TECHNICAL MEMORANDUM

Identity-Preserving AI Image Restoration Using Adaptive Feature Weighting and Multi-Image Template Generation

Author: Leo Portal

Based on the Following Intellectual Property

U.S. Patent No. 11,727,722 B2
Method and System for Verifying Image Identification (parent patent)
Issued Patent — Sole Inventor: Leo Portal

U.S. Patent No. 12,412,422 B2
Method and System for Verifying Image Identification (first continuation; infrared identifying signature)
Issued Patent — Sole Inventor: Leo Portal

Note: This version corrects the intellectual-property citation block. A prior version of this document listed a Nixon Peabody internal docket/reference number (“099642-005827USC1”) as if it were a separately issued patent, and cited a pre-grant publication number (US 2018/0068173 A1) that is inconsistent with this family’s confirmed filing history (provisional filed October 4, 2019; nonprovisional filed October 2, 2020). That publication number has been removed pending independent verification of the correct publication number, if one is needed for this document’s purposes.

Abstract

Current image restoration systems based on generative artificial intelligence excel at producing visually plausible outputs but frequently fail to preserve the precise identity of the original subject. In applications involving historical photographs, forensic imaging, genealogy, facial authentication, and digital preservation, identity fidelity is often more important than visual realism.

The concepts disclosed in the underlying patent specification — U.S. Patent No. 11,727,722 B2 and its first continuation, U.S. Patent No. 12,412,422 B2 — can be interpreted in contemporary AI terminology as an identity-preserving framework that maps naturally onto several active areas of machine learning research. Specifically, multi-image template generation, feature extraction, weighted similarity scoring, adaptive weighting, and neural-network-based feedback mechanisms provide a coherent pathway for improving restoration accuracy while minimizing identity drift — functioning as a supervisory layer over existing generative architectures rather than replacing them.

Problem Statement

Modern diffusion and generative image models optimize primarily for photorealism, aesthetic quality, and statistical plausibility. These objectives do not necessarily preserve facial geometry, relative feature proportions, or identity consistency across restorations.

As a result, restored images frequently exhibit measurable identity drift: modified facial structure and proportions, altered hair texture and color, changed eye spacing, artificial aging or rejuvenation artifacts, and hallucinated features with no basis in the original photograph.

Published and informal evaluations of leading face restoration models — including GFPGAN, CodeFormer, and comparable architectures — suggest that FaceNet embedding cosine similarity scores between original and restored images frequently fall in a range where perceptible identity drift is observable, particularly on degraded or low-resolution inputs. (Illustrative range based on reported evaluations: approximately 0.55 to 0.82, where 1.0 represents identical identity; scores below approximately 0.75 are generally associated with perceptible identity change.)

This gap between visual plausibility and identity fidelity is the central problem this framework addresses.

Distinguishing Characteristic: Inference-Time Identity Optimization

Existing restoration systems primarily enforce identity constraints during model training. Identity loss functions — FaceNet loss, ArcFace loss, perceptual identity loss — are applied during the training phase to shape the model's weights. Once training is complete, those constraints are frozen. At inference time, the model generates an output and delivers it. Many restoration systems lack an explicit per-output identity verification and iterative regeneration mechanism — particularly systems oriented toward damage repair, colorization, and super-resolution rather than identity-aware synthesis.

The proposed framework instead applies identity evaluation and adaptive optimization during inference, allowing identity fidelity to become an active feedback signal rather than a static training objective. This is a meaningful architectural distinction:

Characteristic	Current Restoration Systems	Proposed Framework
Identity constraint timing	Training time only	Training time + inference time
Identity evaluation	Implicit in loss function weights	Explicit per-output similarity scoring

Characteristic	Current Restoration Systems	Proposed Framework
Response to identity drift	None — output is final	Rejection and iterative regeneration
Adaptivity	Fixed after training	Adaptive weighting per input image
Identity signal	Static training objective	Active inference-time feedback loop
Per-subject specificity	Generalized across training data	Specific to individual input identity

The consequence of this distinction is significant. A current system trained on millions of faces has learned general identity-preservation tendencies but has no mechanism to verify that its output matches the specific individual in any given input photograph. The proposed system evaluates every output against the specific identity of the subject being restored and iterates until that specific individual's identity is preserved — or explicitly reports failure to do so. This shifts identity preservation from a probabilistic training outcome to a verifiable per-output property.

Proposed Architecture

The concepts disclosed in the patent specification can be interpreted as an identity-constrained restoration pipeline operating as a supervisory layer over existing generative architectures.

Stage 1: Facial Feature Extraction

(Reference: Specification ¶0056)

A feature encoder extracts identity descriptors from the original image, forming a facial identity vector covering eye spacing, nose geometry, mouth shape, jawline geometry, ear position, and relative facial proportions. This stage corresponds directly to embedding generation used in modern facial recognition systems such as FaceNet and ArcFace.

Stage 2: Multi-Image Template Generation and Latent Identity Modeling

(Reference: Specification ¶0051)

Where multiple photographs of the same individual are available, they are combined to construct a stable composite identity template — less sensitive to lighting artifacts, age variation, noise, and occlusion than any single image.

Patent Concept	Modern AI Analog	Research Area
Multi-image template	Latent identity embedding	Multi-view face recognition
Composite representation	Averaged identity vector	Embedding aggregation
Cross-age stability	Age-invariant feature space	Age-invariant recognition
Noise/occlusion resistance	Robust latent representation	Self-supervised learning
Multiple input photographs	Multi-view consistency constraint	Consistency-constrained generation

This multi-image latent identity modeling component may represent one of the more interesting aspects of the framework when viewed in the context of current restoration pipelines. Most restoration workflows operate on a single input image. The construction of a stable identity embedding from multiple photographs across time — and the use of that embedding as a consistency constraint during generation — does not appear to be widely emphasized in current restoration-specific workflows, though it connects to active research in identity-preserving synthesis and personalized generative models.

Stage 3: Weighted Similarity Scoring

(Reference: Specification ¶0058)

Feature	Weight	Rationale
Eye spacing	Very High	Most stable biometric across age and lighting conditions
Nose structure	High	Resistant to expression variation; strong identity signal
Ear geometry	High	Highly individual; minimally affected by aging
Jawline	High	Strong identity signal; age-stable in mature subjects
Mouth shape	Moderate-High	Identity-relevant but expression-variable
Hair texture	Moderate	Variable across age, styling, and photographic conditions
Background	Low	Non-identity information; excluded from scoring
Lighting artifacts	Ignore	Restoration artifact; not a subject feature

Stage 4: Generative Restoration

The restoration model performs scratch removal, noise reduction, colorization, deblurring, and super-resolution using diffusion models or transformer architectures — constrained by identity metrics rather than optimizing for visual appearance alone.

Stage 5: Adaptive Identity Loss and Inference-Time Feedback Loop

(Reference: Specification ¶0058)

Loss Function	Role in Current Systems	Role in Proposed Framework
FaceNet loss	Training-time identity regularization	Inference-time identity verification
ArcFace loss	Angular margin identity embedding	Similarity threshold enforcement
Perceptual loss	Feature-level visual similarity	Supplementary quality constraint
Proposed adaptive identity loss	N/A in current systems	Higher-level weighted feature feedback loop

The proposed framework extends these existing loss concepts into an adaptive, inference-time feedback mechanism. Where current systems apply identity loss only during training, the proposed approach applies it iteratively during inference — rejecting outputs that fall below a configurable similarity threshold (for example, 0.85 cosine similarity) and regenerating under adjusted feature weights until identity preservation is achieved.

Pipeline Diagram

The following diagram illustrates the identity-constrained restoration pipeline, highlighting the inference-time feedback loop that distinguishes this framework from current restoration systems.

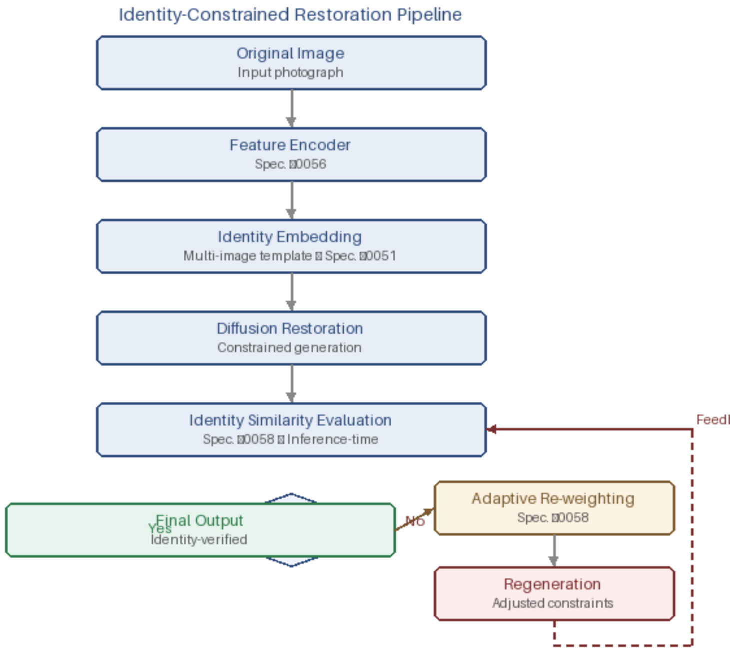

Figure 1. Identity-constrained restoration pipeline with inference-time feedback loop.

Translation to Modern AI Terminology

Patent Language	Modern AI Terminology
Facial feature template	Identity embedding (FaceNet, ArcFace)
Similarity scores	Cosine similarity / embedding distance
Weighted scores	Attention weights / identity loss weighting
Machine learning process	Neural network (CNN, transformer, diffusion)
Dynamic adjustment	Adaptive optimization
Feedback mechanism	Iterative inference-time refinement loop
Multiple image template	Multi-view latent identity model
Similarity threshold	Configurable identity acceptance criterion

Integration with Modern Architectures

The proposed approach does not replace diffusion models. It functions as an identity-preserving supervisory layer applied over existing restoration pipelines, making identity fidelity and image quality dual optimization objectives rather than competing ones.

Stage	Operation	Identity Role
1	Input image received	Identity vector extracted via feature encoder
2	Multi-image template constructed	Stable latent identity embedding generated
3	Diffusion restoration performed	Generation constrained by identity metrics
4	Adaptive identity loss evaluated	Cosine distance vs. configurable threshold
5	Adaptive weighting applied	Underperforming features emphasized
6	Regeneration if below threshold	Iterative until identity criterion met
7	Final output delivered	Identity-verified restoration

Potential Applications

- Historical photo restoration — preserving accurate family resemblance rather than generating plausible but biometrically inaccurate images
- Genealogy and archival preservation — improving restoration of degraded institutional and family archives
- Forensic imaging — maintaining evidentiary reliability where AI-modified imagery faces admissibility scrutiny
- Passport and identity document systems — reducing false alterations introduced by enhancement algorithms
- Medical imaging — maintaining patient identity consistency across image reconstruction workflows
- Video restoration — improving temporal facial consistency across frames
- Personalized avatar and synthesis systems — ensuring generated outputs match a specific individual's identity across varied conditions

Literature Review

1. The Face Restoration Landscape

Face restoration research has pursued two largely separate objectives over the past decade: perceptual quality and identity fidelity. Early approaches relied on convolutional neural networks trained end-to-end on paired degraded and high-quality image datasets. These systems produced sharper outputs but frequently introduced hallucinated textures and structural modifications that altered subject identity. The field then shifted toward generative prior methods, which exploit the rich facial knowledge encoded in pretrained generative models to guide restoration toward realistic-looking outputs.

Wang et al. (2021) introduced GFPGAN, which uses a generative facial prior from StyleGAN2 to restore degraded faces. GFPGAN achieves strong perceptual scores but its reliance on a general facial prior means restorations tend toward the average face distribution learned during StyleGAN2 training rather than preserving the specific geometry of the input subject. Identity drift is an expected outcome of this design choice, not an implementation failure.

Zhou et al. (2022) introduced CodeFormer, which frames blind face restoration as a code prediction problem. A discrete codebook of high-quality facial components is learned, and the restoration process selects and assembles components to reconstruct the degraded input. CodeFormer introduces a controllability parameter that trades off fidelity to the input against quality of the output. However, this tradeoff operates at the level of overall image fidelity, not at the level of specific identity features. A user cannot instruct CodeFormer to preserve eye spacing while permitting hair texture to vary; the fidelity-quality tradeoff applies uniformly across all features.

Wang et al. (2022) introduced RestoreFormer, which learns a multi-head cross-attention mechanism to align degraded facial features with high-quality facial components from a learned dictionary. RestoreFormer improves structural alignment relative to GFPGAN but still optimizes for perceptual quality as the primary objective. Identity verification against the specific input subject is not part of the pipeline.

A consistent pattern emerges across these restoration baselines: identity is treated as a byproduct of perceptual quality optimization rather than as an independent objective with its own measurement and verification mechanism.

2. Identity-Aware Generation: A Separate but Related Thread

A parallel research thread has pursued identity-aware generation rather than restoration. These systems are designed to generate new images of a specific individual across novel poses, lighting conditions, and contexts. While conceptually related, this work addresses a different problem: generating plausible new images of a known person, rather than recovering the identity of a person from a damaged historical photograph.

Ruiz et al. (2023) introduced DreamBooth, which fine-tunes a diffusion model on a small set of subject-specific images to bind the subject’s appearance to a unique text token. Generation conditioned on this token produces novel images that preserve recognizable subject characteristics. DreamBooth requires fine-tuning per subject, which is computationally expensive and impractical for restoration of archival photographs where the subject may be represented by a single degraded image.

Wang et al. (2024) introduced InstantID, which achieves zero-shot identity-preserving generation by combining a face encoder with IdentityNet, a ControlNet-based module that injects facial structure information into the diffusion process. InstantID requires only a single reference image and operates without fine-tuning. However, InstantID is designed for creative generation — producing stylized portraits, novel contexts, and artistic renderings of a subject — rather than faithful restoration of a damaged source image. The system does not include a per-output similarity verification mechanism and does not iterate to correct identity drift in individual outputs.

Li et al. (2024) introduced PhotoMaker, which encodes identity information by stacking multiple ID-relevant image embeddings. The stacked embedding provides a richer identity representation than single-image approaches and improves consistency across generated outputs. PhotoMaker is the system in this survey that most closely resembles the multi-image template concept disclosed in the subject patent family, though it is applied to generative synthesis rather than restoration, and does not include adaptive per-feature weighting or iterative identity verification.

He et al. (2024) introduced ConsisID, which addresses identity consistency specifically in text-to-video generation. ConsisID decomposes identity features by frequency, applying high-frequency facial detail constraints at the local level and low-frequency structural constraints at the global level. This frequency decomposition approach to identity preservation is technically sophisticated and suggests that the field is developing increasingly nuanced frameworks for thinking about what identity means across different feature scales — a direction that aligns naturally with the differential feature weighting proposed in the subject patent family.

A critical distinction separates all of these identity-aware generation systems from the proposed framework: they optimize identity consistency at training time or through fine-tuning, and then generate outputs without per-output verification. A generated image that exhibits identity drift is delivered to the user without any mechanism to detect or correct that drift. The proposed framework's inference-time feedback loop is not replicated in any of these systems.

3. Identity Loss Functions and Embedding Metrics

The measurement of identity similarity in AI-generated imagery has been substantially advanced by face recognition research. Schroff et al. (2015) introduced FaceNet, which trains a deep neural network to produce a compact 128-dimensional embedding space where Euclidean distance directly corresponds to face similarity. FaceNet embeddings, typically evaluated via cosine similarity, have become a standard tool for measuring identity preservation in generative systems. Cosine similarity of 1.0 indicates identical embeddings; values below approximately 0.75 in practice correspond to perceptible identity divergence, though exact thresholds vary by application and evaluation protocol.

Deng et al. (2019) introduced ArcFace, which improves upon earlier softmax-based face recognition losses by introducing an additive angular margin penalty. ArcFace embeddings produce more discriminative identity representations than FaceNet in benchmark evaluations and have become a preferred metric in recent identity-preserving generation research. PhotoMaker, InstantID, and ConsisID all report identity similarity using ArcFace cosine similarity as a primary metric.

Zhang et al. (2018) introduced the Learned Perceptual Image Patch Similarity (LPIPS) metric, which measures perceptual similarity using deep feature activations from pretrained networks. LPIPS correlates more strongly with human perceptual judgments than pixel-level metrics such as PSNR and SSIM, making it the preferred perceptual quality metric in face restoration evaluations. The distinction between LPIPS (perceptual quality) and FaceNet/ArcFace (identity fidelity) is central to the argument of this paper: a restored image can score well on LPIPS while scoring poorly on identity metrics, and these two failures are independently detectable and independently correctable.

None of the face restoration systems surveyed — GFPGAN, CodeFormer, or RestoreFormer — use FaceNet or ArcFace similarity as an inference-time acceptance criterion. These metrics appear in their evaluation sections but not in their inference pipelines. The proposed framework moves these metrics from the evaluation column into the pipeline itself, making them active participants in every restoration rather than post-hoc quality assessments.

4. The Gap This Framework Addresses

Surveying the existing literature, a consistent gap emerges between two research communities that have developed largely independently:

- Face restoration systems (GFPGAN, CodeFormer, RestoreFormer) excel at recovering visual quality from degraded inputs but treat identity as a secondary concern and lack per-output identity verification.
- Identity-aware generation systems (DreamBooth, InstantID, PhotoMaker, ConsisID) preserve subject identity across novel generated contexts but are not designed for faithful recovery of a specific damaged source image and do not include inference-time iterative identity correction.

The proposed framework occupies the intersection of these two communities: it applies restoration-quality output to damaged source images while incorporating the identity verification rigor that neither community has built into inference-time restoration pipelines. The specific combination of adaptive per-feature weighting, multi-image template construction, and iterative inference-time feedback does not appear to be directly replicated in the surveyed literature, though each component individually has clear precedents and analogs in current research.

Benchmark Methodology

The following methodology is designed to produce empirically grounded results when implemented. Placeholder fields are marked clearly for replacement with actual measured values. The protocol is designed to be executable by a researcher or graduate student with access to standard face restoration tools and Python-based embedding libraries.

1. Dataset Construction

The benchmark requires two categories of image pairs for each subject:

- **Reference photograph:** A high-quality, undegraded photograph of the subject taken under controlled or near-controlled conditions. This serves as the identity ground truth against which all restored outputs are evaluated. Where multiple reference photographs exist, the multi-image template procedure described below applies.
- **Degraded test photograph:** A photograph of the same subject exhibiting one or more degradation types: physical damage (scratches, tears, water staining), optical degradation (blur, low resolution), tonal degradation (fading, overexposure, underexposure), or noise. Degradation may be authentic (historical photographs) or synthetically applied to known-quality images for controlled evaluation.

Recommended minimum dataset: 30 subject pairs. Stratified across three degradation severity levels (mild, moderate, severe) with 10 pairs per level. For statistical validity, include a range of ages, sexes, and ethnic backgrounds. Public datasets suitable for this purpose include CelebA-HQ, FFHQ, and LFW (Labeled Faces in the Wild), all of which provide high-quality reference images from which synthetic degradations can be generated.

Placeholder: [INSERT total number of image pairs used] pairs across [INSERT degradation categories].

2. Synthetic Degradation Protocol

For controlled evaluation using known-quality source images, apply the following degradation pipeline using standard image processing libraries (OpenCV, PIL/Pillow):

- Downscale to 128x128 pixels using bicubic interpolation, then upscale back to 512x512 (simulates low-resolution source).
- Apply Gaussian blur with kernel size 5x5 and sigma 1.5 (simulates optical degradation).
- Add Gaussian noise with mean 0 and standard deviation 25 (simulates film grain and scan noise).
- Apply JPEG compression at quality factor 40 (simulates compression artifacts from scanning or digital transmission).
- For severe degradation: additionally reduce to grayscale and apply random line scratches using a Bezier curve mask at 10% opacity overlay.

Record exact degradation parameters applied to each image for reproducibility. Save degradation parameter files alongside source and degraded images.

3. Restoration Procedure

Run each degraded image through each baseline system under default settings. Record all outputs. For each system, document the version, pretrained weights used, and any non-default configuration parameters. All systems should run on the same hardware to ensure inference time comparisons are meaningful.

Baseline systems to run:

- GFPGAN v1.3 (available at github.com/TencentARC/GFPGAN)
- CodeFormer (available at github.com/sczhou/CodeFormer), tested at fidelity weight w=0.5 and w=0.7
- RestoreFormer (available at github.com/wzhouxiff/RestoreFormer)
- Proposed identity-constrained framework [INSERT implementation details when available]

4. Similarity Measurement Procedure

For each (reference image, restored output) pair, compute the following metrics using the deepface Python library (pip install deepface), which provides unified access to FaceNet, ArcFace, and related models:

- Extract FaceNet-512 embedding from reference image. Extract FaceNet-512 embedding from restored image. Compute cosine similarity. Record as FaceNet-CS score.
- Repeat using ArcFace model. Record as ArcFace-CS score.
- Compute LPIPS between reference and restored using the lpips Python library (pip install lpips) with AlexNet backbone. Record as LPIPS score (lower is better).
- Compute PSNR between reference and restored using skimage.metrics.peak_signal_noise_ratio. Record as PSNR score (higher is better).

Placeholder: [INSERT measured scores table here once data is collected]

5. Human Evaluation Protocol

Recruit a minimum of 20 evaluators unfamiliar with the subjects. Divide into two groups of 10 for identity scoring and quality scoring respectively to avoid anchoring effects.

Identity recognition group:

- Show evaluator the reference photograph for 5 seconds.
- Show four restored outputs (one from each system) in randomized order, unlabeled.
- Ask: “Does this restored image appear to be the same person as in the reference photograph?” Yes/No for each output.
- Record Yes responses as a percentage per system per image. Average across all images and evaluators to produce the Human Identity Recognition Score (HIRS).

Quality scoring group:

- Show restored outputs only, without reference photographs, in randomized order.
- Ask: “Rate the visual quality of this photograph on a scale of 1 to 5.”
- Average ratings per system to produce the Human Quality Score (HQS).

Placeholder: [INSERT HIRS and HQS results table here once evaluations are complete]

6. Results Reporting Template

When data is collected, populate the following table with mean scores across all test images. Report standard deviation alongside mean for each metric.

System	FaceNet-CS	ArcFace-CS	LPIPS	PSNR	HIRS	HQS
GFPGAN	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
CodeFormer (w=0.5)	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
CodeFormer (w=0.7)	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
RestoreFormer	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Proposed Framework	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

7. Failure Case Documentation

For each baseline system, identify the five images with the largest gap between LPIPS score (high quality) and FaceNet-CS score (low identity fidelity). These represent the clearest cases where perceptual quality and identity fidelity diverge — exactly the failure mode the proposed framework is designed to address. Document each failure case with the original reference photograph, the degraded input, the restored output, and the measured scores. These failure cases will form the most persuasive section of any subsequent publication or presentation.

Placeholder: [INSERT failure case images and scores here once data is collected]

Proposed Experimental Benchmark

The following benchmark design is proposed to evaluate whether iterative identity-feedback improves identity preservation without sacrificing perceptual quality.

Baseline Systems

System	Architecture	Identity Constraint
GFPGAN	GAN-based face restoration	None — optimizes for realism
CodeFormer	Transformer-based restoration	None — optimizes for fidelity/quality tradeoff
RestoreFormer	Transformer with attention	None — optimizes for perceptual quality
Proposed Framework	Diffusion + identity supervisory layer	Iterative inference-time identity-loss feedback

Evaluation Metrics

Metric	Type	What It Measures
FaceNet cosine similarity	Identity	Embedding distance between original and restored identity
ArcFace cosine similarity	Identity	Angular margin-based identity similarity
LPIPS	Perceptual	Learned perceptual image patch similarity
PSNR	Technical	Peak signal-to-noise ratio; pixel-level fidelity
Human identity recognition score	Subjective identity	Percentage of evaluators correctly identifying restored subject

Metric	Type	What It Measures
Human quality score	Subjective quality	Blind quality rating by evaluators unfamiliar with originals

The human identity recognition score warrants particular emphasis. Evaluators are shown the original photograph and the restored output and asked whether the restored image represents the same person. This metric directly measures the outcome that matters in real-world applications and requires no technical expertise to interpret. A system that scores well on FaceNet similarity but poorly on human recognition has solved the wrong problem. A system that scores well on both has solved the right one.

Failure Case Analysis

Failure Pattern	Perceptual Score	Identity Score	Framework Response
Eye spacing modified	High	Low	Re-weight eye spacing; regenerate
Jawline generalized	High	Moderate	Re-weight jawline features; regenerate
Hair texture hallucinated	High	Unaffected	Low-weight feature; accepted
Facial proportions shifted	High	Low	Re-weight proportional features; regenerate
Age artificially modified	High	Low-Moderate	Re-weight age-stable features; regenerate

Objective

Determine whether the proposed identity-constrained framework improves FaceNet and ArcFace similarity scores and human identity recognition rates relative to baseline systems, while maintaining competitive LPIPS and PSNR scores — demonstrating that identity preservation and perceptual quality are complementary rather than competing objectives when the pipeline is correctly designed.

Formal Notation

The following notation formalizes the framework for readers familiar with machine learning conventions. Let:

- $E(x)$ = identity embedding function mapping an image to a normalized feature vector in embedding space (e.g., FaceNet-512 or ArcFace embedding)
- $G(x)$ = generative restoration model mapping a degraded input image to a restored output image
- $S(a, b)$ = cosine similarity between embedding vectors a and b, where $S(a, b) \in [-1, 1]$ and $S(a, b) = 1$ indicates identical embeddings
- x_{orig} = original or reference image providing the identity ground truth
- x_{deg} = degraded input image to be restored
- τ = configurable identity acceptance threshold (e.g., $\tau = 0.85$)
- W_f = adaptive weight vector over facial feature dimensions $f \in$ {eye spacing, nose, jawline, ear, ...}
- T = maximum iteration count before reporting failure

The multi-image identity template is constructed as:

$E_{template} = normalize(\sum_i E(x_i))$ *for all reference images* x_i *of the subject*

The acceptance condition for a restored output is:

Accept $G(x_{deg})$ *if* $S(E_{template}, E(G(x_{deg}))) \geq \tau$

If the condition is not met, the adaptive feedback step updates the generation constraints:

$W_f \leftarrow W_f + \alpha \cdot \nabla_f \cdot (\tau - S(E_{template}, E(G(x_{deg}))))$

$G(x_{deg}) \leftarrow G'(x_{deg}; W_f)$

where α is the learning rate for weight adaptation and ∇_f is the per-feature gradient of the similarity deficit. The iteration terminates when $S \geq \tau$ or when the iteration count reaches T, at which point the system reports identity preservation failure for that input.

This formulation makes clear that the proposed framework adds an outer optimization loop operating at the identity level, distinct from and external to the inner optimization loop of the generative model itself. Current restoration systems have only the inner loop. The proposed framework adds the outer loop.

Limitations

A credible research framework requires honest acknowledgment of its limitations. The following objections are anticipated and should be addressed in any subsequent implementation or publication.

1. Increased Inference Cost

Each regeneration cycle adds inference cost proportional to one full forward pass through the generative model. In the worst case — T iterations before timeout — total inference cost is T times the baseline. For large diffusion models this is non-trivial. The framework mitigates this through three factors: embedding similarity evaluation is computationally negligible relative to the generative step; the target applications are not latency-sensitive; and empirical observation suggests most cases converge within two to four cycles. However, cases involving severe degradation or extreme identity drift may require more iterations, and the computational budget must be designed accordingly. A practical implementation should expose T as a configurable parameter and allow users to trade identity accuracy for speed where appropriate.

2. Potential Over-Constraining

Strong identity constraints may conflict with the generative model's ability to fill in genuinely missing or unrecoverable information. A heavily damaged photograph may contain insufficient facial information to reconstruct identity-accurate geometry. In these cases, the system may iterate without converging, or may produce outputs that satisfy the identity constraint but exhibit visual artifacts introduced by repeated regeneration under tight constraints. Over-constraining is most likely to occur when the degraded input has lost a large proportion of identity-relevant facial information and the reference template is derived from photographs taken under very different conditions. The framework should include a graceful failure mode that reports low-confidence restoration rather than delivering an over-constrained output as if it were reliable.

3. Threshold Selection

The identity acceptance threshold τ is presented as configurable, but selecting an appropriate value in practice requires calibration. A threshold that is too high may cause the system to iterate indefinitely or time out on inputs where the maximum achievable identity fidelity is constrained by the degradation severity. A threshold that is too low may accept outputs with perceptible identity drift. The $\tau = 0.85$ value used illustratively in this document is derived from general face recognition literature and may not be optimal for restoration-specific applications, particularly for cross-age or cross-condition comparisons where cosine similarity between genuine matches is structurally lower. Empirical calibration of τ against application-specific data should be part of any implementation process.

4. Insufficient Identity Information in Degraded Inputs

The framework assumes the degraded input contains sufficient residual identity information for the feature encoder to extract a meaningful identity vector. For severely damaged photographs — those with large occluded facial regions, extreme blur, or significant structural loss — the initial embedding $E(x_{deg})$ may be unreliable, and the feedback loop may optimize against a noisy or misleading signal. The multi-image template construction partially addresses this by providing an independent reference embedding from undegraded photographs, but the system remains dependent on the generative model's ability to recover plausible identity geometry from the available signal. A practical implementation should include a degradation severity assessment step that estimates the reliability of the extracted embedding before committing to iterative regeneration.

5. Conflict Between Identity Preservation and Realism

Identity preservation and perceptual realism may conflict in specific cases. A subject with unusual facial geometry — proportions that fall outside the generative model's training distribution — may produce identity-accurate outputs that appear visually implausible to evaluators unfamiliar with the subject. Conversely, forcing a model trained on typical facial geometry to preserve atypical proportions may introduce visual artifacts that reduce perceptual quality scores. The proposed framework treats identity fidelity as the primary objective and perceptual quality as secondary, which is the correct priority ordering for the target applications. However, users should be informed that in edge cases this priority ordering may produce outputs that score lower on perceptual metrics than baseline systems, even when identity accuracy is higher.

6. Embedding Model Dependency

The framework's identity verification quality is bounded by the quality of the embedding model used. FaceNet and ArcFace are strong general-purpose face recognition embeddings, but they were trained primarily on contemporary photographs and may exhibit reduced sensitivity to identity characteristics in historical photographs with unusual lighting, photographic chemistry, or severe tonal degradation. An embedding model specifically fine-tuned on historical or archival photographs would likely improve framework performance for the primary target application. This represents an area for future work rather than a fundamental limitation of the framework architecture.

Future Work

The framework described in this memorandum represents a foundation architecture. Several directions for extension and application are identified below.

1. Video Restoration

Current video restoration systems process frames independently, producing temporal flickering and identity inconsistency that is perceptible even when individual frames score well on per-frame metrics. The proposed identity embedding framework extends naturally to video by treating the per-subject identity template as a temporal consistency constraint applied across all frames simultaneously. A restored video would be required not only to pass per-frame identity verification but to maintain consistent embedding similarity across the temporal sequence, eliminating the identity drift that accumulates frame-by-frame in current approaches. The feedback loop would operate at the sequence level rather than the frame level.

2. Personalized Avatars and Synthetic Identity

The multi-image template construction and adaptive identity loss mechanism are directly applicable to personalized avatar generation, where the requirement is to produce novel synthetic images of a specific individual that remain recognizable across varied poses, lighting, and styles. The framework's inference-time verification loop could serve as a quality gate for avatar generation pipelines, rejecting outputs that fail identity verification before delivery to the user. This application is commercially significant and technically adjacent to the restoration use case.

3. Medical Image Reconstruction

AI-based reconstruction of medical images — MRI, CT, and X-ray enhancement — faces an analogous problem to facial restoration: generative models optimize for visual plausibility at the cost of patient-specific anatomical fidelity. A misidentified or smoothed anatomical feature in a reconstructed medical image carries clinical consequences that a misidentified facial feature in a historical photograph does not. The identity-constrained supervisory layer concept translates directly: replace the facial identity embedding with an anatomical feature embedding, replace the facial similarity threshold with a clinically calibrated anatomical fidelity threshold, and apply the same iterative verification and regeneration loop. The framework architecture is domain-agnostic; the domain-specific component is the choice of embedding model and threshold calibration.

4. Authentication and Forensic Systems

In forensic and authentication contexts, the verifiability of the identity preservation claim is as important as the preservation itself. The framework's explicit similarity scoring produces a quantifiable, auditable record of identity fidelity for each restored output — a property that baseline systems do not provide. A forensic restoration system built on this framework could deliver not only a restored image but a signed similarity score report documenting the measured identity fidelity of the output, the threshold applied, and the number of regeneration cycles required. This audit trail could be material to evidentiary admissibility proceedings.

5. Multimodal Identity Embeddings

Current identity embeddings are derived from visual features alone. Future embedding models may incorporate multimodal identity signals — voice characteristics, gait patterns, biometric measurements, known biographical constraints — to construct richer identity templates that are more robust across severe degradation conditions. For historical photographs where visual information is severely compromised, a multimodal template that incorporates documentary evidence (known height, distinctive features recorded in historical records, comparative photographs) could provide a stronger identity constraint than visual embeddings alone. The framework architecture accommodates multimodal templates without modification; the embedding function E(x) is replaced with a multimodal embedding function E(x, m) where m represents auxiliary identity evidence.

6. LLM-Guided Adaptive Feature Weighting

The current framework applies fixed adaptive weighting rules based on predefined feature importance hierarchies. A more sophisticated implementation could use a large language model as a reasoning layer to interpret contextual information about the subject and dynamically adjust feature weights accordingly. For example: if documentary evidence indicates the subject was known for distinctive eye coloring, increase weighting on periocular features; if the subject is photographed in childhood and the reference template is from adulthood, reduce weighting on features known to change with age and increase weighting on age-stable biometrics. LLM-guided weighting would transform the framework from a fixed-rule system into a contextually adaptive one, potentially improving performance significantly on difficult cross-age and cross-condition restoration tasks.

Conclusion

Identity preservation remains an important and only partially solved problem in generative image restoration. Current systems have largely addressed the realism problem. The identity problem — ensuring that a restored image accurately represents the specific individual in the original photograph rather than a plausible generalization — has not been systematically solved.

The central contribution of the framework disclosed in the underlying patent specification (U.S. Patent No. 11,727,722 B2 and its first continuation, U.S. Patent No. 12,412,422 B2) is the shift from training-time identity constraints to inference-time identity verification. Rather than relying on a model's learned tendencies to preserve identity on average, the proposed approach evaluates every output against the specific identity of the subject being restored and iterates until that identity is preserved. This transforms identity preservation from a probabilistic training outcome into a verifiable per-output property.

The concepts disclosed in the specification can be interpreted as a framework that, implemented as an identity-constrained supervisory layer over existing diffusion architectures, may provide a framework for addressing this gap. The framework connects naturally to active research in identity loss functions, multi-view and consistency-constrained generation, adaptive inference-time optimization, and identity-preserving personalized synthesis.

The proposed experimental benchmark provides a concrete path from concept to evaluation. If iterative identity-feedback demonstrably improves FaceNet and ArcFace similarity scores and human recognition rates while maintaining competitive perceptual quality scores, the case for identity-constrained restoration becomes empirically grounded rather than theoretically argued.

The result is not merely more visually appealing — it is measurably more accurate to the identity of the original person. As image generation systems become increasingly capable, treating identity fidelity as an explicit, verifiable, per-output property rather than an incidental training outcome is the logical next step in the evolution of generative image restoration.

References

Identity Embedding and Loss Functions

[1] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. CVPR 2015. https://arxiv.org/abs/1503.03832

[2] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. CVPR 2019. https://arxiv.org/abs/1801.07698

[3] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric (LPIPS). CVPR 2018. https://arxiv.org/abs/1801.03924

Face Restoration Baselines

[4] Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior (GFPGAN). CVPR 2021. https://arxiv.org/abs/2101.04061

[5] Zhou, S., Chan, K., Li, C., & Loy, C. C. (2022). Towards robust blind face restoration with codebook lookup transformer (CodeFormer). NeurIPS 2022. https://arxiv.org/abs/2206.11253

[6] Wang, Z., Chen, J., & Hoi, S. C. H. (2022). RestoreFormer: High-quality blind face restoration from undegraded key-value pairs. CVPR 2022. https://arxiv.org/abs/2201.06374

Identity-Aware Generation and Synthesis

[7] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. CVPR 2023. https://arxiv.org/abs/2208.12242

[8] Wang, J., Cao, Y., Chen, C., He, Y., Jiang, B., Zhu, H., ... & Han, X. (2024). InstantID: Zero-shot identity-preserving generation in seconds. arXiv 2024. https://arxiv.org/abs/2401.07519

[9] Li, Z., Cao, M., Wang, X., Qi, Z., Cheng, M. M., & Shan, Y. (2024). PhotoMaker: Customizing realistic human photos via stacked ID embedding. CVPR 2024. https://arxiv.org/abs/2312.04461

[10] He, Y., Wang, J., Cao, Y., Yan, J., & Zhou, Y. (2024). ConsisID: Identity-preserving text-to-video generation by frequency decomposition. arXiv 2024. https://arxiv.org/abs/2411.17440

Image Quality Metrics

[11] Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. ICPR 2010. https://doi.org/10.1109/ICPR.2010.579

[12] Portal, L. U.S. Patent No. 11,727,722 B2: Method and System for Verifying Image Identification (parent patent). United States Patent and Trademark Office. Sole Inventor.

[13] Portal, L. U.S. Patent No. 12,412,422 B2: Method and System for Verifying Image Identification (first continuation; infrared identifying signature limitation). United States Patent and Trademark Office. Sole Inventor.

Note on reference [14]: an earlier version of this document cited a pre-grant publication, “U.S. Patent Application Publication No. US 2018/0068173 A1,” as the publication underlying references [12] and [13]. That publication number implies a filing roughly two to three years earlier than this patent family’s confirmed origin (provisional filed October 4, 2019; nonprovisional filed October 2, 2020), and has accordingly been removed from this reference list pending independent verification of the correct publication number, if a pre-grant publication citation is needed for this document’s purposes.