Table 16: Same-identity minus different-identity graph overlap on rendered images, with reported bias-corrected and accelerated (BCa) 95% bootstrap intervals. The main plot shows V/14^2. Grid coordinates (same node index), 49-node subsample, cyclic different-identity pairing; identities are resampled jointly, so that the B and D values of one identity stay paired, preserving the dependence among observations from the same identity.

Model	Resolution	15°	45°	75°
V	28^2	.0883 [.0834,.0935]	.0198 [.0159,.0238]	.0089 [.0058,.0121]
V	14^2	.0839 [.0786,.0894]	.0176 [.0135,.0218]	.0050 [.0017,.0083]
V	7^2	.0320 [.0300,.0341]	.0059 [.0047,.0072]	.0021 [.0008,.0032]
H	14^2	.0665 [.0622,.0710]	.0156 [.0125,.0186]	.0064 [.0038,.0091]

Scope of the identity-substituted control. The control substitutes another identity at the same pose while retaining the supplied maps, testing whether the matching-location advantage depends

on identity. Accordingly, *correspondence specificity* denotes only the advantage over this control; it is not coordinate-prediction or point-matching accuracy. A complementary control would preserve identity while shuffling spatial locations, matching displacement, visibility, and facial-region distributions where possible. Results for that control are unavailable. Feature information at supplied correspondences does not establish that the graph itself recovers those locations.

Long-range edge correspondence. At 14^2, a long-range edge spans more than three cells along at least one grid axis (Chebyshev distance). Such edges constitute 14.1% of edges on real CFP-FP images. A separate rendered-image analysis transports endpoints using valid correspondences and records a hit when the destination neighborhood lies within one cell of the mapped target. Hit rates are 0.8100 for true maps (357,966 edges), 0.7891 for identity-shuffled maps (353,353 edges), and 0.2677 for random long-range pairs (357,966 pairs). The true-minus-shuffled difference is 0.0209, substantially smaller than the contrast with random pairs.

Because the eligible edges differ between arms, paired inference requires their intersection. Edges associated with the same identity are also statistically dependent. For reference, short-range true-map agreement is 0.9783 over approximately 1.78M edges. These results demonstrate structured connectivity without establishing identity-specific correspondence or excluding its learnability.

Under one fixed evaluation seed, the hypergraph model has a 14^2 grid/content gap of $+0.003$ at $15°$ and -0.046 at $75°$, providing another instance of the pose-dependent trend. Stochastic clustering and different neighborhood cardinalities preclude direct comparison of absolute overlaps with k-NN graphs or generalization to all graph architectures.

E Formal results and their scope

We first characterize the invariance of a spatially weighted readout under feature relabeling. Let x contain one channel's spatial features, w its GDC weights, and P a permutation matrix acting on positions.

Proposition 1 (Spatial readout under relabeling)**.** *For a fixed permutation P and a single-channel GDC readout $y(x) = w^\top x$, $y(Px) = y(x)$ for every x if and only if $P^\top w = w$. Invariance for all permutations is equivalent to $w = a\mathbf{1}$ for some scalar a.*

Proof. The difference is $(P^\top w - w)^\top x$, which vanishes for every vector x exactly when its coefficient is zero. If invariance holds for every transposition, exchanging any pair of coordinates leaves w unchanged, so all coordinates are equal. Conversely a constant vector is invariant to every permutation. □

For a fixed permutation, weights must be constant within its cycles; invariance to all permutations requires spatially constant weights. The proposition concerns a single GDC channel for arbitrary input features. Physical pose changes additionally alter appearance and visibility, and subsequent layers may suppress readout variation. The result therefore motivates a GDC-versus-average-pooling control rather than identifying the cause of the empirical recognition deficit. Proposition 1 concerns a single pre-projection channel; the multi-channel readout followed by a linear projection and ℓ_2 normalization is a relaxed setting in which channel-wise uniformity is sufficient but not shown to be necessary.

Lemma 1 (Feature-only neighbor construction under relabeling)**.** *Let $A(X)$ be feature-only k-NN adjacency, with a strict gap at each row's selection boundary. Then for a permutation P,*

$$D(PX) = PD(X)P^\top, \qquad A(PX) = PA(X)P^\top. \tag{15}$$

Proof. Permuting feature rows only reindexes the pairwise distances. The strict boundary gap makes each selected set unique; its indices undergo the same permutation. □

With tied distances, equivariance additionally requires a tie-breaking rule that respects relabeling; selection by the lowest node index can violate it. Positional terms must also transform consistently. The lemma concerns relabeling an existing feature set, whereas viewpoint changes alter visibility and appearance as well. The fixed absolute and relative positional terms of Section 2 are indexed by

Table 17: **Mean target and query displacement at 14^2.** Cell width is 8 pixels, and both distances are averaged over visible nodes. Their ratio is a displacement statistic, not tracking coverage. Ratios are computed from the displayed rounded values. Mean demand averages over all visible nodes, including those whose projected target remains in its original cell (demand 0); the lower bound $h/2$ of Corollary 1 applies only to the displaced nodes. At $15°$ the fraction of nodes with demand 0 is 56.2%, so the conditional mean over displaced nodes is approximately $0.44/0.438 \approx 1.0$ cell; at $75°$ the non-displaced fraction is 4.9%. Baseline GraphFace samples at fixed grid centers; these displacement statistics evaluate the auxiliary landmark-anchored resampling of Table 3 applied at evaluation time, not a mechanism of the baseline.

Yaw	Mean demand (cells)	Mean query displacement (cells)	Ratio (%)
$15°$	0.44	0.189	43.0
$30°$	0.85	0.186	21.9
$45°$	1.39	0.169	12.2
$60°$	2.02	0.136	6.7
$75°$	2.65	0.099	3.7

grid coordinate and do not transform under node relabelling, so the operational models do not satisfy the premise of Lemma 1; they are a candidate source consistent with the grid preference observed at 14^2 in Section 3.1; positional ablations would be needed to establish it.

Corollary 1 (Distinct quantized centers)**.** *For square pitch h, let $\bar{q}_i = r_i + hk$, where $k \in \mathbb{Z}^2 \setminus \{0\}$. Under the original-cell constraint,*

$$\min_{p \in C_i} \|p - \bar{q}_i\|_2 = h\sqrt{\sum_{a=1}^{2} (|k_a| - 1/2)_+^2} \geq h/2. \tag{16}$$

Proof. Apply Equation (12). At least one integer coordinate has magnitude at least one, so its residual is at least $h/2$. An adjacent center attains equality. □

The lower bound applies to targets at distinct quantized centers. A continuous target just outside the original cell can be arbitrarily close to its boundary. The theorem uses a closed square; discrete implementations must additionally specify the assignment of points on cell boundaries.

The reported sampler constructs 85 candidates from five landmarks: each landmark contributes its center and eight directions at two radii. A candidate replaces the sampling coordinate of its containing cell, preserving the original-cell constraint. Multiple candidates in one cell require a specified collision rule. Varying the radius changes which cells receive candidates rather than the displacement allowed outside a cell. Table 17 reports mean distances, without the individual displacement vectors required to estimate tracking coverage. An unchanged quantized index is not equivalent to accurate continuous tracking.

General admissible regions and tracking coverage. For any nonempty compact admissible set S_i and $R_i = \max_{p \in S_i} \|p - r_i\|_2$,

$$e_i^*(q) = \operatorname{dist}(q, S_i) \geq (\|q - r_i\|_2 - R_i)_+. \tag{17}$$

The bound follows from the triangle inequality. For a rectangle with side lengths h_x, h_y, the maximum radius is $R_i = \sqrt{h_x^2 + h_y^2}/2$, and coordinate-wise projection onto the allowed intervals gives the exact minimum error. If queries are independently selectable, the maximum fraction of visible points trackable within tolerance ε is

$$\text{Coverage}^*(\varepsilon) = \frac{1}{|V|} \sum_{i \in V} \mathbf{1}\{\operatorname{dist}(q_i, S_i) \leq \varepsilon\}. \tag{18}$$

For a realized sampler, coverage instead uses the achieved error $\|p_i - q_i\|_2$. Neither quantity follows from a ratio of mean distances. Additional joint constraints on queries may reduce simultaneously attainable coverage. These expressions define prospective measurements; the requisite per-point data are unavailable.

Sampling constraints versus recognition capacity. A permutation exchanging distant cells preserves a bijection while allowing displacements much larger than one cell; bijectivity alone therefore does not imply the theorem's constraint. Separately, features $f_y(i) = z_y$ can repeat an identity-specific vector at every location. Distinct z_y permit discrimination without local correspondence. These counterexamples establish the logical scope of the bound, not the solutions obtained by a particular training procedure.

Arbitrarily small spatial perturbations can change neighbor rankings. In one dimension, candidates at $-h$ and $+h$ are equidistant from a query at zero; a positive perturbation makes $+h$ closer. Nearly tied distances admit the same behavior. Grid spacing thus provides no positive lower bound on displacement required to alter selection, and a sampling-reach bound does not rule out effects of positional bias or edge selection.

Alignment loss and spatial specificity. Identical unit features yield zero cosine-alignment loss. Spatially uniform, identity-specific features can also yield zero true-alignment loss while retaining identity discrimination. In both cases, the mapped-location advantage over the unchanged location is zero. These examples motivate separate measurements of D_{corr} and recognition; they do not determine the solution learned by the combined ArcFace and alignment objective.

Lemma 2 (Stable k-nearest-neighbor selection)**.** *Let distances be ordered increasingly and let $k < n$. If $\left\|\widetilde{d} - d\right\|_\infty \le \epsilon$ and $d_{(k+1)} - d_{(k)} > 2\epsilon$, the selected set is unchanged.*

Proof. For any originally selected i and unselected j, $\widetilde{d}_i \le d_{(k)} + \epsilon < d_{(k+1)} - \epsilon \le \widetilde{d}_j$. No unselected candidate can cross the selection boundary. □

Failure of the strict gap condition permits, but does not require, a neighbor change. Threshold selection $d_j < t$ has an analogous boundary: distance perturbations bounded by ϵ and threshold perturbations bounded by η preserve selection if $\min_j |d_j - t| > \epsilon + \eta$. The observed difference between top-k and threshold graphs therefore does not establish numerical immunity of threshold selection.

F Intervention details and experimental scope

The original landmark-sampling experiment did not supply landmarks during training because the input gate blocked them. The trained model was therefore equivalent to V. A subsequent train–test mismatch does not evaluate the intended landmark-conditioned training procedure. Although the proposed sampling rule admits geometric analysis, that analysis cannot validate the failed training configuration.

Warping positional embeddings reduces CFP-FP d' from 3.558 to 3.494, a change of -0.064 with reported interval $[-0.110, -0.018]$. This intervention modifies positional signals. In a separate distance-bias experiment, 97.6% of edge changes were attributed to the spatial prior and 2.4% to landmarks. An internal enabled-versus-disabled comparison did not establish improvement over an independently trained, budget-matched V-equivalent baseline.

The unrestricted-offset pilot permits sampling beyond the original cell. At 14^2 and 7^2, respectively, 71% and 91% of offsets exceed 0.707 cells. CFP-FP accuracy is 83.757%, compared with 84.957% for the V-equivalent control. This configuration lies outside the theorem's assumptions. Its single seed and five-epoch schedule do not exclude benefits under other supervision or training schedules.

The augmentation study retained 39,249 usable pairs from 40,000 rendered candidates. Each pilot uses seed 2026, 56,860 steps (five epochs), training batch size 512, and evaluation batch size 128. The no-render reference uses the same number of steps. Short 500-step calibrations considered $\lambda \in \{0, 0.25, 1, 4\}$; this is not an exhaustive search. The weight was first set to 0.25 from this calibration and revised to 1.0 before launch under a pre-registered correction recorded in the project ledger. The alignment term is applied to the residual output of the final block at the 14^2 (C=160) and 7^2 (C=400) stages; visible nodes from both stages are pooled into a single mean, and no normalization is applied beyond the cosine. Each training batch contains 512 real images and 32 rendered pairs

(64 rendered images), each pair drawn at one yaw from $\{15, \ldots, 75\}^\circ$. The map-permuted alignment arm keeps the same rendered images and replaces the correspondence map of each rendered pair with that of another identity by a per-image permutation. In the five-epoch augmentation-only condition, the grid/content gap changes from -0.0495 to -0.0115 relative to its matched reference. Comparison with the twenty-epoch value of -0.089 would confound augmentation with training duration.

Additional five-epoch tests yield CFP-FP accuracy/GAR of $83.914/62.469$ for landmark distance bias and $82.929/61.481$ for 14^2 windowing, compared with $84.957/62.716$ for the V-equivalent reference. These short-schedule results are not directly comparable to the twenty-epoch benchmarks.

Training-time and evaluation-time locality. Restricting a globally trained model's 14^2 graph to local windows at evaluation reduces CFP-FP accuracy by 0.21 pp. Training with this restriction for five epochs instead reduces accuracy by 2.03 pp relative to its matched reference. Evaluation-time sensitivity therefore does not predict the effect of imposing the same restriction during training; the underlying learning mechanism remains unresolved.

A separate evaluation-only experiment at 28^2 uses four windows and achieves $1.159\times$ batch-32 throughput, with LFW/CFP-FP changes of $+0.033/+0.029$ pp. Sixteen windows yield $1.225\times$ throughput and a 0.214 pp CFP-FP reduction. Single-image speed is $0.98\times$ the reference. These measurements are specific to one GPU setup. The 14^2 training experiment does not validate the 28^2 inference setting, and the small positive accuracy differences require repeated evaluation before interpretation as gains.

Full-schedule and cross-dataset results. At full reported precision, twenty-epoch accuracy is $96.819 \pm 0.073\%$, with extreme-pose GAR $92.02 \pm 0.99\%$. The deficit relative to R50's fixed 93.58% GAR is 1.56 ± 0.99 pp. CASIA-WebFace contains approximately 490K images of 10.5K identities. Its added render stream uses a 12.5% proportion and 7,000 near-frontal source images. For seeds 2026/2027/2028, paired augmentation gains are $+0.90/+1.90/+1.43$ pp on CFP-FP and $+5.06/+6.54/+7.04$ pp in extreme-pose GAR. Yaw-averaged grid/content gaps are approximately $0.000/0.005/0.002$, and 75° gaps are $-0.033/-0.035/-0.037$. These rounded summaries do not replace full-precision source arrays.

Scope of the theoretical interpretation. An earlier cosine-law extension beyond the fixed lattice was inconsistent with reported graph-overlap measurements and is not retained. Results about mapping an entire closed surface into a plane do not directly constrain recognition from a visible surface subset. Borsuk–Ulam, continuous canonicalization, and tri-plane synthesis involve distinct settings and assumptions; the present argument does not rely on those analogies. Successful pose augmentation likewise does not establish that it is the only effective intervention.

F.1 Implementation of the four grid-preserving interventions

The four interventions summarized in Table 3 and above share one landmark input and one training recipe; their implementations are specified here so that the scope statements of Section 4 can be checked against what was actually run.

Shared landmark infrastructure. All three landmark interventions consume the same input: five facial landmarks per image (left eye, right eye, nose tip, left and right mouth corner) detected with SCRFD over every one of the 5.8 M MS1MV2 (Guo et al., 2016) training images and cached as a $5.8\text{M} \times 5 \times 2$ array in the 112-px aligned-crop coordinate frame, kept synchronized with the training-time horizontal flip. The canonical ArcFace five-point template serves as the reference frame: $(38.29, 51.70)$, $(73.53, 51.50)$, $(56.03, 71.74)$, $(41.55, 92.37)$, $(70.73, 92.20)$. Evaluation images use per-benchmark landmark caches built the same way. A per-sample gate drops the intervention for any image whose landmarks are non-finite; that image then follows the unmodified grid path. Every landmark backbone is written so that "no landmarks" is bit-exact with GraphFace-V, which is what makes the within-checkpoint enabled-versus-disabled comparisons of Appendix H possible.

Training recipe (identical for every run below). The GraphFace-V recipe: $\alpha = 0$ (feature-distance k-NN, absolute and relative positional terms kept), ArcFace loss with margins $(1.0, 0.5, 0.0)$, 512-d embedding, SGD with learning rate 0.1, momentum 0.9, weight decay 5×10^{-4}, one warm-up epoch, per-GPU batch 256 on two GPUs (512 effective), fp16, Partial-FC sampling rate 0.1, seed 2026, a 20-epoch schedule. Pilots are the same schedule read at five epochs; the positional-embedding warp is a full 20-epoch run.

(1) Positional-embedding warp. The absolute positional embedding added after the stem (one tensor of shape $1 \times C \times 28 \times 28$, i.e. the 28^2 stage) is resampled per image with the similarity transform (Umeyama) that maps the image's five detected landmarks onto the canonical template, so that a grid cell carries the canonical positional code of the face part it actually contains. The transform is purely geometric—zero learnable parameters, so it cannot collapse the way a learned pose-conditioning path had—and the graph construction itself (k-NN, aggregation, relative position) is untouched. A frontal face has landmarks close to the template, so the transform is near identity and the warped embedding near the vanilla one. The first pilot had a flip bug (landmarks not mirrored with the image); it was fixed before the full run, which used the V schedule end to end. Outcome: d' 3.494 against 3.558 for V, $\Delta d' = -0.064$, CI $[-0.110, -0.018]$, $p = 0.003$. The diagnosis ran six variants: three independent landmark sources agree within ± 0.005 d', so landmark quality is not the cause; the Umeyama residual is 9.22 px on the clean evaluation images against 3.14 px on training images; warping frontal faces only is indistinguishable from OFF; and beyond 70° the false-non-match rate at $1\,\%$ false-match worsens by 3.7–4.8 pp. Closed as "warp-dependence plus the limits of 2-D geometry": a 2-D similarity cannot undo an out-of-plane rotation.

(2) Landmark-anchored lattice. Applied at stage 1 only (14^2 nodes, 8 px per cell), the stage at which the probe of Section 3.1 localizes the anchoring. A hybrid lattice is built from the five landmarks: each landmark contributes its own position plus eight points at radius 8 px and eight at 16 px ($5 \times 17 = 85$ points). Each point overrides the sampling position of the grid cell that contains it; every other cell keeps its center, so the node count stays 196 and the node–cell correspondence is kept by construction, so the reach bound (Section 4) applies. At stage-1 entry the feature map is bilinearly resampled once at these positions, and both stage-1 graph layers and everything downstream operate on the anchored frame. Zero parameters; with positions equal to cell centers the resampling returns the exact input, so OFF is bit-exact with V. The construction gate (step-0 exactness, engagement, confound-matched control) passed before the five-epoch pilot was launched.

Why the run is invalid. The training script delivered landmarks to a backbone only when the network name contained the substring `lmwarp`; the anchored-lattice and distance-bias configurations did not, so both pilots trained without a single landmark—as pure GraphFace-V. The guard that existed to catch exactly this ("do the landmarks engage at step 1?") was itself placed behind the condition "landmarks were supplied", so it printed nothing and passed in silence. The fault was found on 2026-07-13, after the pilot had been evaluated. The fix gates the loader on the configuration declaring a landmark path and runs the engagement assertion unconditionally for every landmark-consuming backbone, aborting the run if the measured anchor displacement is zero. The -0.686 pp the pilot reported is therefore a train/test mismatch cost—landmark-free weights evaluated with landmarks—not the method's. The pilot was not repeated: the displacement measurement of Table 17 and the reach bound of Section 4 showed that the sampling rule could not reach displaced content (the anchored positions are sub-cell by construction, capped at 0.707 cells); whether the intervention would nonetheless have changed recognition was not tested.

(3) Landmark bias on the k-NN distance. Same stage, same 85-point lattice, but the feature path is left entirely unchanged: only the spatial term of the stage-1 neighbor-selection distance is computed on the anchored coordinates. The selection distance is the hybrid rule of Equation (2) with the anchored coordinates substituted, $d = 0.5\,\mathrm{minmax}(d_{\text{spatial}}) + 0.5\,\mathrm{minmax}(d_{\text{feature}})$, followed by the usual relative-position term and top-k. Only the stage-1 neighbor-selection module is swapped; parameters and initialization are those of V, and the selection runs without gradient, so the intervention changes which edges exist and nothing else. With no landmarks the module delegates to the parent and is bit-exact with V. After the loader fault above, the pilot was re-run with landmarks verified to engage. Against its own OFF arm the intervention passes its registered gate (CFP-FP $82.686 \to 83.914$, $+1.228$ pp; d' $+0.080$); against a budget-matched V control trained without the prior it is 1.04 pp *behind* (83.914 vs 84.957). Decomposing the change in the edge set attributes

$2.4\,\%$ of it to the landmarks and $97.6\,\%$ to the spatial prior itself, which is why this arm is recorded above as an ablation of a stage-1 spatial prior rather than as a landmark intervention.

(4) Graph-branch variant. The same anchored resampling confined to the input of the graph branch, leaving the feed-forward and residual paths on the grid. It was specified but not implemented; the same sampling rule applies, and its recognition effect is untested.

The two routes outside the reach bound, for completeness. *Free offset field:* a learned per-node sampling offset at the 14^2 and 7^2 stages with the maximum displacement per axis set to 1.5 cells—deliberately above the 0.707-cell cap—trained for five epochs under a gate registered before launch, with no retry permitted. *Edge quota:* stage 1 is partitioned into 7×7-node windows; inside each window the ordinary $k = 9$ selection runs unchanged, and each node receives $m = 3$ additional edges reserved for nodes in *other* windows, chosen by routing to the window whose mean token is most similar and then taking the m nearest by feature distance. The two null arms use the same quota with grid-nearest cross-window cells (proximity) and with random cross-window nodes. All quota arms are evaluated zero-shot on a trained V checkpoint.

G Corruption evaluation and uncertainty

The five pixel-noise types are Gaussian, impulse, shot, speckle, and salt-and-pepper. Because some are signal-dependent or replace pixels, the family is not uniformly additive. The remaining thirteen types include blur, weather, illumination and color changes, compression, pixelation, and occlusion. Glass and zoom blur were omitted from the repeated eighteen-type evaluation for computational reasons; the earlier twenty-type results describe a different protocol.

Table 18: Mean accuracy (%) across severity-5 pixel-noise types. Augmented results average three training seeds. The final column compares augmentation with the fixed CNN reference, not with V.

Benchmark	R50	V	V + renders	Augmented − R50
LFW	74.59	75.88	73.02	−1.57
CFP-FP	59.54	62.25	60.74	+1.20

For pixel noise, SD across training seeds is 1.43 pp on LFW and 1.10 pp on CFP-FP; SD across corruption draws is 0.50 and 0.55 pp, respectively. Figure 4 uses reported combined SEs of $0.39/0.32$ for noise and $0.09/0.09$ for other types. Independent verification requires the estimator, aggregation order, and raw arrays. Three seeds evaluated under three draws produce nine seed–draw combinations rather than nine independent training runs.

Corruption seeds depend on type, view, image index, and draw. The later manuscript reports zero checksum mismatches across 498 settings, indicating identical input bytes for compared models. This check does not independently reproduce the evaluation. Across three draws, H exceeds the CNN on 5, 5, and 4 of the five noise types. Its noise-family advantage over R50 is 1.81 ± 0.30 pp for frontal pairs and 2.06 ± 0.82 pp for cross-pose pairs, with variation across corruption draws rather than training seeds.

The high-frequency ratio is computed on the stage-end feature maps at 14^2, 7^2, and 4^2. For each image we take the 2-D discrete Fourier transform of every channel, average the power spectra over channels, and sum the power at radial frequencies above two-thirds of the Nyquist radius; the ratio is this high-frequency energy for the Gaussian-noise (severity-5) image divided by that of the clean image, averaged over 512 CFP-FP images (256 near-frontal, 256 profile) corrupted once and shared by all models. The reported changes are the mean over the three augmented seeds of the ratio minus that of V; the pre-registered prediction concerned 14^2 and 7^2, and the 4^2 value is reported from the same run without having been part of that test. After augmentation, high-frequency ratios change by $-0.067, -0.092, -0.112$ at $14^2, 7^2, 4^2$, respectively. Greater high-frequency suppression co-occurs with lower noise accuracy, contradicting the specific prediction of weaker suppression. This result rejects that prediction without resolving all frequency-based explanations.

H ATTRIBUTION CONTROLS AND REPRODUCIBILITY REQUIREMENTS

Identity dependence of the diagnostic score. The exploratory score is the grid/content overlap gap $\Delta_{\text{anchor}} = J_{\text{content}} - J_{\text{grid}}$ of Section 3.1, read at 14^2 and 75° on a five-epoch checkpoint trained without landmarks, after applying landmark-anchored resampling at evaluation time with the landmarks of each arm (true, another identity's at the same yaw, spatially permuted, or frontal at every yaw). The quoted values are increases of that gap relative to the twenty-epoch V reference (-0.0895); the enabled-versus-disabled reference ($+0.0400$) is the same checkpoint with resampling switched off, so it isolates the checkpoint difference from the intervention. An exploratory score increases by $+0.0489$ with true landmarks and $+0.0486$ with identity-shuffled landmarks. The shuffled input reproduces approximately 99.4% of the apparent effect. A within-checkpoint enabled-versus-disabled comparison attributes $+0.0400$ to checkpoint/training-budget differences and $+0.0089$ to the intervention. The true-minus-shuffled difference of $+0.0003$ is contained within the latter effect rather than an additional independent contribution. The five arms are listed in Table 19.

Table 19: Five-arm specificity test of the exploratory score at $14^2/75^\circ$: landmark-anchored resampling applied at evaluation time to the same five-epoch checkpoint, 298 rendered identities. The gap is $J_{\text{content}} - J_{\text{grid}}$; the reference is the twenty-epoch V value of -0.0895.

Arm	Landmarks supplied to the lattice	Gap	Δ versus V
Disabled	none (bit-exact vanilla path, same checkpoint)	-0.0495	$+0.0400$
True	the render's own landmarks	-0.0406	$+0.0489$
Substituted identity	another identity's landmarks at the same yaw	-0.0409	$+0.0486$
Permuted field	the true displacement field permuted across cells	-0.0517	$+0.0378$
Pose-frozen	frontal landmarks at every yaw	-0.1017	-0.0122

Recovery of removed connections. Content-based rewiring changes CFP-FP accuracy by $+0.129$ pp after window restriction but by -0.128 pp on the intact global graph. The fractions of original edges restored are 0.648 for content-based, 0.291 for proximity-based, and 0.020 for random rewiring. Across these three conditions, restored-edge fraction (0.648, 0.291, 0.020) and accuracy gain ($+0.129$, $+0.086$, -0.086 pp) increase together; three conditions do not support an inferential test, and we report the ordering only. This pattern is consistent with recovery of removed connections, but three conditions cannot establish a unique causal mechanism statistically.

Evaluation randomness and batch-size sensitivity. The batch-size comparison uses batches of 64 and 128. Unchanged accuracy does not imply identical intermediate tensors. On CFP-FP the batch-size shift is $+0.129$ pp for V, $+0.157$ for H, and -0.186 for the posedeform variant (Table 20); the 0.467 pp maximum occurs on CPLFW for V. ViHGNN introduces a separate source of variation through random cluster initialization at evaluation. Repeated evaluation at fixed batch size yields CFP-FP accuracies of $95.757/95.586/95.686\%$, a range of 0.171 pp. These three values are from the determinism probe (unseeded, one benchmark, fixed batch); the Table 1 accuracy is a separate single fixed-batch evaluation and its d' comes from the seeded three-evaluation-seed adapter, so the three are different measurements. Its $d' = 3.321 \pm 0.006$ uses a fixed-seed adapter and three evaluation seeds. This stochastic variation is distinct from batch-dependent numerical effects.

Readout control. A convolutional baseline fitted with the same GDC readout reaches 97.83% on CFP-FP against 97.67% for the standard readout ($+0.16$ pp, batch 64), so the readout alone does not reduce cross-pose accuracy on that backbone; its d' and extreme-yaw rate were not measured, and the interaction between the readout and the graph backbone remains untested.

H.1 PRE-REGISTRATION RECORDS

Every gate and prediction below was fixed before the measurement it governs. Since the project repository was created on 2026-07-21, registrations are committed to version control, either as a pre-registration file or in the docstring of the script that produces the number, so that the commit

Table 20: Evaluation batch-size sensitivity (batch 128 minus batch 64; identical checkpoint, pairs, precision, and evaluation code) across the seven benchmarks of Table 8, grouped by neighbor-selection mechanism. Zero entries are exact. The clean-mean shift is not consistently signed across the top-k models.

Backbone	Selection mechanism	Benchmarks moved (of 7)	Max $\lvert\Delta\rvert$ (pp)	Clean-mean Δ (pp)
R50	none (CNN)	0	0.000	0.000
ViT-S	none (transformer)	0	0.000	0.000
GreedyViG-S	distance threshold	0	0.000	0.000
V	top-k ranking	7	0.467	-0.047
H	top-k ranking	6	0.157	$+0.020$
Static resampling	top-k ranking	6	0.186	-0.051
ViHGNN[†]	fuzzy clustering (random initialization)	7	0.517	-0.075

[†]Listed for completeness but excluded from the batch-size band: its variation at a fixed batch is intrinsic (unseeded cluster initialization; 0.171 pp spread over three repeated evaluations, see text), so its entries do not isolate a batch-dependent numerical effect.

timestamp rather than the authors' memory is the evidence. Two measurements reported in Appendix D (the long-range edge prevalence and the transport agreement) and three gates of Section 5, namely the three-arm gate, the second-seed reconfirmation and the twenty-epoch promotion gate, predate the repository and carry their rules in the script header without a preceding commit; they are marked as such in Table 21 and in the project ledger. Table 21 lists the registrations for the experiments reported in this manuscript.

H.2 Materials required for empirical reproduction

Each result requires a checkpoint hash, configuration and execution command, data and preprocessing versions, exact pair list, evaluation batch size and precision, threshold procedure, and random seeds. Connectivity analyses additionally require fitted face models, visibility masks, correspondence maps, coordinate-quantization and collision rules, extraction layers, and graph-overlap aggregation procedures. Evaluation reports should specify checkpoint, batch size, precision, backend, and tie-breaking rule.

The affine experiment realigns the feature map extracted from a warped image. Reproduction requires the extraction layer, interpolation and padding conventions, and the procedure for recomputing the graph and identity embedding, together with a round-trip resampling control. The reported similarity summaries do not uniquely determine these choices.

Intervention validity requires confirming that the intended input is active during training and that disabling the modification numerically recovers the stated baseline. Preregistration claims require timestamped records rather than manuscript descriptions alone. Additional view generation also requires checking for identity overlap between training and evaluation sets.

The supplementary package contains the manuscript, the analysis scripts that read the stored records, the records themselves, the retired-values ledger, and the pre-registration files. It does not contain training code, checkpoints, raw pair scores, or renderer assets. No new numerical experiment was performed for this revision. Proposed controls remain future work, and unavailable measurements have not been imputed.

Table 21: Pre-registered gates and predictions for the experiments reported here. "Script header" denotes a rule recorded in the header of the producing script before the repository existed (no preceding commit); commit hashes refer to the project repository.

Experiment	Registered	Form	Prediction or decision rule	Outcome
Three-arm alignment test (Section 5, Table 4)	2026-07-15, before any arm finished	Script header and specification; pre-repository	Success requires $(d) > (b)$ and $(d) > (c')$ on CFP-FP, where (b) is augmentation alone, (c') map-permuted alignment and (d) true alignment; mechanism signal $D_{\mathrm{corr}}(d) > D_{\mathrm{corr}}(b)$ and $D_{\mathrm{corr}}(d) > +0.12$	Fails on the first conjunct: $(d) - (b) = -0.986$ pp, $(d) - (c') = +0.457$ pp; $D_{\mathrm{corr}}(d) = 0.034$ against 0.156 for (b)
Second-seed reconfirmation of augmentation alone (seed 2027, five epochs)	2026-07-15	Script header and configuration header; pre-repository	Thresholds fixed at half the seed-2026 effect against the frozen five-epoch reference: ΔCFP-FP $\geq +2.0$ pp and ΔGAR($> 70^\circ$) $\geq +6.0$ pp; LFW ≥ 0.967	$+4.01$ pp and $+14.32$ pp; reconfirmed
Twenty-epoch promotion gate (seed 2026 full run, Table 5)	2026-07-17	Script header; pre-repository	Promote if CFP-FP $\geq$ V three-seed mean $+0.5$ pp and GAR($> 70^\circ$) improves by $\geq +4.0$ pp, with LFW ≥ 0.967; an alternative "specialist" condition required IJB-C TAR at FAR 10^{-6} to improve by $\geq +1.5$ pp; promotion launches seeds 2027 and 2028	$+1.21$ pp CFP-FP and $+6.91$ pp GAR: promote (IJB-C $+0.24$ pp, specialist condition not met)
Unrestricted-offset pilot (Table 3, Appendix F)	2026-08-25, commit `c73940b`	Pre-registration file	At 56,860 steps: CFP-FP $-84.957 \geq +1.0$ pp and GAR($> 70^\circ$) $-62.716 > 0$ gives "promising", otherwise terminate with no retry. Annotations: the run does not test the hypothesis if fewer than 5% of nodes exceed 0.707 cells; saturation flagged if more than 50% exceed $0.9s$. Predicted: terminate, with offsets leaving the cell	Terminate: -1.200 pp and -0.494 pp; 71% and 91% of offsets beyond 0.707 cells at 14^2 and 7^2 (in-cell annotation not raised; saturation 44%, not raised)
CASIA-WebFace replication (Section 3.1, Table 5)	2026-08-25, commit `9332fa5`	Pre-registration file	Gate 1: augmentation minus baseline > 0 on CFP-FP and on GAR($> 70^\circ$), sign only. Gate 2: baseline 14^2 grid/content gap averaged over 15–75° is < 0, with the identity-bootstrap interval deciding "reproduced" (excludes 0) or "direction only". Predicted: both gates pass	Gate 1 passes ($+0.90$ pp, $+5.06$ pp); Gate 2 fails (gap $+0.000$, interval $[-0.002, +0.002]$); the connectivity diagnosis is scoped to MS1MV2
CASIA-WebFace at three seeds (Section 3.1)	2026-08-28, commit `2adc104`	Pre-registration file	Read-out on the baseline 14^2 gap: three-seed mean with 95% t-interval, in the precedence sign split $>$ upper bound < 0 (diagnosis extends) $>$ lower bound > 0 (reversal) $>$ interval contains 0 (scope kept). Predicted: scope kept	Scope kept: $+0.002 \pm 0.003$, t-interval $[-0.004, +0.009]$; Gate 1 sign reproduced at all three seeds
Frequency-domain account of the noise cost (Appendix G)	2026-08-06, commit `27e9a66`	Script docstring	H1: under Gaussian noise the high-frequency ratio of the augmented model exceeds that of V at both 14^2 and 7^2 (weaker suppression), with the same sign at the seed minimum, and the pattern does not recur under fog; pass, partial or fail	Fail with the sign reversed: -0.067 and -0.092 at 14^2 and 7^2; the account is withdrawn
Non-noise corruption types, three seeds by three draws (Section 5, Figure 4)	2026-08-25, commit `419fdc8`	Pre-registration file	P1: the 13-type CFP-FP mean of augmentation minus V stays positive and quotable ($\lvert\mathrm{mean}\rvert > 2\,\mathrm{s.e.}$). P2: the 13-type LFW mean stays within ± 0.5 pp. P3: frost, occlusion, pixelate, saturate and hue stay positive on all three draws; the count of positive types is not expected to hold. P4: six types deterministic, seven stochastic	All four confirmed: $+0.748$ (s.e. 0.091); -0.285 (s.e. 0.087; within ± 0.5 but above the quotation threshold, so "neutral" is not claimed); five named types positive on every draw, count $11/12/11$; 6 deterministic and 7 stochastic
Severe-noise re-measurement of H against R50 at three corruption draws (Appendix G)	2026-08-09, commit `0031a24`	Pre-registration file	Registered verbatim: "P1 — the family mean survives. H − R50 stays positive and lands within ± 0.7 pp of $+1.89$. P2 — the 5/5 count does not survive. At least one of impulse/gaussian flips sign on at least one draw. P3 — the ordering H > V > R50 on the family mean holds." Read-out: a per-type number is quoted only if $\lvert\mathrm{mean}\rvert > 2\,\mathrm{s.e.}$ over draws; the count of types won is reported per draw, never averaged; the family mean is quoted as mean $\pm$ draw-SD; no seed term is quoted, since this checkpoint has one training seed	P1–P3 confirmed: frontal family mean $+1.81 \pm 0.30$ pp; counts $5, 5, 4$ across draws; ordering H > V > R50. P2 held for a reason other than the registered one: the count broke at shot noise, while impulse and Gaussian grew and stayed positive on all draws