AGE SUBSPACES IN 3D FACE REPRESENTATIONS: READABILITY, PREDICTIVE DEPENDENCE, AND TRANSFERABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding age-related face representations requires distinguishing whether age information is linearly readable, whether a fitted predictor relies on the identified directions, and whether those directions transfer across independently trained representations. We investigate these questions in 3DMM-derived shape representations using correlation-ranked PCA directions, controlled direction removal, and cross-run comparisons. Correlation-ranked PCA directions retain age-predictive performance with fewer dimensions than variance-ranked PCA directions. Under the evaluated removal protocol, removing the top ten correlation-ranked directions reduces the fixed predictor's R^2 from 0.588 to approximately 0.16, whereas dimension-matched random removal reduces R^2 by only 0.017. Across training runs, the discovered subspaces exhibit overlap close to the random baseline in native coordinates, and directly transferring these directions across runs produces limited removal effects. These findings establish predictive dependence within the measured representations without implying complete age-information removal or intrinsic semantic differences across models. We further introduce Age Subspace Learning (ASL), a PCA-based decomposition with identity supervision, for age-robust face representation. Together, the results distinguish age readability, predictor dependence, and direct transferability, clarifying the interpretation and scope of age-subspace analysis.

1 INTRODUCTION

Understanding how learned representations encode high-level semantic attributes is a central problem in representation learning (Bengio et al., 2013). Facial age provides an informative setting for studying this problem because age-related variation affects facial geometry and appearance while interacting with identity-relevant information. Advances in face editing and generative modeling have improved visual realism and controllability Shen et al. (2020a); Ho et al. (2020); Zhang et al. (2023); Zhao et al. (2019); Yang et al. (2023). However, successful generation or recognition alone does not reveal where age information resides in a representation, how a predictor uses it, or whether its geometric organization is consistent across training runs.

Addressing these questions requires distinguishing three complementary properties: *readability*, *predictive dependence*, and *transferability*. Readability concerns whether age information can be extracted from selected representation directions by a linear probe. Predictive dependence concerns whether a fitted predictor relies on those directions, as assessed by controlled removal. Transferability concerns whether directions identified in one representation retain their predictive relevance when applied to another. Evidence for one property does not establish the others: a readable signal need not be essential to a particular predictor, and directions that support prediction within one representation need not transfer directly across independently trained representations.

We investigate these properties in 3D Morphable Model (3DMM)-derived shape representations Sanyal et al. (2019). Our analysis uses principal components as an orthonormal set of candidate directions and ranks them by their absolute correlation with chronological age. We first compare this ranking with explained-variance ranking to assess how efficiently the selected directions retain age-predictive performance. We then remove selected directions while keeping the age

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: Qualitative visualization obtained by moving along an age-related direction discovered in the analyzed representation. The generated faces show progressive age-related appearance changes while largely retaining identity cues. This example illustrates the semantic interpretability of the discovered direction within this representation; it is not evidence that the same direction is shared across independently trained models or representation spaces.

predictor fixed and compare the resulting performance changes with dimension-matched random and variance-ranked removal. Finally, we quantify cross-run subspace overlap and compare the effects of removing directions identified within the target run with those transferred from another run. Figure 1 illustrates age-related variation obtained by traversing a discovered direction within one analyzed representation.

Our experiments reveal three findings. **First**, correlation-ranked principal components retain age-predictive performance in fewer dimensions than variance-ranked components, demonstrating that explained variance and age predictiveness are distinct selection criteria. **Second**, removing the selected age-related directions substantially degrades the fixed predictor's performance relative to dimension-matched random removal, establishing predictive dependence beyond correlation alone. **Third**, the discovered subspaces exhibit overlap close to the random baseline and limited direct transfer across training runs in native coordinates. This last finding establishes a limitation of direct reuse; without explicit alignment, it does not distinguish coordinate misalignment from differences in the underlying semantic organization.

To connect this analysis with cross-age face representation, we additionally evaluate Age Subspace Learning (ASL), a PCA-based decomposition with identity supervision, as an application-level case study. The diagnostic analysis evaluates age-related directions, whereas the downstream experiments assess the utility of the ASL decomposition for identity representation. These evaluations provide complementary evidence without assuming that reduced age predictability necessarily improves identity preservation.

Our main contributions are:

- **A systematic analysis of age subspaces.** We evaluate readability, predictive dependence, and transferability as distinct properties of age-related directions in 3DMM-derived shape representations.
- **Controlled evaluation of predictive dependence.** We compare correlation-ranked, variance-ranked, and random direction removal, demonstrating that the fixed age predictor depends strongly on the correlation-ranked directions.
- **Cross-run evaluation of subspace overlap and transfer.** We quantify geometric overlap and compare within-run and cross-run removal effects, revealing limited direct transferability under the native-coordinate protocol.

2 Related Work

Cross-age face recognition. Cross-age face recognition aims to mitigate the influence of facial aging on identity recognition Huang et al. (2021). Early approaches relied on age normalization

or age synthesis; more recently, deep face recognition models Huang et al. (2020); Boutros et al. (2022) such as ArcFace Deng et al. (2019) and MagFace Meng et al. (2021) have substantially improved cross-age recognition, and several works further combine 3D Morphable Models (3DMMs) or adversarial learning to suppress age-related variation Bouritsas et al. (2019); Bas et al. (2017). However, these methods treat age as an undesirable factor to be removed, and focus on recognition accuracy rather than the intrinsic geometry of age representation.

Face age editing. Face age editing aims to generate photorealistic facial appearances across ages while preserving identity An et al. (2021); Zhang et al. (2017); Wang et al. (2018). Early methods relied on GAN inversion and latent space manipulation Xia et al. (2022), while recent diffusion-based approaches have improved generation quality and controllability Fuest et al. (2026). Representative models such as SDXL Podell et al. (2024) and Cradle2Cane Liu et al. (2026) achieve realistic lifelong aging through conditional generation. Yet age is still represented as a conditional label or an implicitly learned latent variable, and how age is geometrically encoded in the underlying representation is rarely investigated.

Disentanglement and representation geometry. A central goal of representation learning is to identify latent factors that explain data variation (Bengio et al., 2013), and a substantial body of work has pursued this through disentanglement Locatello et al. (2019); Chen et al. (2020); Higgins et al. (2017); Kumar et al. (2017); Chen et al. (2016). In the visual domain, methods such as InterFaceGAN Shen et al. (2020b), GANSpace Härkönen et al. (2020), and recent 3D-aware generative models Attaiki et al. (2025); Wang et al. (2024); He et al. (2020) show that high-level semantics can be associated with interpretable latent directions and manipulated for controllable editing. However, these methods primarily focus on *discovering* semantic directions, treating each attribute as an independent linear axis while paying little attention to the *geometric organization* of these factors: whether an attribute is encoded by a low-dimensional subspace rather than a single axis, whether it is stable and transferable across independently trained representations, and whether it plays a functional role or merely correlates with the attribute.

Summary. Prior work establishes that age can be manipulated or suppressed for downstream tasks, but it does not clearly separate three representation-level questions: whether age directions are linearly readable, whether a trained predictor functionally depends on them, and whether they transfer across independently trained representations. Our analysis focuses on this distinction in 3DMM shape space Gerig et al. (2018); Li et al. (2017).

3 METHOD

3.1 RESEARCH QUESTIONS AND SCOPE

We organize the study around three complementary questions.

RQ1: Can age-related directions be discovered? We ask whether chronological age can be linearly read from individual directions and whether correlation-based ranking identifies predictive directions more efficiently than ranking by feature variance.

RQ2: Do the discovered directions have predictive functional importance? We remove selected directions while keeping the fitted age predictor fixed. A larger performance decrease than matched random removal indicates that the predictor relies on those directions. This is a model-behavior intervention; it should not be interpreted as evidence that the removed directions constitute the complete biological aging mechanism.

RQ3: Are the directions stable and directly transferable across runs? We compare independently trained runs that produce same-dimensional 3DMM-derived shape representations and test whether a subspace discovered in one run can be removed from another run without re-estimating or aligning the coordinates. Failure of this direct transfer indicates limited native-coordinate transferability. Without an explicit cross-representation alignment and a corresponding alignment-quality check, it does not establish that the underlying age semantics are intrinsically different.

3.2 Problem Formulation

Given an 80-dimensional 3DMM-derived shape representation $\mathbf{z} \in \mathbb{R}^{80}$, ASL decomposes it into an age-related linear component and an orthogonal complement,

$$\mathbf{z} = \mathbf{z}_{\text{age}} + \mathbf{z}_{\text{comp}}. \tag{1}$$

Let $V_k = [\mathbf{v}_1, \ldots, \mathbf{v}_k] \in \mathbb{R}^{80\times k}$ contain an orthonormal set of selected directions. The projection matrices are

$$P_{\text{age}} = V_k V_k^\top, \qquad P_{\text{comp}} = I - P_{\text{age}}, \tag{2}$$

and

$$\mathbf{z}_{\text{age}} = P_{\text{age}}\mathbf{z}, \qquad \mathbf{z}_{\text{comp}} = P_{\text{comp}}\mathbf{z}. \tag{3}$$

The two projected components are orthogonal by construction. This algebraic orthogonality should not be interpreted as a guarantee of semantic disentanglement. In particular, the complementary component may still contain residual age-related information, and identity preservation is evaluated empirically in the downstream task.

3.3 Age Subspace Learning

Application subspace. In ASL, PCA is applied to the learned shape representation to obtain orthonormal geometric directions $\{\mathbf{v}_i\}_{i=1}^{80}$ ordered by explained variance. The correlation-guided ranking used in the diagnostic analysis is intended to characterize age predictability, whereas the PCA decomposition provides the explicit orthogonal coordinate system adopted by ASL for the downstream application. The application model uses the first k directions as an age-associated subspace,

$$S_{\text{age}}^{\text{ASL}} = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k), \qquad S_{\text{comp}} = (S_{\text{age}}^{\text{ASL}})^\perp. \tag{4}$$

Unless otherwise stated, the task experiments use $k = 10$. We do *not* interpret this choice as evidence that the first ten PCs contain a fixed percentage of all age information; predictive compactness is evaluated separately with K_{80} and K_{90} in Section 4.2.

Identity enhancement. A pretrained ArcFace network provides an identity supervision signal

$$\mathbf{f}_{\text{id}} = \text{ArcFace}(I), \tag{5}$$

and a lightweight mapping network $\hat{\mathbf{f}}_{\text{id}} = h(\mathbf{z}_{\text{comp}})$ maps the complementary 3DMM representation into the identity embedding space. Identity consistency is optimized with cosine distance,

$$L_{\text{id}} = 1 - \frac{\hat{\mathbf{f}}_{\text{id}}^\top \mathbf{f}_{\text{id}}}{\|\hat{\mathbf{f}}_{\text{id}}\| \, \|\mathbf{f}_{\text{id}}\|}. \tag{6}$$

The application objective is written compactly as

$$L_{\text{total}} = L_{\text{task}} + \lambda_{\text{id}} L_{\text{id}} + \lambda_{\text{age}} L_{\text{age}}, \tag{7}$$

where $L_{\text{task}} = L_{\text{reg}} + L_{\text{rec}} + L_{\text{perc}}$ and L_{age} denotes the auxiliary age term used by the application model. The age-related objective provides auxiliary supervision for matching the generated output to the target age. It is evaluated together with the downstream task objective and is not interpreted as a direct intervention that removes all age information from the representation.

3.4 Correlation-Guided Diagnostic Protocol

The mechanism analysis uses the full PCA basis as a set of orthonormal candidate directions and ranks those directions by age association rather than by explained variance. For centered representation $\tilde{\mathbf{z}} = \mathbf{z} - \boldsymbol{\mu}$, the score of direction $\mathbf{v}_i$ is

$$a_i = \left|\text{corr}(\mathbf{v}_i^\top \tilde{\mathbf{z}}, y_{\text{age}})\right|. \tag{8}$$

Let π sort a_i in descending order and $U_K = [\mathbf{v}_{\pi(1)}, \ldots, \mathbf{v}_{\pi(K)}]$. We compare this correlation-ranked subspace with (i) the top-K variance-ranked PCs and (ii) random K-dimensional orthonormal subspaces of the same ambient dimension.

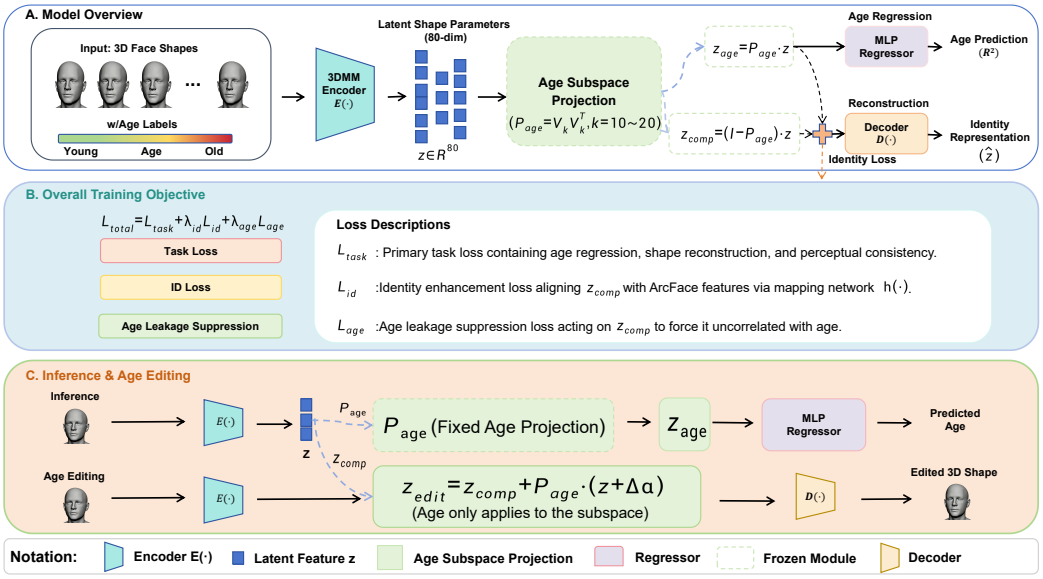


Figure 2: Overview of the Age Subspace Learning (ASL) framework. The representation analysis in Section 4.2 is used to validate and delimit the role of age-related directions; it does not assume a universal cross-model subspace.

For functional validation, the age predictor is kept fixed and the selected component is removed,

$$\mathbf{z}^{(-K)} = \mathbf{z} - U_K U_K^\top (\mathbf{z} - \boldsymbol{\mu}). \tag{9}$$

The resulting change in R^2 measures dependence of the fitted predictor on the removed directions. It does not answer whether age information could be recovered after retraining a new predictor on the residual representation.

For two orthonormal K-dimensional bases U and W expressed in the same 80-D 3DMM coefficient coordinates, we measure subspace similarity using the normalized principal-angle similarity:

$$\mathrm{sim}_\theta(U, W) = \frac{1}{K} \sum_{j=1}^{K} \left(1 - \frac{\theta_j}{\pi/2}\right), \tag{10}$$

where θ_j are the principal angles between the two subspaces. Random baselines use independently sampled orthonormal K-subspaces in the same $d = 80$ ambient space and are estimated empirically using repeated random sampling.

For cross-run transfer, a source subspace is applied directly to the target representation without any explicit cross-run alignment. We refer to this as *native-coordinate transfer*. Cross-architecture or cross-parameterization comparisons are not used as main evidence unless a mathematically valid mapping to a common coordinate system is available.

Why correlation ranking and matched controls? Variance ranking answers which directions explain the largest amount of total representation variation, whereas our diagnostic question is which directions are most informative about age. The two criteria need not agree: an age-predictive direction can have modest feature variance, and a high-variance direction can be dominated by identity, pose, or other factors. We therefore use variance-ranked PCs as a structured baseline rather than as the definition of the age subspace. Random orthonormal subspaces provide a dimensionality-matched control for the fact that removing any K directions necessarily perturbs the representation. The comparison among correlation-ranked, variance-ranked, and random removal isolates whether the observed degradation is specific to age-associated directions rather than a generic consequence of deleting coordinates.

Readability, dependence, and recoverability. The diagnostic protocol deliberately distinguishes three levels of evidence. A high correlation or high R^2 establishes *readability*: age information can

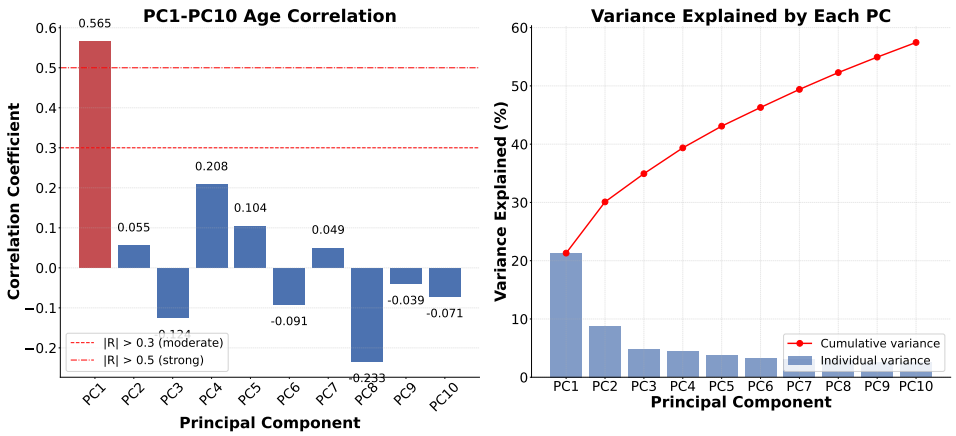


Figure 3: Age association and explained variance of the first ten principal components, displayed in descending order of explained variance. The left panel reports signed Pearson correlations with chronological age; the right panel reports individual and cumulative explained variance. Absolute correlations are used for direction ranking in the diagnostic analysis.

be extracted by the chosen linear probe. A larger drop after targeted removal establishes *predictor dependence*: the fixed probe uses the removed directions. Neither statement implies that all age information has been erased from the residual representation. Testing *recoverability* would require fitting a new predictor after removal and measuring how much age information can be reconstructed from the remaining coordinates. We keep this distinction explicit because fixed-predictor intervention and retrained-predictor evaluation answer different questions and should not be pooled into a single notion of age information removal.

4 Experiments

4.1 Experimental Settings

Datasets. Experiments are conducted on FFHQ Bai et al. (2023), CelebA-200 Serianni et al. (2025), and AgeDB-400 Moschoglou et al. (2017). FFHQ is used for the application training protocol, where approximately 42,500 adult face images (20–60 years old) are retained after face alignment and resized to 512×512. CelebA-200 and AgeDB-400 are used for representation analysis and cross-age evaluation under their corresponding protocols. Results obtained under different sample sizes or domains are not pooled into a single compactness claim.

Implementation details. The framework is implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU using Adam with a learning rate of 1×10^{-4} and weight decay of 0.01 Paszke et al. (2019); Kingma & Ba (2014). Training is performed for 10 epochs with FP16 mixed precision and a batch size of 2. The application experiments use $k = 10$ unless otherwise stated, whereas the diagnostic analysis scans K and reports predictive-retention statistics separately. Seeded mechanism experiments use five runs $\{42, 123, 456, 789, 999\}$.

Evaluation metrics. We report Pearson age correlation, age-prediction R^2, cross-age identity similarity, predictive retention $R^2(K)/R^2_{\text{full}}$, and the principal-angle similarity in Eq. 10. We define K_{80} and K_{90} as the smallest dimensions reaching 80% and 90% of full predictive R^2, respectively. These quantities measure *predictive compactness*; they are distinct from explained feature variance and are not interchanged in the discussion.

4.2 Mechanism Analysis

The analysis follows the three research questions in Section 3.1. Fig. 3, Fig. 4, Fig. 5, and Table 1 summarize the main evidence; secondary curves and per-seed tables are moved to the appendix.

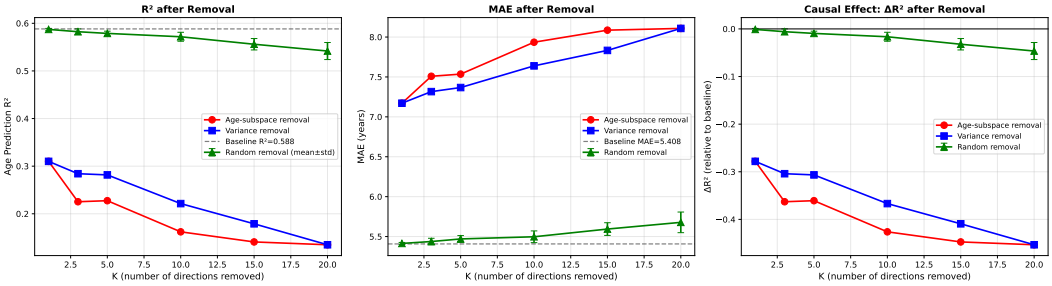


Figure 4: Predictive removal effect (RQ2). Removing correlation-ranked age directions causes a larger decrease in the fixed predictor's R^2 than removing variance-ranked or random directions of the same dimensionality.

Table 1: Representation analysis under the reported protocols. K_{80} and K_{90} refer to predictive retention, not explained feature variance.

Question	Quantity	Result
RQ1	K_{80}: variance / correlation	20 / 13
RQ1	K_{90}: variance / correlation	37 / 23
RQ2	ΔR^2: correlation / random removal	-0.426 / -0.017
RQ3	Cross-run similarity vs. random	$\approx$ random
RQ3	ΔR^2: own / direct transfer	-0.510 / -0.018

RQ1: readability and ranking. Correlation-ranked directions retain predictive information with fewer dimensions than variance-ranked directions. Under the reported protocol, K_{80} decreases from 20 to 13 and K_{90} from 37 to 23 (Table 1). This shows that explained feature variance and age predictiveness are not equivalent criteria. The result supports a distributed linear age signal with a stronger leading component and additional predictive directions, but it does not imply that a fixed number of PCs universally captures a fixed percentage of age information.

RQ2: predictive functional importance. Fig. 4 evaluates the fixed predictor after removing top-K directions. The full representation gives $R^2 = 0.588$ in this protocol; after removing the top ten correlation-ranked directions, R^2 falls to approximately 0.16, corresponding to $\Delta R^2 = -0.426$. Removing matched random directions changes R^2 by only -0.017. Variance-ranked removal is also weaker than correlation-ranked removal. These results show that the fitted predictor relies strongly on the selected directions. We use the terms *removal effect* and *predictive functional importance*, rather than claiming a causal mechanism underlying biological facial aging.

RQ3: stability and native-coordinate transfer. Fig. 5 shows that cross-run native-coordinate subspace similarity is close to the matched random baseline in the native coordinate system. In the direct transfer experiment, removing a run's own selected subspace reduces R^2 by 0.510 on average, whereas directly removing a subspace obtained from another run changes R^2 by only 0.018 (Table 1). Thus, the discovered directions transfer poorly *without alignment*. This result does not by itself show that age semantics are intrinsically different across representations: an alignment experiment such as orthogonal Procrustes, together with an overall alignment-quality measurement, is required to distinguish coordinate mismatch from genuinely non-shared structure.

4.3 Comparison with State-of-the-Art

We compare ASL with representative face-recognition and 3DMM baselines under the original downstream evaluation protocol. These results serve as application-level evidence and are deliberately separated from the age-prediction removal analysis: a direction being predictive of age, or its removal hurting an age predictor, does not automatically prove a gain in cross-age recognition.

Table 2 reports the downstream accuracies under this protocol. ASL achieves competitive performance compared with the listed baselines under this evaluation protocol. We use these task results,

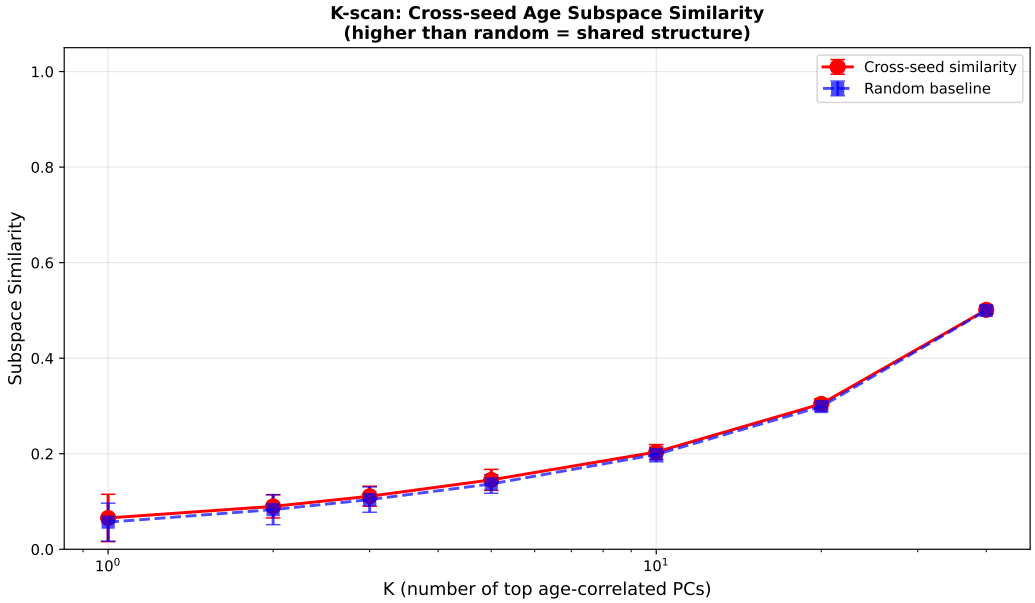


Figure 5: Cross-run native-coordinate subspace similarity (RQ3). Cross-run subspace overlap is compared with the matched random-subspace baseline across K. Without an explicit cross-representation alignment, this figure measures native-coordinate overlap rather than direct transfer performance or intrinsically different semantics.

Table 2: Comparison with state-of-the-art methods.

Method	LFW	CALFW	AgeDB	MORPH	CACD
ArcFace	99.20	95.45	97.35	98.19	97.86
MagFace	99.30	95.86	97.68	98.31	98.13
AdaFace	99.35	95.92	97.75	98.42	98.21
3DMM-Baseline	97.80	91.53	94.25	96.19	95.37
ASL (Ours)	**99.45**	**96.20**	**97.88**	**98.56**	**98.45**

together with the representation analysis above, as complementary evidence: the former evaluates application performance, while the latter explains which linear directions a fitted age predictor uses and where direct cross-run transfer breaks down.

4.4 Ablation Studies

The ablations serve a different purpose from the diagnostic experiments above. The diagnostic analysis asks what can be read from a representation and how a fixed predictor responds to controlled removal; the ASL ablations ask whether particular design choices improve the original cross-age application. Keeping these roles separate avoids using an age-prediction result as a surrogate for recognition performance.

Subspace dimension and projection. The original ASL application varies $k \in \{1, 5, 10, 15, 20\}$ and compares the original representation, PC1-only removal, and projection with the selected ASL subspace. These experiments test whether the downstream method benefits from operating on more than the leading variance direction and whether performance is sensitive to the chosen dimensionality. We retain $k = 10$ as the application setting used in the original protocol, but we do not reinterpret this hyper-parameter as a universal statement that ten components contain a fixed proportion of age information. The diagnostic K_{80}/K_{90} analysis is the appropriate place to discuss predictive compactness.

Identity enhancement and joint optimization. The identity-enhancement ablation evaluates whether explicit ArcFace supervision adds information beyond geometric projection alone. Under the original task protocol, removing identity supervision reduces cross-age identity consistency, while the full model improves identity preservation. The joint-optimization study in Appendix A.4 further separates the task, identity, and auxiliary age terms. These results support the application design, but they are not used to infer that the same age directions are shared across runs or architectures. Detailed ablation tables are reported in Appendix A.3.

5 DISCUSSION

What the removal experiment establishes. Removing the selected age-related directions causes a substantially larger decrease in R^2 than removing random directions of the same dimensionality. This result demonstrates that the fixed age predictor depends strongly on the selected directions within the analyzed representation. Such predictive dependence does not establish complete age-information removal: a newly trained predictor may recover age information from the remaining coordinates. Fixed-predictor removal and residual-information recovery therefore address different questions.

What weak cross-run transfer means. Our cross-run analysis provides two complementary observations: the discovered subspaces exhibit overlap close to the random baseline in native coordinates, and removing directions transferred from another run has a much smaller effect on the target predictor than removing its own selected directions. Together, these findings indicate limited direct transferability under the evaluated protocol. They do not establish that age semantics are intrinsically different across representations, because coordinate misalignment may contribute to both weak overlap and weak transfer. We do not report transfer after Procrustes alignment and therefore leave open whether explicit alignment can improve cross-run transfer.

Relation to downstream cross-age recognition. The representation analysis and the ASL application experiments address related but distinct questions. The former evaluates the readability, predictive dependence, and direct transferability of correlation-ranked age directions. The latter evaluates a variance-ranked PCA decomposition with identity supervision for cross-age face representation. These results should not be interpreted as evidence that correlation-ranked removal directly produces the reported recognition gains. More generally, reducing age-prediction performance does not necessarily improve identity preservation, and recognition gains alone do not establish semantic disentanglement.

Limitations. Our findings are limited to the measured 3DMM-derived representations, datasets, and evaluation protocols. The analysis focuses on linear directions and a fixed age predictor; it does not establish whether residual age information remains recoverable by retrained linear or nonlinear predictors. Chronological age is also an imperfect proxy for visible facial aging. Finally, our cross-run conclusions concern comparisons and removal operations in native coordinates. Evaluating transfer after explicit alignment, together with alignment quality on held-out samples, is necessary to assess whether the observed transfer limitation persists beyond the current protocol.

6 CONCLUSION

We studied age subspaces in 3DMM-derived face representations by distinguishing readability, predictive dependence, and transferability. Correlation-ranked principal components retain age-predictive performance in fewer dimensions than variance-ranked components, and their removal substantially degrades a fixed age predictor relative to dimension-matched random removal. Across independently trained runs, the discovered subspaces exhibit near-random overlap in native coordinates and limited direct transfer, as measured by cross-run removal effects. These findings establish within-run predictive dependence while leaving open whether explicit alignment can improve transfer. We additionally evaluated ASL, a PCA-based decomposition with identity supervision, as an application to cross-age face representation. Together, our results show that reading out age information, establishing a predictor's dependence on selected directions, and transferring those directions across representations require distinct empirical evidence.

AI USE STATEMENT

The authors used AI-assisted tools for language and formatting support. All scientific claims, experimental results, and final manuscript decisions remain the responsibility of the authors.

ETHICS STATEMENT

This work uses publicly available face datasets and introduces no new human-subject data collection. Face recognition and age editing are dual-use technologies; the experiments are intended for representation analysis and should not be interpreted as validating deployment in surveillance or other high-risk settings. Dataset demographic bias may affect the measured geometry, and the present study does not establish demographic fairness.

REPRODUCIBILITY STATEMENT

Implementation details, evaluation metrics, and the main hyper-parameters are provided in the experimental settings. The seed-based mechanism experiments are conducted over five independent training runs with seeds $\{42, 123, 456, 789, 999\}$. The released code will include the core ASL implementation and the main analysis scripts for age-direction discovery, subspace removal, and cross-run evaluation. Some auxiliary components may not be included in the public release. Third-party pretrained models and weights are subject to their original licenses and will not be redistributed.

REFERENCES

Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang, et al. Partial fc: Training 10 million identities on a single machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1445–1449, 2021.

Souhaib Attaiki, Paul Guerrero, Duygu Ceylan, Niloy Mitra, and Maks Ovsjanikov. Ganfusion: Feed-forward text-to-3d with diffusion in gan space. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 3985–3995, 2025.

Haoran Bai, Di Kang, Haoxian Zhang, Jinshan Pan, and Linchao Bao. Ffhq-uv: Normalized facial uv-texture dataset for 3d face reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 362–371, 2023.

Anil Bas, Patrik Huber, William AP Smith, Muhammad Awais, and Josef Kittler. 3d morphable models as spatial transformer networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 904–912, 2017.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

Giorgos Bouritsas, Sergiy Bokhnyak, Stylianos Ploumpis, Michael Bronstein, and Stefanos Zafeiriou. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7213–7222, 2019.

Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1578–1587, 2022.

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

Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems*, 29, 2016.

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*, pp. 4690–4699, 2019.

Michael Fuest, Pingchuan Ma, Ming Gui, Johannes Schusterbauer, Vincent Tao Hu, and Björn Ommer. Diffusion models and representation learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Luthi, Sandro Schönborn, and Thomas Vetter. Morphable face models-an open framework. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 75–82. IEEE, 2018.

Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. *Advances in neural information processing systems*, 33:9841–9850, 2020.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5901–5910, 2020.

Zhizhong Huang, Junping Zhang, and Hongming Shan. When age-invariant face recognition meets face age synthesis: A multi-task learning framework. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7282–7291, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Abhishek Kumar, Prasanna Sattigeri, and Avinash Balakrishnan. Variational inference of disentangled latent concepts from unlabeled observations. *arXiv preprint arXiv:1711.00848*, 2017.

Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017.

Tao Liu, Dafeng Zhang, Gengchen Li, Shizhuo Liu, Senmao Li, Shiqi Yang, Boqian Li, Kai Wang, Yaxing Wang, et al. From cradle to cane: A two-pass framework for high-fidelity lifespan face aging. *Advances in Neural Information Processing Systems*, 38:20787–20818, 2026.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pp. 4114–4124. PMLR, 2019.

Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14225–14234, 2021.

Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 51–59, 2017.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, volume 2024, pp. 1862–1874, 2024.

Soubhik Sanyal, Timo Bolkart, Haiwen Feng, and Michael J Black. Learning to regress 3d face shape and expression from an image without 3d supervision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7763–7772, 2019.

Aaron Serianni, Tyler Zhu, Olga Russakovsky, and Vikram V Ramaswamy. Attention iou: Examining biases in celeba using attention maps. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 4386–4397, 2025.

Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9243–9252, 2020a.

Yujun Shen, Ceyuan Yang, Xiaoou Tang, and Bolei Zhou. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE transactions on pattern analysis and machine intelligence*, 44(4):2004–2018, 2020b.

Xin Wang, Hong Chen, Si'ao Tang, Zihao Wu, and Wenwu Zhu. Disentangled representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9677–9696, 2024.

Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7939–7947, 2018.

Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. Gan inversion: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(3): 3121–3138, 2022.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM computing surveys*, 56(4):1–39, 2023.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3836–3847, 2023.

Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5810–5818, 2017.

Jian Zhao, Yu Cheng, Yi Cheng, Yang Yang, Fang Zhao, Jianshu Li, Hengzhu Liu, Shuicheng Yan, and Jiashi Feng. Look across elapse: Disentangled representation learning and photorealistic cross-age face synthesis for age-invariant face recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 9251–9258, 2019.

A Additional Results and Ablation Studies

This appendix provides supplementary analyses of predictive retention and per-run removal effects, together with additional ablations of the ASL framework. We report the representation-level diagnostic analyses separately from the application-level ASL ablations because they address different questions and may follow different evaluation protocols.

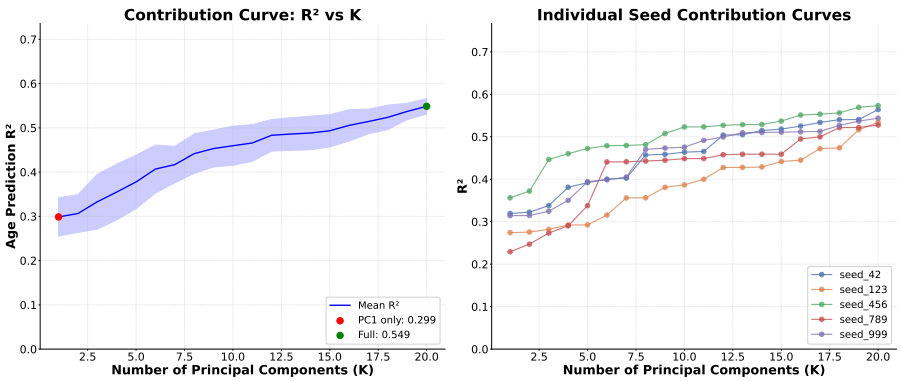


Figure 6: Age-prediction performance as the retained subspace dimension increases. Predictive retention is reported separately from explained feature variance; K_{80} and K_{90} denote the numbers of retained directions required to preserve 80% and 90% of the full predictive performance, respectively.

Table 3: Effect of removing PC1 on age prediction across the five reported runs.

Seed	Original R^2	After Removing PC1	Decrease
seed_42	0.678	0.523	0.155
seed_123	0.684	0.531	0.153
seed_456	0.542	0.399	0.143
seed_789	0.621	0.482	0.139
seed_999	0.660	0.512	0.148
Average	0.637	0.489	0.148

A.1 Predictive Retention Across Subspace Dimensions

Figure 6 shows how age-prediction performance changes as the retained subspace dimension increases. Predictive retention measures how much age-predictive capability is preserved by a subset of directions and is therefore distinct from the percentage of total feature variance explained by those directions. In the main text, we accordingly characterize predictive compactness using K_{80} and K_{90}, i.e., the numbers of directions required to retain 80% and 90% of the full predictive performance.

A.2 Per-Run Effect of PC1 Removal

To complement the top-K removal analysis in the main text, Table 3 reports the effect of removing only the leading principal component for each of the five runs. The decrease is consistent across runs, showing that PC1 contains substantial age-predictive information. At the same time, the stronger top-K analysis demonstrates that age-related predictive information is not restricted to a single dominant direction.

A.3 ASL Application Ablations

The following experiments evaluate individual design choices of ASL under the downstream application protocol. They complement the representation-level analyses in the main text but should be interpreted as application ablations rather than as evidence about cross-representation subspace transfer.

Effect of age-subspace dimension. We vary the age-subspace dimension over $k \in \{1, 5, 10, 15, 20\}$ while keeping the remaining settings fixed. This experiment evaluates the sensitivity of the downstream ASL model to the amount of age-associated variation retained by the subspace. Its role differs from the diagnostic K_{80}/K_{90} analysis, which measures predictive retention under the representation-analysis protocol.

Table 4: Ablation of the optimization objectives in ASL under the downstream application protocol. The last row corresponds to the complete ASL pipeline, including the full training objective and the proposed projection mechanism.

Method	Age Corr.	R^2	ID Sim.	Accuracy
Task Loss Only	0.85	0.605	0.43	93.12
Task + Identity Loss	0.51	0.25	0.51	94.85
Task + Age Loss	0.48	0.18	0.55	95.63
Task + Identity + Age Loss + Projection	0.45	0.12	0.59	96.41
Full ASL	**0.45**	**0.12**	**0.62**	**96.72**

Effect of age-aware projection. We compare three variants: the original representation, removal of PC1 only, and projection using the full ASL age subspace. This ablation examines whether modeling multiple age-associated directions provides additional benefit over suppressing only the leading PCA direction. The result is interpreted within the downstream ASL setting and does not imply that the same subspace can be transferred directly across independently trained representations.

Effect of identity enhancement. We further remove identity supervision while keeping the remaining components unchanged. Comparing this variant with the complete model evaluates the contribution of identity enhancement to cross-age identity preservation. This downstream result is complementary to, but conceptually distinct from, the age-prediction removal experiments used to analyze the functional importance of age-related directions.

A.4 Joint Optimization Ablation

We additionally study the contribution of the task objective, identity supervision, and the auxiliary age objective under the ASL application protocol. Table 4 reports the corresponding age- and identity-related metrics. Because these values are obtained under the application-level training and evaluation setting, the reported R^2 values are interpreted within this ablation and are not directly compared with the diagnostic R^2 values used in the representation-level removal analysis.