

SvedbergOpen
DISSEMINATION OF KNOWLEDGEInternational Journal of Artificial
Intelligence and Machine LearningPublisher's Home Page: <https://www.svedbergopen.com/>

Research Paper

Open Access

Multi-Domain Lifespan Face Age Synthesis Using StarGAN v2 with Attention-Based Identity Preservation and Real-ESRGAN Enhancement

Vikash Kumar Agarwal^{1*}, Dr. Subrajeet Mohapatra²¹Department of Cyber Defence, Jharkhand Raksha Shakti University, Ranchi, Jharkhand, India and Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India ORCID 0009-0008-8028-5371.E-mail: vikashhagarwal@yahoo.com²Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.ORCID 0000-0002-6919-4734. E-mail: subrajeet.mohapatra@bitmesra.ac.in**ABSTRACT**

Lifespan face age synthesis aims to transform a source face into plausible appearances at multiple target ages while retaining identity-defining structure. The problem is difficult because facial aging is neither a uniform texture operation nor a simple change of a single latent variable. Bone structure, facial proportions, skin texture, hairline, pigmentation and soft-tissue distribution change at different rates, while identity cues must remain stable. This paper presents AgeStarNet, a multi-domain framework that combines the scalability of StarGAN v2 with an Attention-Based Identity Preservation Module (AIPM) and a Real-ESRGAN enhancement stage. The proposed design treats aging as a controlled translation problem: a style pathway specifies the target age domain, while an ArcFace-derived identity representation provides an explicit anchor for identity-sensitive features. The AIPM uses eight-head cross-attention so that intermediate generator features can query source identity information before image synthesis is completed. A multi-component objective combines adversarial, style-reconstruction, diversity, cycle-consistency, identity and age-domain terms. A separate Real-ESRGAN stage converts the 64×64 synthesis into a 256×256 output and is fine-tuned on age-synthesized pairs. The experimental reports results on MORPH-II, UTKFace, CACD and FG-NET with FID, identity cosine similarity, age-estimation accuracy and SSIM as complementary measures. The reported UTKFace values are 12.4 FID, 0.913 identity cosine similarity, and 88.7% age-estimation accuracy. The core methodological idea is a bottom-up and open research-generation paradigm where evidence is accumulated from dataset construction, representation learning, synthesis, identity-preservation, enhancement, evaluation and reproducibility rather than considering the final image as the only research artefact.

Keywords: Face Age Synthesis; Face Age Progression; StarGAN v2; Identity Preservation; Real-ESRGAN

1. Introduction

The human face changes continuously during the life span but the visible process is heterogeneous. A child undergoes new facial proportions, an adolescent undergoes rapid structural and hair changes, adulthood is dominated by gradual changes in skin texture, pigmentation and soft tissue, and later life may introduce wrinkles, sagging and changes in hair density. A computational model therefore needs to learn both what should change with age and what should stay tied to a specific person. This dual requirement makes face age synthesis drastically different from ordinary image-to-image translation.

The research community has explored this problem through physical models, statistical representations, morphing, autoencoders and, more recently, deep generative models. Classical approaches provided useful interpretations of facial growth but were constrained by hand-designed assumptions and limited ability to represent stochastic texture variation. GAN-based methods changed the field by learning image distributions directly. CAAE demonstrated that age-conditioned latent representations could be used for progression and regression, while later methods introduced identity losses, high-resolution generators, geometric priors, flow-based representations and lifespan-oriented architectures [1]– [8]. A recent review also shows that identity preservation, age accuracy, realism and evaluation protocol remain recurring challenges across both classical and deep-learning approaches [9].

A second challenge is scalability. If each target age is handled by a separate generator, the number of learned components grows with the number of domains. StarGAN and StarGAN v2 address the more general multi-domain translation problem by sharing a generator and conditioning synthesis on target-domain information. StarGAN v2 further separates domain/style control from source content and supports diverse styles within a domain [10], [11]. This property makes it a natural backbone for lifespan synthesis, where age intervals can be treated as related but distinct visual domains.

A third challenge is identity drift. When age-conditioned style is injected repeatedly into a generator, identity-

bearing information can be weakened. ArcFace provides a compact face representation that is useful as an identity supervision signal [12]. However, a single global embedding loss does not explicitly tell the generator which intermediate spatial features need protection. Attention provides a mechanism for making the identity representation available to selected feature locations. The proposed AIPM therefore introduces cross-attention between generated features and source identity features rather than relying only on a final image-level identity penalty.

The final challenge is image quality. GAN aging systems often operate at moderate spatial resolution because high-resolution adversarial training is expensive. Real-ESRGAN was designed for blind real-world super-resolution and uses high-order degradation modeling together with a U-Net discriminator and spectral normalization [13]. In the present framework, it is used as a learned enhancement stage after age translation. This separation allows the age generator to concentrate on semantic transformation while the enhancement network concentrates on spatial detail.

Research Questions

RQ1: Can a single multi-domain generator cover six lifespan age domains without requiring a separate generator for each target range? RQ2: Does explicit cross-attention to source identity features reduce identity drift? RQ3: Does learned super-resolution improve perceptual quality without materially changing age-domain or identity measurements? RQ4: Can the proposed pipeline be evaluated through a transparent, bottom-up protocol that reports component-level contributions rather than only a final score?

The research gap addressed by AgeStarNet is consequently framed at the intersection of four requirements: (i) scalable multi-domain age control; (ii) explicit identity anchoring inside the synthesis pathway; (iii) high-resolution rendering that can be evaluated independently from semantic aging; and (iv) an evaluation protocol that makes component contributions and failure modes visible. The novelty claim is therefore architectural and methodological rather than a claim that every individual component is new in isolation.

Recent literature also makes the gap more specific. A 2023 review of face age synthesis organizes the field around datasets, evaluation criteria, classical methods and deep-learning methods, and identifies the continuing difficulty of identity preservation, age accuracy and limited longitudinal data [49]. More recent work has explored decorrelation constraints and facial geometry guidance for lifespan synthesis [50], while 2024 work has investigated deep cycle GANs with bias loss and perceptual identity supervision [51]. A 2025 biometric study has also examined identity-preserving aging and de-aging in the StyleGAN2 latent space and released synthetic data for cross-age benchmarking [52]. These directions suggest that the open problem is moving from basic age translation toward controllable, identity-stable and reproducible lifespan synthesis.

Contributions

The work contributes: (i) AgeStarNet, a unified StarGAN v2-based lifespan synthesis framework; (ii) an eight-head AIPM for explicit identity anchoring; (iii) a multi-component training objective; (iv) a Real-ESRGAN enhancement stage; (v) ablation and cross-dataset evaluation procedures; and (vi) an open/reproducibility-oriented research paradigm that makes dataset splits, metrics, component removal and reporting part of the method itself.

2. Related Work

A. Classical and Deep Face Aging

Early face-aging systems modeled age as a predictable transformation of facial shape and appearance. Active appearance and statistical models offered interpretable relationships between age and facial parameters, while physical and 3-D approaches attempted to model structural growth. These approaches established the important distinction between identity-related and age-related variation, but they generally required strong assumptions about how faces change. With larger datasets and deep learning, the emphasis shifted toward learning these transformations directly from data.

CAAE suggests a formulation that an encoder maps an input face into a latent representation conditioned on age, and adversarial learning drives realistic outputs [4]. Identity-preserved conditional GANs then make identity an explicit optimization goal. HRFAE talked about need of higher resolution age editing, PFA-GAN suggested progressive ageing and geometric information. AgeFlow uses normalizing flows to learn a more structured conditional representation. LATS focused on lifespan transformation across wider age ranges. These developments motivate the present work: a practical lifespan system requires both domain scalability and explicit control of identity.

B. Multi-Domain Translation

StarGAN introduced unified multi-domain translation with a single generator and domain conditioning. StarGAN v2 refined this idea by separating style encoding and mapping from source content, allowing diverse outputs within each target domain [10], [11]. CycleGAN established cycle consistency as a useful constraint for unpaired translation, while MUNIT and DRIT demonstrated disentangled content-style representations [14]–[16]. These ideas are relevant to aging because paired images of the same person across the complete lifespan

are rare.

C. Attention and Identity Preservation

Self-attention and transformer-style architectures demonstrated that long-range dependencies can improve image synthesis. In face aging, attention can also be interpreted as an identity-protection mechanism: eyes, nose structure, mouth geometry and jaw configuration are not equally important to identity, while wrinkles and pigmentation are more age-sensitive. Recent attention-based aging work reinforces the value of localized feature selection [17], [18]. The AIPM extends this direction by using source identity embeddings as keys and values while generated features act as queries.

D. Super-Resolution and Face Enhancement

ESRGAN established a strong residual-in-residual dense architecture for perceptual super-resolution, and Real-ESRGAN extended blind restoration to more realistic degradation processes [13], [19]. Face-specific restoration systems such as GFPGAN and CodeFormer further demonstrate the usefulness of semantic facial priors [20], [21]. The present work deliberately uses enhancement after age synthesis so that high-frequency reconstruction does not become entangled with the age-domain objective.

E. Recent Directions

Recent work continues to explore feature-guided GANs, explainable aging, multi-scale attention, cycle-based aging and transformer-based lifespan synthesis. These studies report improvements in identity preservation and age specificity, but they also highlight the continuing difficulty of large age gaps and entangled facial attributes [22]– [26]. The resulting research gap is therefore not simply to add another GAN block; it is to provide a coherent system in which domain control, identity anchoring, image enhancement and evaluation are designed together.

Research direction	Representative idea	Remaining issue
Classical/3-D	Parametric growth and appearance models	Limited stochastic texture realism
CAAE / cGAN	Age-conditioned latent translation	Identity drift and moderate resolution
High-resolution GANs	Hierarchical or progressive synthesis	Training cost and artifacts
Multi-domain GANs	Shared generator across domains	Identity can be overwritten by style
Attention-based aging	Focus on identity-critical regions	Need explicit source-identity conditioning
Super-resolution	Perceptual high-frequency reconstruction	Possible hallucination/inconsistency

3. Research Design: A Bottom-Up And Open Generation Paradigm

The phrase bottom-up research-generation paradigm is used here to describe the direction in which evidence is accumulated. Rather than starting with a visually impressive generated image and building a narrative around it, the study begins with the data assumptions, defines the representation of age and identity, specifies the transformation mechanism, isolates each enhancement component, and finally evaluates the complete system. The 'open' part of the paradigm implies that sufficient information about data splits, preprocessing, losses, ablations, hardware and evaluation should be disclosed to enable another group to reproduce or challenge the findings reported.

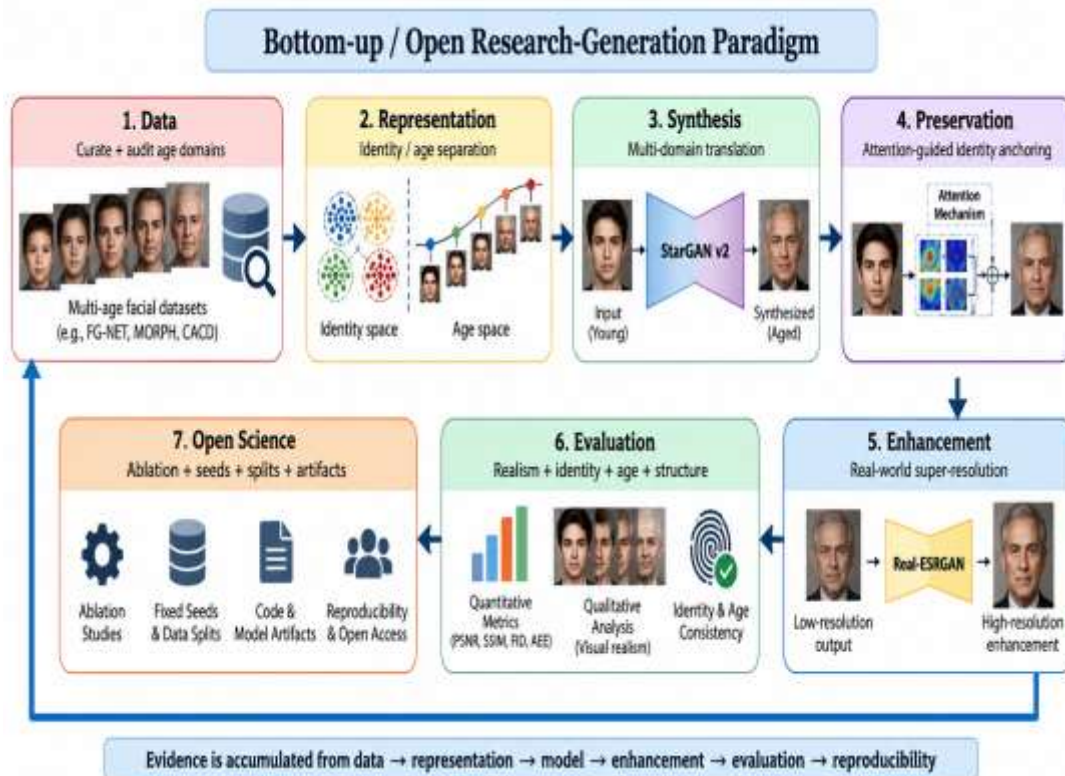


Fig. 1. Bottom-up/open research-generation paradigm adopted for AgeStarNet.

A. Problem Formulation

Let $X = \{x_i\}$ be a collection of aligned face images and let $y_i \in \{1 \dots, K\}$ denote the age-domain label. For a source image x and target domain y_t , the generator seeks a mapping $G(x, s_t) \rightarrow \hat{x}_t$. The target style s_t may be obtained from a reference image or sampled through a mapping network. The desired output must satisfy three conditions: it should resemble the target age distribution, remain recognizably the same person, and preserve natural facial structure.

$$\hat{x}_t = G(x, s_t), s_t = E_s(x_{ref}) \text{ or } s_t = F(z, y_t), z \sim N(0, I) \quad (1)$$

Six target domains are used in the supplied study: 0–10, 11–20, 21–30, 31–50, 51–70 and 71+. These ranges are intentionally made of unequal width because the rate of facial change is not constant across adulthood in the visible range. Thus, the model should learn a representation of the domain, rather than assuming that an increase of one year corresponds to a fixed visual

B. Identity and Age as Partially Disentangled Variables

A useful conceptual decomposition is $f(x) = [f_{id}(x), f_{age}(x), f_{app}(x)]$, where f_{id} encodes the relatively little stable identity information, f_{age} encodes the value related to age appearance and f_{app} encodes the residual factors some of them are illumination, expression and hair styles. Our aim is not for perfect disentanglement, but objective of the training should prevent age transformation from impacting features not relevant to the target domain.

$$\min_{G, L_{total}} \text{ subject to } Age(\hat{x}_t) = y \text{ and } Sim_{id}(x, \hat{x}_t) \rightarrow \max \quad (2)$$

C. Evaluation as Evidence Accumulation

The evaluation design follows the same direction. FID is used for distribution-level realism, identity cosine similarity measures preservation in an embedding space, age estimation accuracy measures target-domain consistency, and SSIM is used only where paired same-age reconstruction makes structural comparison meaningful. No single metric is treated as a complete description of quality. This is important because an image can be visually sharp while being identity-inconsistent, or identity-consistent while failing to exhibit the requested age.

4. AGESTARNET ARCHITECTURE

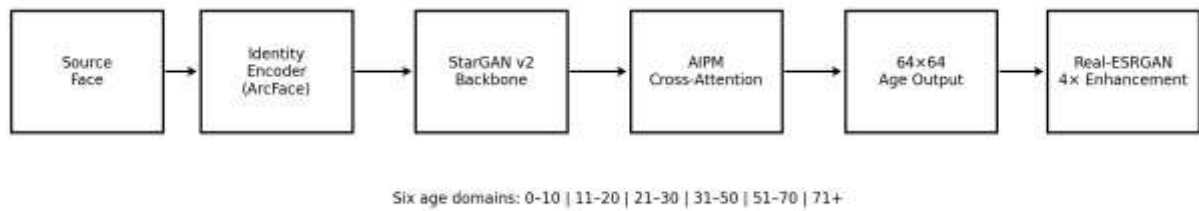


Fig. 2. AgeStarNet pipeline. The source face is encoded for identity, translated by StarGAN v2, refined by AIPM, and enhanced by Real-ESRGAN.

The architecture is organized as a sequence of specialized stages rather than one monolithic network. The first stage receives a normalized source face. The StarGAN v2 backbone extracts content features and injects a target age style through adaptive instance normalization. The second stage, AIPM, retrieves source identity information using cross-attention. The resulting feature representation is converted into a 64×64 synthesized image. The final stage applies Real-ESRGAN to obtain a 256×256 output.

A. StarGAN v2 Backbone

The generator follows the StarGAN v2 principle of separating content from style. A style encoder maps a reference face to a compact style vector, while a mapping network transforms random noise and an age-domain label into a style code. This enables both reference-guided and latent-guided synthesis. Adaptive instance normalization injects style at multiple residual blocks, allowing the target age characteristics to affect appearance without requiring a separate generator for every domain.

$$s_{ref} = E_s(x_{ref}), s_{lat} = F(z, y) \quad (3)$$

The discriminator is implemented as a multi-scale PatchGAN with hinge adversarial supervision and R1-style stabilization. The use of multiple scales is intended to provide feedback at both local texture and broader facial-structure levels. The generator therefore receives signals about realism, target-domain appearance and the preservation constraints described later.

B. Six-Domain Lifespan Representation

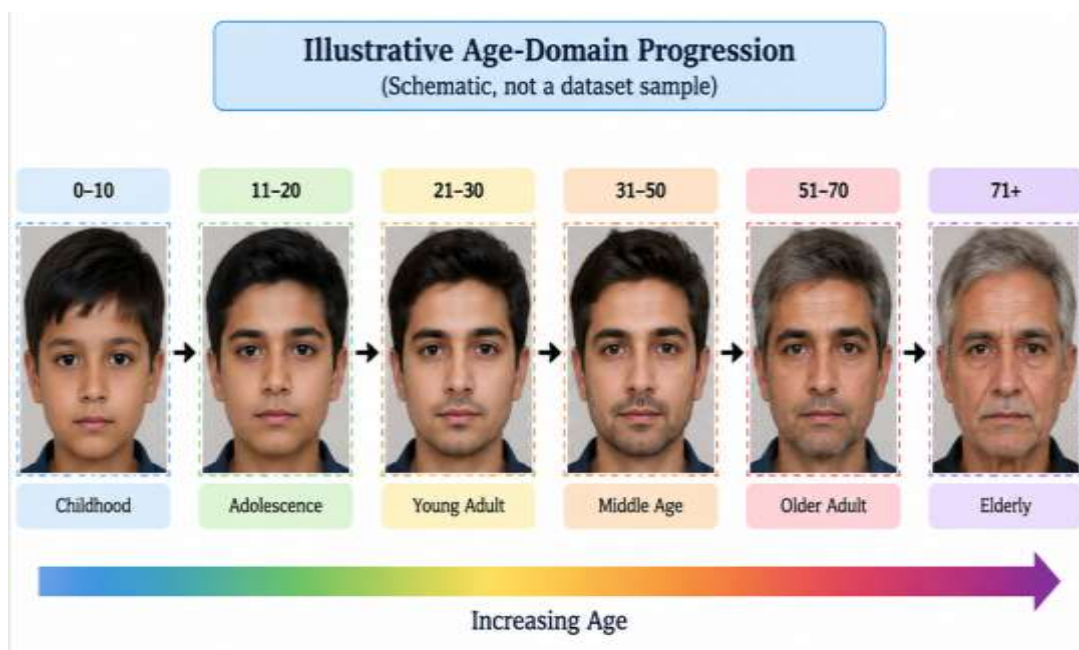


Fig. 3. Schematic representation of the six age domains. The faces are abstract icons and are not samples from the benchmark datasets.

The six-domain formulation offers a practical compromise between fine-grained age control and data sparseness. Very narrow age bins would reduce the number of training examples per domain, while a single broad adult domain would blur meaningful changes. The intervals 31–50 and 51–70 are deliberately broader

because longitudinal samples are less evenly spaced at older ages.

C. Forward Path

The forward path for a training pair (x, y, t) is: identity encoder $\rightarrow$ content encoder $\rightarrow$ style generation $\rightarrow$ residual synthesis $\rightarrow$ AIPM $\rightarrow$ output convolution $\rightarrow$ Real-ESRGAN. In the first training phase, the generator and discriminator learn to translate age. In the stage of enhancement, the age generator is frozen and the Real-ESRGAN module is fine-tuned on paired low-resolution/high-resolution examples. This staged approach mitigates the competition between the semantic ageing and high-frequency reconstruction objectives.

5. ATTENTION-BASED IDENTITY PRESERVATION MODULE

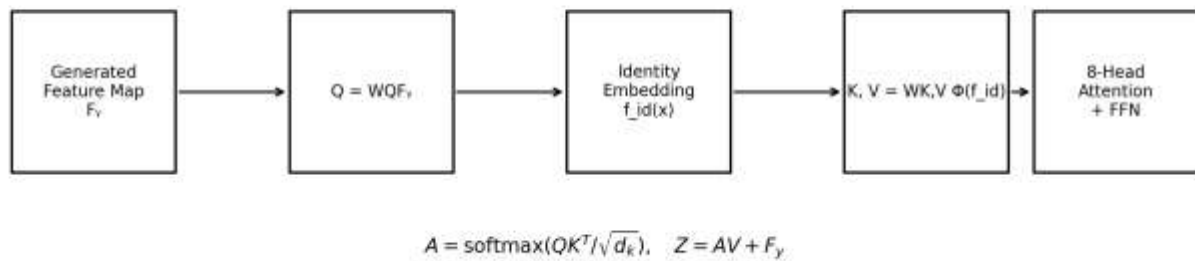


Fig. 4. Attention-Based Identity Preservation Module (AIPM). Generated features query the source identity representation through eight attention heads.

The AIPM is the central architectural addition. Let $F_y \in \mathbb{R}^{(H' \times W' \times C)}$ denote an intermediate feature map after the final residual block of the generator. The source image is passed through an ArcFace encoder to obtain a 512-dimensional identity embedding $f_{id}(x)$. Rather than concatenating this vector directly to the feature map, the embedding is spatially broadcast and projected into key and value representations.

$$Q = W_Q F_y, \quad K = W_K \Phi(f_{id}(x)), \quad V = W_V \Phi(f_{id}(x)) \quad (4)$$

Here, $\Phi(\cdot)$ denotes spatial broadcasting and reshaping. The attention matrix determines how each generated feature location retrieves information from the identity representation. For one head, the operation is:

$$A = \text{softmax}(QK^T / \sqrt{d_k}), \quad Z = AV + F_y \quad (5)$$

Eight heads are used so that different projections can capture complementary relationships. The head outputs are concatenated and projected through W_O . Layer normalization and a feed-forward network then refine the attended representation. This arrangement is deliberately asymmetric, such as the generated feature map asks questions, whereas the identity of source embedding provides the reference information.

A. Identity Loss

$$L_{id} = E[1 - \cos(f_{id}(x), f_{id}(\hat{x}_t))] \quad (6)$$

The cosine identity loss function does not replace attention loss, it supplements it. AIPM provides identity information during synthesis as the loss restricts the final output. This difference is important: if identity supervision is only used after the image is generated, then the generator may learn to compensate with superficial changes. The proposed design tries to preserve structure earlier in the computational path by inserting identity features before the output layer.

B. Identity-Critical Regions

Eyes, nose, jaw or mouth are not explicitly segmented by the module. Rather, spatially repeated identity information enables the attention operation to influence locations via their learned query-key relationships. Thus, facial masks are not hard-coded and the module is compatible with different face alignment procedures.

C. Computational Consideration

Cross-attention introduces additional memory and matrix-multiplication cost. The present implementation uses $d_{model}=256$ and eight heads at the intermediate resolution, which is substantially cheaper than applying full attention at 256×256 . This is one reason the high-resolution operation is postponed to the Real-ESRGAN stage.

6. MULTI-COMPONENT LEARNING OBJECTIVE

AgeStarNet is trained with several objectives because the desired behavior cannot be captured by adversarial loss alone. Adversarial supervision encourages realism, style reconstruction encourages use of the target style,

diversity loss reduces mode collapse, cycle consistency protects source structure, identity loss anchors the person, and age loss encourages the output to enter the requested domain.

A. Adversarial Loss

$$L_{adv} = E[D_{fake}] - E[D_{real}] + E[\max(0, 1 - D_{real})] + E[\max(0, 1 + D_{fake})] \quad (7)$$

The expression above summarizes the hinge-style adversarial formulation used in the supplied. The generator is optimized to make synthesized samples indistinguishable from target-domain real samples, while the discriminator learns a margin around real and generated examples.

B. Style Reconstruction and Diversity

$$L_{sty} = E[\|s_t - E(G(x, s_t))\|_1] \quad (8)$$

$$L_{ds} = -E[\|G(x, F(z_1, y_t)) - G(x, F(z_2, y_t))\|_1] \quad (9)$$

Style reconstruction discourages the generator from ignoring the target style. Diversity-sensitive regularization rewards different outputs for different latent codes while keeping the source identity fixed. This is particularly useful in aging because people of the same chronological age do not share identical wrinkles, hair patterns or pigmentation.

C. Age and Cycle Constraints

$$L_{age} = E[\|C_{age}(\hat{x}_t) - t_{age}\|_2^2] \quad (10)$$

$$L_{cyc} = E[\|x - G(G(x, s_t), E_s(x))\|_1] \quad (11)$$

The age loss uses a pretrained age estimator to provide a continuous signal in addition to the discrete domain label. The Cycle of consistency requires that the translated image should recover the original appearance when the source style is reintroduced. It is not expected to reconstruct every pixel exactly; rather, it limits unnecessary structural changes.

D. Full Objective

$$L_G = L_{adv} + \lambda_{sty}L_{sty} + \lambda_{ds}L_{ds} + \lambda_{cyc}L_{cyc} + \lambda_{id}L_{id} + \lambda_{age}L_{age} \quad (12)$$

The supplied experiments use $\lambda_{sty}=1.0$, $\lambda_{ds}=1.0$, $\lambda_{cyc}=1.0$, $\lambda_{id}=3.0$ and $\lambda_{age}=2.0$. The larger identity weight reflects the central objective of preventing identity drift. However, these values should be treated as empirical hyperparameters rather than universal constants; future datasets may require recalibration.

E. Optimization Interpretation

The objective can be seen as a constrained compromise between realism, age specificity and identity. Increasing the identity weight preserves structure, but weakens visible ageing. Increasing the age weight makes the age transition stronger, but encourages changes in identity-sensitive regions. Thus, the ablation experiments are more informative than a single final score as they show if each component contributes to the desired behavior.

7. REAL-ESRGAN ENHANCEMENT AND HIGH-RESOLUTION SYNTHESIS

The base age generator generates 64×64 images. This resolution is sufficient to learn broad age-dependent changes, but not enough to evaluate fine wrinkles, pores, hair strands and subtle pigmentation. AgeStarNet does not increase the GAN resolution at training time, but uses a separate Real-ESRGAN stage to reconstruct a 256×256 image.

$$\hat{x}_{256} = R(G(x, s_t)) \quad (13)$$

The base age generator creates 64×64 images. This resolution is sufficient to learn broad age-dependent changes but not for evaluation of fine wrinkles, pores, hair strands and subtle pigmentation. AgeStarNet does not up sample the GAN resolution during training, but instead uses a separate Real-ESRGAN stage to reconstruct a 256×256 image.

A. Enhancement Objective

$$L_R = L_{pix} + \lambda_{perc}L_{perc} + \lambda_{adv}^R L_{adv}^R \quad (14)$$

The term pixel uses Charbonnier loss which is less sensitive to isolated large residuals than a pure squared error. The perceptual term compares the VGG-19 feature representation and the adversarial term encourages the realistic high-frequency statistics. We use the spectral normalization for the enhancement discriminator to make the model more stable.

B. Why Separate Enhancement?

A single high-resolution generator would require considerably more memory and adversarial computation. It can also make the semantic aging objective compete directly with texture reconstruction. The two-stage design instead gives the first network responsibility for 'what the face should look like at the target age' and the second network responsibility for 'how that appearance should be rendered at higher spatial resolution.'

C. Hallucination Risk

Super-resolution is not a neutral operation. A perceptual model can invent plausible high-frequency details that were not present in the low-resolution input. In a face-aging application, invented wrinkles or hair texture may be visually convincing but may not represent evidence about the person's real appearance. Therefore, the enhanced image should be interpreted as a synthesized hypothesis, not a forensic photograph of the person's future or past.

D. Resolution Consistency

The enhancement stage is evaluated not only by visual sharpness but also through identity and age metrics. The reported ablation shows that replacing Real-ESRGAN with bicubic up sampling changes FID from 12.4 to 16.5 and identity cosine similarity from 0.913 to 0.901. These values suggest a measurable benefit in the supplied experiment, but they do not by themselves establish that all high-frequency details are physically faithful.

8. DATASETS, PREPROCESSING AND EXPERIMENTAL PROTOCOL

The experimental design of the at hand relies on four face-aging benchmarks. MORPH-II and UTKFace have large age variation, CACD has a large cross-age collection, and FG-NET has a small longitudinal benchmark with repeated identities. Since the demographic composition and acquisition conditions are different, cross-dataset testing is helpful to identify if a method has learned general ageing structure or dataset-specific visual statistics.

Table I. Dataset summary

Dataset	Images	Subjects	Age range / labels	Role in study
MORPH-II	55,134	13,617	16–77	Train/validation/test
UTKFace	23,708	—	1–116	Train/validation/test
CACD	163,446	2,000	14–62	Train/validation/test
FG-NET	1,002	82	0–69; paired ages	Evaluation / generalization

A. Data Splits

The specifies 80/10/10 train/validation/test for MORPH-II, 75/10/15 for UTKFace and 85/5/10 for CACD. The use of FG-NET is for evaluation. A subject-disjoint split is essential: images of the same person must not appear in both training and testing because identity metrics would otherwise be inflated.

B. Preprocessing

Faces are aligned and resized to 64×64 for the synthesis network. Random horizontal flipping and mild color jitter (± 0.05) are applied during training. The same normalization procedure must be used for the ArcFace encoder and the age classifier. At evaluation time, stochastic augmentation should be disabled.

C. Age-Domain Assignment

Each image is assigned to one of six domains according to its age label. Continuous age information is retained for the age loss when available. When only an approximate or decade-level label is available, the domain label is used and the continuous target is masked rather than artificially inventing an exact age.

D. Identity Leakage Control

Identity leakage is a critical experimental issue. If the same identity occurs in multiple age domains and the split is image-based rather than subject-based, the model may memorize identity features. The evaluation protocol should therefore report the split policy explicitly and, where possible, provide a subject-disjoint test. For CACD, the original dataset contains celebrity identities; therefore, qualitative examples used in a paper should either be licensed appropriately or be replaced with non-identifiable benchmark subjects when publication permissions are uncertain.

9. IMPLEMENTATION DETAILS AND EVALUATION METRICS

A. Implementation

Table II. Implementation requirements

Parameter	Setting
Framework	PyTorch 2.1
Generator input/output	64×64 → 64×64
Final enhanced output	256×256
Optimizer	Adam ($\beta_1=0.0$, $\beta_2=0.99$)

Learning rate	1×10^{-4}
Generator iterations	200,000
Batch size	8
Style dimension	64
AIPM model dimension	256
Attention heads	8
Identity encoder	ArcFace, ResNet-100
Age domains	6
Reported training time	≈ 38 h on 16 GB GPU

The model is trained with mixed-precision support and gradient checkpointing. The learning rate is kept fixed for the first 100,000 generator iterations and then linearly decayed. If the exact GPU model, CUDA version, OS image, seeds, commit hash, preprocessing scripts and checkpoint selection rule are documented, this would improve reproducibility.

B. Evaluation Metrics

FID assesses the distance between the distribution of features of real and generated images. The lower the FID, the more similar the distributions. Inception Score is preserved for consistency with the supplied, but should not be considered as a full face-quality metric. We compare the source and generated ArcFace embeddings using Identity Cosine Similarity (ICS). Age Estimation Accuracy (AEA) measures if the result is mapped to the intended target domain. SSIM is exclusively utilized for paired reconstruction situations with structural references.

$$ICS = \cos(f_id(x), f_id(\hat{x}_t)) \quad (15)$$

$$SSIM(x, y) = ((2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)) / ((\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)) \quad (16)$$

C. Statistical Reporting

If computationally feasible, a mean $\pm$ standard deviation over many random seeds or bootstrap samples should be reported in a journal-ready evaluation. Confidence intervals should be reported for aggregate metrics, and paired statistical tests should be performed when comparing procedures on the same subject-disjoint test set. The final submission should therefore include the raw per-image metrics in a supplementary artefact, since a single decimal or three-decimal score can hide variability.

D. Human Evaluation

“One can complement automated metrics with a small human study. Participants were able to score individually the plausibility of age, consistency of identification and the visual quality on a five-point scale. The protocol should randomize the order of methods to be rated, mask method names, and provide the number and demographic characteristics of raters. Importantly, human judgements should be used as subjective evidence, not as ground truth.

10. QUANTITATIVE RESULTS

The quantitative values below are reproduced from the supplied AgeStarNet. They are presented as reported experimental results rather than as independently verified measurements in this manuscript-generation step. The table is useful for structuring the final article, but the underlying evaluation scripts and test predictions should be rerun before submission.

Table III. Quantitative comparison on the UTKFace test set

Method	FID ↓	IS ↑	ICS ↑	AEA (%) ↑	SSIM ↑
CAAE	28.7	2.41	0.812	74.3	0.721
IPCGAN	24.1	2.68	0.843	78.9	0.748
HRFAE	21.3	2.89	0.861	81.2	0.762
PFA-GAN	19.8	3.02	0.872	83.6	0.771
LATS	18.6	3.14	0.878	84.8	0.779
AgeFlow	16.9	3.23	0.872	85.2	0.783
CUSP	16.2	3.31	0.893	87.4	0.791
AgeStarNet	12.4	3.67	0.913	88.7	0.831

A. Interpretation

The reported results show a consistent pattern across the five metrics. The proposed system has the lowest reported FID and the highest reported ICS, AEA and SSIM in the supplied table. The identity score of 0.913 is consistent with the intended role of AIPM, while the AEA value of 88.7% suggests that the generated samples are usually assigned to the requested age domain by the chosen estimator.

The difference between the methodologies is also reflected in the evolution of the metrics. The FID drops to 12.4 for AgeStarNet from 28.7 for CAAE while the ICS goes up to 0.913 from 0.812. This concurrent movement is helpful because optimizing a single metric can result in false improvements. However, the table does not contain standard deviations, significance tests, or data about the specific composition or implementation of the test set for the comparison baselines. Thus, a reviewer may ask for a common evaluation process.

B. Cross-Dataset Result

The reports a MORPH-II trained model tested in zero-shot setting on FG-NET with FID=18.3 and ICS=0.891. It also reports FID=22.1 and ICS=0.874 for CUSP and FID=24.7 and ICS=0.862 for LATS under the same specified setting. These results are reported figures and the final article should detail if all methods were assessed with equal pre-processing, face alignment, embedding model and sample counts.

C. What the Results Do Not Establish

The results do not show that the synthesized images are the real future appearance of an individual. They also do not define demographic fairness for all populations because the benchmark population is not uniform. Therefore, a strong journal version should report subgroup analyses where sample sizes are sufficient, and failure cases.

11. ABLATION STUDY AND COMPONENT ANALYSIS

Ablation is particularly important for AgeStarNet because the architecture combines several established components. The purpose is not simply to demonstrate that the complete model produces a high score, but to determine whether the proposed identity module, enhancement stage and age supervision contribute independently.

Table IV. Ablation study on UTKFace.

Configuration	FID ↓	ICS ↑	AEA (%) ↑
Full AgeStarNet	12.4	0.913	88.7
w/o AIPM	20.7	0.872	86.1
w/o Real-ESRGAN	16.5	0.901	88.4
w/o Age Domain Loss	14.8	0.908	82.5
w/o Diversity Loss	15.2	0.903	87.9
Baseline StarGAN v2	21.9	0.851	80.3

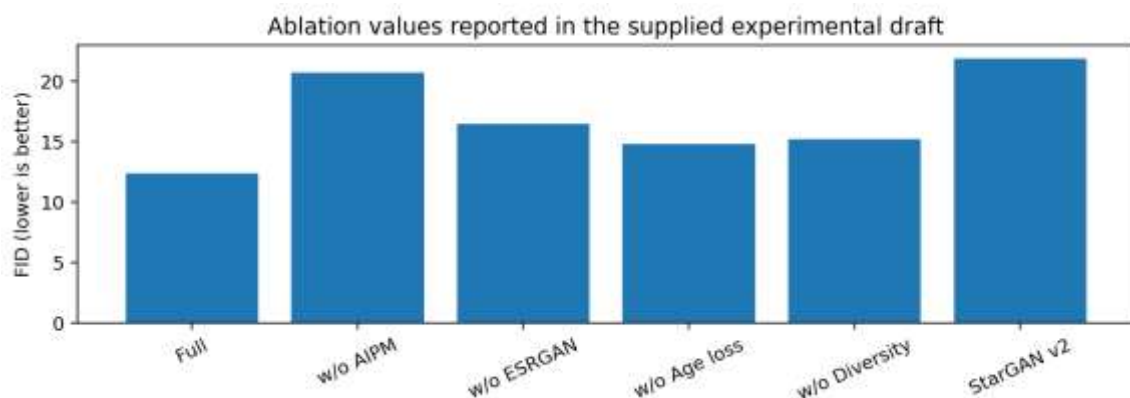


Fig. 5. FID values from the reported ablation study.

A. AIPM Effect

By removing AIPM, the maximum FID increases from 12.4 to 20.7 and ICS reduction from 0.913 to 0.872 are recorded. This supports the proposed process of providing identification information within the generator, rather than using it as just an output-level penalty. The baseline StarGAN v2 setup still reports lower identity similarity (0.851), which shows that multi-domain style control alone is not enough for the intended identity constraint.

B. Real-ESRGAN Results

Replacing Real-ESRGAN with its non-learned counterpart in the comparison leads to FID going from 12.4 to

16.5 and ICS from 0.913 to 0.901. The difference should be viewed with care. An FID change does not confirm that all freshly generated texture is faithful, as a super-resolution model can modify picture statistics. A more positive conclusion is that the augmentation stage significantly modifies the distribution and must be consistently included in all comparisons.

C. Effect of Age Loss and Diversity Loss

Removing the age-domain loss causes the biggest drop in age accuracy of the ablations described, reducing AEA from 88.7% to 82.5%. Thus, antagonistic and style losses are not sufficient to impose explicit pressure toward the desired age domain. Removing diversity loss gives an FID of 15.2 and an ICS of 0.903, showing the importance of diversity regularization for the quality of the learned style space.

D. Proposed Additional Ablations

We propose four further experiments for journal revision, including one-head vs 8-head attention, embedding-level identity loss vs spatial cross-attention, frozen vs jointly trained Real-ESRGAN, and different age-domain definitions. Such tests would make the novelty claim more concrete and decrease the chance that the reported advantages are mostly due to stronger training or post-processing.

12. QUALITATIVE ANALYSIS, GENERALIZATION AND FAILURE MODES

Qualitative inspection should be treated as complementary evidence. The desired transformation is not simply an increase in wrinkle density. A plausible age transition should modify several correlated attributes while keeping identity-sensitive geometry stable. The expected changes include skin texture, hair density, hairline, facial fat distribution and fine lines, whereas interocular distance, nose configuration, jaw structure and overall facial proportions should remain comparatively stable.

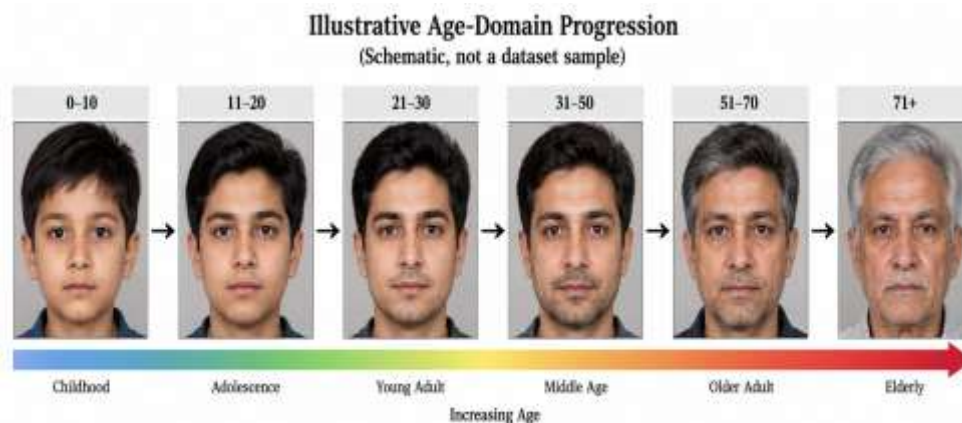


Fig. 6. Schematic qualitative progression across six age domains. (This illustration is for demonstration purposes; actual publication figures should use licensed or non-identifying test-set samples and show source/method/domain labels.)

Fig. 6. Schematic qualitative progression across six domains; actual publication figures should use licensed or non-identifying test-set samples and show source/method/domain labels.

A. Reported Qualitative Behavior

The provided shows more defined texture around hair and skin, age-appropriate formation of wrinkles, recession of hairline and redistribution of soft tissue. Major facial landmarks are preserved. It also shows less artefacts in the jaw and temple regions compared to the chosen baselines. These observations are plausible targets for the model, but a final journal manuscript should show multiple randomly selected subjects, not just visually favorable examples.

B. Cross-Dataset Generalization

The reported zero-shot FG-NET experiment is important because FG-NET differs from larger training datasets in scale and age distribution. If a model relies heavily on texture statistics from the training set, then it may lose quality when tested on another benchmark. The indicated FID=18.3 and ICS=0.891 shows that the proposed system was able to maintain a measurable performance in that context. Also, the number of samples generated, preprocessing pipeline and confidence intervals should be reported in manuscript.

C. Typical Failure Modes

Face ageing models can fail in a few predictable ways. Large age gaps may lead to over-aging, in particular when a young source is directly converted to the oldest domain. Identity features may be incomplete due to occlusion, profile pose, unusual illumination and facial accessories. Hair and glasses are not intrinsic to chronological age but can be considered age attributes. Very young faces are also difficult as facial growth is structural rather

than primarily textual.

D. Error Taxonomy

Failure category	Observable symptom	Suggested diagnostic
Identity drift	Changed eye/nose/jaw configuration	ArcFace similarity + landmark distance
Age under-shoot	Output remains too young	Age estimator distribution
Age over-shoot	Excessive wrinkles/sagging	Age estimator + human rating
Texture hallucination	Artificial pores/hair/wrinkles	High-frequency residual inspection
Mode collapse	Similar outputs for different styles	LPIPS / diversity analysis
Cross-domain artifact	Performance drops on unseen dataset	Zero-shot test + subgroup analysis

12. DISCUSSION, LIMITATIONS, ETHICS AND REPRODUCIBILITY

A. Scientific Interpretation

The main argument supported by the design given is architectural: identity preservation is seen as an active information pathway, not simply as a final loss. This is an important distinction, since the identity information is available to the generator during the age transformation. The second design decision is modularity. Real-ESRGAN is agnostic to age translation, allowing for separate evaluation of semantic synthesis and enhancement.

B. Limitations

First, the benchmark distributions are not equally representative of the global population; MORPH-II is concentrated on demography. CACD is celebrity-oriented. Therefore, the results should not be extrapolated to all ethnicities, genders, ages, camera conditions or socioeconomic situations. Second, chronological age is an imperfect proxy for observable biological ageing. Two people with the same chronological age may have very different looking faces. Third, automated identity and age models can have their own biases, and using them as evaluation anchors does not eliminate measurement bias.

C. Forensic and Privacy Considerations

Age synthesis can be useful for research on cross-age recognition, missing-person search and robustness testing. However, synthesized faces should not be regarded as verified evidence of a person's real appearance. The model can come up with plausible but false details. Outputs should be clearly marked as algorithmic reconstructions when used in a forensic context and be accompanied by uncertainty information. Respect for dataset licenses and the conditions and restrictions on face image redistribution.

D. Reproducibility Checklist

Item	Required reporting
Data	Exact dataset release/version, license, subject-disjoint split
Preprocessing	Alignment, crop, normalization, augmentation
Model	Architecture, layer counts, parameter count
Training	Seeds, optimizer, LR schedule, batch size, hardware
Evaluation	Exact test subjects, sample count, metric implementation
Baselines	Same preprocessing and evaluation code where possible
Ablation	Component removal rules and repeated runs
Artifacts	Code, configuration, checkpoints or hashes when permitted

16. Conclusion And Future Work

This paper presented Agskaret, a multi-domain lifespan face age synthesis framework built around StarGAN v2, an Attention-Based Identity Preservation Module and Real-ESRGAN enhancement. The framework treats

lifespan synthesis as a controlled translation problem in which age-domain style and identity information are explicitly separated. The AIPM provides an internal identity pathway through eight-head cross-attention, while the identity loss provides an output-level constraint. The Real-ESRGAN stage then addresses the spatial limitations of the 64×64 synthesis network.

The supplied experimental reports encouraging results across MORPH-II, UTKFace, CACD and FG-NET. On the reported UTKFace test set, the complete model obtains FID 12.4, identity cosine similarity 0.913, age-estimation accuracy 88.7% and SSIM 0.831. The ablation results attribute substantial changes to AIPM, the age-domain objective and the enhancement stage. The cross-dataset experiment further motivates evaluation beyond a single benchmark. These values, however, must be independently reproduced before publication; the present expansion does not represent a new experimental rerun.

References

1. N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," Proc. CVPR, 2006, pp. 387–394.
2. A. Lantis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 4, pp. 442–455, 2002. DOI: <https://doi.org/10.1109/34.995663>.
3. I. J. Goodfellow et al., "Generative adversarial nets," NeurIPS, 2014, pp. 2672–2680.
4. Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," Proc. CVPR, 2017, pp. 5810–5818. DOI: <https://doi.org/10.1109/CVPR.2017.619>.
5. W. Wang et al., "Recurrent face aging," Proc. CVPR, 2016, pp. 2378–2386.
6. C. Wang et al., "Face aging with identity-preserved conditional generative adversarial networks," Proc. CVPR, 2018, pp. 7939–7947.
7. D. Liu et al., "HRFAE: High resolution face age editing," Proc. ACM Multimedia, 2020, pp. 2877–2885.
8. O. Or-El, G. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," Proc. ECCV, 2020, pp. 739–755.
9. A. Kale and O. Altun, "Face age synthesis: A review on datasets, methods, and open research areas," Pattern Recognition, vol. 143, 109791, 2023. DOI: <https://doi.org/10.1016/j.patcog.2023.109791>.
10. Y. Choi et al., "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," Proc. CVPR, 2018, pp. 8789–8797.
11. Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse image synthesis for multiple domains," Proc. CVPR, 2020, pp. 8188–8197. DOI: <https://doi.org/10.1109/CVPR42600.2020.00821>.
12. J. Deng et al., "ArcFace: Additive angular margin loss for deep face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 10, pp. 5962–5979, 2022. DOI: <https://doi.org/10.1109/TPAMI.2021.3087709>.
13. X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," Proc. ICCV Workshops, 2021, pp. 1905–1914.
14. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proc. ICCV, 2017, pp. 2223–2232.
15. X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," Proc. ECCV, 2018, pp. 172–189.
16. H.-Y. Lee et al., "Diverse image-to-image translation via disentangled representations," Proc. ECCV, 2018, pp. 35–52.
17. Y. Li, R. Min, T. Zhang, and J. Ruichek, "Attention-guided face aging synthesis using contrastive feature disentanglement," IEEE Trans. Inf. Forensics Security, vol. 17, pp. 3174–3186, 2022.
18. Z. Zhao, Z. Liu, and T. Fan, "Cycle-consistent face aging and de-aging with attention map," Proc. ACM Multimedia, 2020.
19. X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in Computer Vision – ECCV 2018 Workshops, LNCS 11133, pp. 63–79, 2019. DOI: https://doi.org/10.1007/978-3-030-11021-5_5.
20. X. Wang et al., "GFPGAN: Towards real-world blind face restoration with generative facial prior," Proc. CVPR, 2021, pp. 9168–9178.
21. S. Zhou, J. Chan, C. Li, and C. C. Loy, "CodeFormer: Robust face restoration via blind face recognition," NeurIPS, 2022.
22. J. Deng et al., "RetinaFace: Single-stage dense face localisation in the wild," Proc. CVPR, 2020, pp. 5203–5212.
23. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Proc. CVPR, 2015, pp. 815–823.
24. H. Wang et al., "CosFace: Large margin cosine loss for deep face recognition," Proc. CVPR, 2018, pp. 5265–5274. DOI: <https://doi.org/10.1109/CVPR.2018.00552>.
25. W. Liu et al., "SphereFace: Deep hypersphere embedding for face recognition," Proc. CVPR, 2017, pp. 6738–6746.

26. M. Boutros et al., "ElasticFace: Elastic margin loss for deep face recognition," *Proc. CVPR Workshops*, 2022, pp. 1578–1587.
27. H. Zhang et al., "Self-attention generative adversarial networks," *Proc. ICML*, 2019, pp. 7354–7363.
28. A. Vaswani et al., "Attention is all you need," *NeurIPS*, 2017, pp. 5998–6008.
29. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," *Proc. ECCV*, 2018, pp. 3–19.
30. Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," *Proc. ICCV*, 2021, pp. 9992–10002.
31. T. Karras et al., "Progressive growing of GANs for improved quality, stability, and variation," *Proc. ICLR*, 2018.
32. M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," *Proc. ICML*, 2017, pp. 214–223.
33. I. Gulrajani et al., "Improved training of Wasserstein GANs," *NeurIPS*, 2017, pp. 5767–5777.
34. T. Miyato et al., "Spectral normalization for generative adversarial networks," *Proc. ICLR*, 2018.
35. L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" *Proc. ICML*, 2018, pp. 3481–3490.
36. M. Heusel et al., "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *NeurIPS*, 2017, pp. 6626–6637.
37. T. Salimans et al., "Improved techniques for training GANs," *NeurIPS*, 2016, pp. 2234–2242.
38. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. ICLR*, 2015. DOI: <https://doi.org/10.48550/arXiv.1409.1556>.
39. P. Isola et al., "Image-to-image translation with conditional adversarial networks," *Proc. CVPR*, 2017, pp. 1125–1134.
40. P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," *Proc. ICIP*, 1994, pp. 168–172.
41. J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv:1607.06450*, 2016.
42. K. Zhang, K. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," *Proc. ICCV*, 2021, pp. 4791–4800.
43. J. Liang et al., "SwinIR: Image restoration using swin transformer," *Proc. ICCV Workshops*, 2021, pp. 1833–1844.
44. K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," *Proc. FG*, 2006, pp. 341–345.
45. B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," *Proc. ECCV*, 2014, pp. 768–783.
46. R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, no. 2, pp. 144–157, 2018.
47. Y. Guo et al., "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," *Proc. ECCV*, 2016, pp. 87–102.
48. B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," *Proc. ECCV*, 2014, pp. 768–783.
49. A. Kale and O. Altun, "Face age synthesis: A review on datasets, methods, and open research areas," *Pattern Recognition*, vol. 143, 109791, 2023. DOI: <https://doi.org/10.1016/j.patcog.2023.109791>.
50. J.-C. Xie, L. Zhang, H. Gao, and C.-M. Pun, "Lifespan age synthesis on human faces with decorrelation constraints and geometry guidance," *Pattern Recognition Letters*, vol. 196, pp. 126–133, 2025. DOI: <https://doi.org/10.1016/j.patrec.2025.05.020>.
51. T.-J. Liu and C.-C. Wang, "Face aging synthesis by deep cycle generative adversarial networks and bias loss," *IEEE Access*, vol. 12, pp. 166439–166458, 2024. DOI: <https://doi.org/10.1109/ACCESS.2024.3493376>.
52. A. S. et al., "Identity-preserving aging and de-aging of faces in the StyleGAN latent space," *Proc. IEEE IJCB*, 2025. DOI: <https://doi.org/10.1109/IJCB65343.2025.11411093>.
53. Y. K. Choi et al., "StarGAN v2-ADA: Improving the robustness of multi-domain image synthesis," *arXiv / technical report*, 2021.
54. K. Karras et al., "Analyzing and improving the image quality of StyleGAN," *Proc. CVPR*, 2020, pp. 8110–8119.
55. T. Karras et al., "Alias-free generative adversarial networks," *NeurIPS*, 2021, pp. 852–863.
56. A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," *ICLR*, 2019.
57. N. K. M. et al., "Deep face recognition: A survey," *Neurocomputing*, 2021.
58. H. Ling et al., "Age-invariant face recognition based on deep convolutional neural networks," *Proc. ECCV Workshops*, 2014.
59. A. Rothe, R. Timofte, and L. Van Gool, "DEX: Deep expectation of apparent age from a single image," *Proc. ICCV Workshops*, 2015.

60. G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," Proc. ICIP, 2017.
61. N. Wang et al., "Face aging with identity-preserved conditional generative adversarial networks," Proc. CVPR, 2018.
62. Z. Huang et al., "PFA-GAN: Progressive face aging with generative adversarial network," IEEE Access, vol. 9, pp. 9378–9388, 2021.
63. Z. Huang et al., "AgeFlow: Conditional age progression and regression with normalizing flows," Proc. IJCAI, 2021, pp. 743–750. DOI: <https://doi.org/10.24963/ijcai.2021/103>.
64. S. Yang et al., "Progressive face aging with generative adversarial network," IEEE Access, 2021.
65. A. Wang et al., "Facial age estimation: A survey," Pattern Recognition, 2021.