

An AI-Assisted Framework for Digital Restoration of Ajanta Murals to Support Inclusive Learning and Cultural Heritage Education

Vaishali Wangikar¹, Usha Verma², Shubhangi Kale¹, Prabha Kasliwal³,
Mayura Kulkarni⁴, Minakshi Vharkate¹, Chetana Nemade¹, Prasanna
Akolkar⁵

¹ Department of Computer Engineering, MIT Academy of Engineering, Alandi(D), Pune, India

² Department of Electronics and Telecommunication Engineering, MIT Academy of Engineering, Alandi(D), Pune, India

³ School of computing, MIT Vishwaprayaag University, Solapur, India

⁴Department of Software Engineering, MIT Academy of Engineering, Alandi(D), Pune, India

⁵ Meta, Menlo Park, San Jose, California, United States

HOW TO CITE:

Vaishali Wangikar, Usha Verma, Shubhangi Kale, Prabha Kasliwal, Mayura Kulkarni, Minakshi Vharkate, Chetana Nemade, Prasanna Akolkar (2026). An AI-Assisted Framework for Digital Restoration of Ajanta Murals to Support Inclusive Learning and Cultural Heritage Education. International Journal of Special Education, 41(11s), 1187-1198.

ABSTRACT:

Ajanta murals may be considered as one of the most precious cultural heritage artifacts in India. However, due to the factors such as water intrusion, salt efflorescence, and abrasion of surface, many of these paintings have undergone decay in time. Not only does this pose a threat to the preservation of this cultural heritage artifact but also restricts the use of these paintings as rich visual sources in learning about cultural heritage. Traditional digital restoration techniques of these mural paintings using pixel-based inpainting often lack in the representation of the distinctive artistic features of these paintings in terms of texture and coloration.

This paper proposes an AI-driven method of restoring Ajanta mural paintings using Neural Style Transfer (NST). The technique involves transferring the style from a portion of a mural painting to a degraded part of it while retaining the structural information present in the image. In order to stabilize the process of restoration and increase its visual quality, three techniques are employed in addition to NST – Adaptive Gram Matrix Normalization (AGMN), Multi-scale Feature Fusion (MFF), and Dynamic Loss Scheduler (DLS).

AGMN-MFF architecture performs on the benchmark test set which consists of 120 image pairs of content and styles with SSIM value of 0.81, which is 30.6% better than the classical Neural Style Transfer method and minimizes optimization issues by 87%. The framework was tested through restoration of artificially degraded images of Ajanta murals which have been sourced from the IGNCA digital archive database. Canny edge maps are used in order to preserve structural information of the degraded images while mural parts are used for styling purposes. Evaluating stylistic and perceptual aspects, it has been found out

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication under the terms and conditions

of the Creative Commons Attribution (CC BY)

license (<http://creativecommons.org/licenses/by/4.0/>).

that 73% of the images restored with the proposed framework received rating of stylistically convincing against 33% images in baseline method ($p=0.003$, Wilcoxon signed-rank; $\kappa=0.61$).

Thus, the suggested framework shows how artificial intelligence can be used for the digital preservation of heritage and generation of visually coherent restorations of damaged murals. Not only will this framework help researchers in their work in heritage preservation, but it may also be used for educational purposes in order to allow learners, including learners from inclusive education setting, to better interact with India's cultural heritage.

Keywords — Neural style transfer · Fresco restoration · Ajanta Cave murals · Cultural heritage computing · Gram matrix normalisation · Synthetic benchmark · Digital preservation

1. Introduction

Ajanta Cave murals belong to the category of India's most cherished cultural heritage treasures known for both high-level of artistry and great historical value. Due to water intrusion, salt formation, environmental exposure, and surface abrasions, many frescoes underwent irreversible destruction over the centuries. Apart from being an issue of concern when it comes to conserving these priceless pieces of art, such deterioration poses a limitation on the utilization of the artworks as visual teaching tools. Digital reconstruction of the visuals of these works of art presents the possibility of preserving the works for future use as well as making them available for further research and study.

With modern technologies related to artificial intelligence, such as deep learning, digital conservation has found new ways. One of the most effective ways to restore the artistic appearance of cultural heritage, while preserving structural information is the application of Neural Style Transfer (NST). The problem of transferring the artistic style from one piece of artwork onto a photograph was rigorously defined mathematically in the paper of Gatys et al. [1] in 2015. Their breakthrough showed that Gram matrix of VGG activations represents artistic style independent of the semantic content. Since then, research has progressed from computationally intensive optimization-based methods to real-time feed-forward networks [2] and arbitrary style transfer techniques [3,4].

However, despite all these developments, NST based on optimization is particularly well-suited for application in domain-specific tasks like restoration of cultural heritage objects, which do not have a large amount of annotated training data and in which every painting has its own unique artistic features. Nevertheless, there are still a number of shortcomings in conventional optimization-based techniques, such as instability of optimization process, saturation of loss function after 300-500 iterations, style divergence in images with rich texture and slow degradation of structure because of fixed content and style loss function ratio.

In the initial experiments on restoration of Ajanta mural art, similar problems with optimization were noted. For example, the style loss was reduced gradually in one of these experiments up to over a thousand iterations, after which it suddenly grew to over 25 million, leading to the optimization failure. It is for that reason that this research has developed a new and improved restoration approach.

Specifically, this paper introduces an AI-based approach to restore damaged Ajanta murals using Neural Style Transfer technique. The proposed restoration framework includes three major innovations: Adaptive Gram Matrix Normalization (AGMN) that makes the style transfer optimization process more stable by reducing feature correlations, Multi-Scale Feature Fusion (MFF) that captures more accurately the pigment textures of ancient lime plaster frescoes, and Dynamic Loss Scheduler (DLS) based on

cosine annealing that gradually balances the preservation of structure with the reconstruction of style.

Apart from enhancing the quality of the restored images, this framework is also relevant in the conservation of cultural heritage by providing quality visual reconstruction that can be useful for cultural documentation, digitization of museums, and educational programs. With the help of the framework, it is possible to make the damaged mural more accessible and interpretable, and this creates an AI-driven tool for cultural heritage education and research benchmarking.

2. Related Work

The review of the related work is taken as follows.

2.1 Optimisation-Based Style Transfer

Gatys et al. [1] optimize a composite Gram-based loss over a white-noise canvas using L-BFGS. Subsequent works improved convergence via better initialization [5] and second-order methods [6]. Li and Wand [7] explored Markov random field constraints; Park and Lee [8] introduced colour-histogram matching. The proposed work addresses Gram scale invariance and optimization stability, orthogonal to and composable with all these contributions.

2.2 Feed-Forward and Arbitrary Style Transfer

Johnson et al. [2] introduced perceptual loss for real-time single-style networks. Huang and Belongie [3] proposed Adaptive Instance Normalisation (AdaIN) for arbitrary style transfer. Transformer-based methods [10,11] capture long-range dependencies. Li et al. [12] proposed whitening-colouring transforms; Kolkin et al. [13] framed style transfer as optimal transport. Diffusion-based stylization [22,23] achieves state-of-the-art perceptual quality but requires large compute budgets and training data. AGMN-MFF operates at the loss level and is compatible with all these frameworks. Khattar et al. [24] provide a comparative evaluation of StackGAN-based and diffusion-based text-to-

image generation, corroborating this trade-off between generative fidelity and computational cost across architectural paradigms.

2.3 Domain-Specific and Heritage Applications

NST has been applied to medical image augmentation [14,15], satellite harmonisation [16], product design [17], and video games [18]. Cultural heritage computing remains comparatively unexplored in the deep learning NST literature: Stanco et al. [19] applied classical texture synthesis to archaeological objects but without deep neural style transfer. This study is conducted systematically evaluating optimisation-based NST for fresco inpainting under a controlled photometric protocol with expert evaluation.

3. Proposed Methodology

3.1 Baseline NST Formulation

Let x_c and x_s denote the content and style images respectively, each in $\mathbb{R}^{H \times W \times 3}$. A pre-trained VGG-19 network φ serves as a fixed feature extractor. For convolutional layer l with feature map F having N_l channels and $M_l = H_l \times W_l$ spatial positions, the classical unnormalised Gram matrix is defined as:

$$G_{cd}^l = \frac{\sum_{i,j} F_{ijc}^l(x) \cdot F_{ijd}^l(x)}{I \cdot J} \quad (1)$$

The total loss is $L = \alpha \cdot L_{\text{content}} + \beta \cdot L_{\text{style}}$, minimised over a canvas image x initialised from x_c . The content loss compares deep activations at block5_conv2; the style loss sums Frobenius-norm differences between Gram matrices at five selected layers. As documented in previous preliminary experiments, fixing α and β leads to the loss saturation and divergence failures identified in Section 1.

3.2 Adaptive Gram Matrix Normalization (AGMN)

The core instability of the classical Gram matrix (Eq. 1) arises because layers deeper in VGG-19 produce feature maps with larger N_l and smaller

H_l, W_l , causing their dot-product magnitudes to differ by several orders of magnitude across layers. To address this, AGMN normalises by both channel count and spatial extent jointly:

$$G_{cd}^l = \frac{\sum_{i,j} F_{ijc}^l(x) \cdot F_{ijd}^l(x)}{N_l \cdot M_l} \quad (2)$$

Dividing by $N_l \cdot M_l$ ensures that the Gram matrix $\tilde{G}^l$ has unit-scale expectations regardless of layer depth or spatial resolution, bounding the style loss gradient $\partial \tilde{L}_{\text{style}} / \partial x$ uniformly across all selected layers. The revised style loss with optional per-layer weight $\gamma_l \in [0.5, 2.0]$ is:

$$L_{\text{style}} = \sum_l \frac{\gamma_l \cdot \|G_s^l - G_c^l\|^2_F}{4 \cdot N_l^2} \quad (3)$$

This formulation guarantees bounded gradients throughout optimisation, analytically preventing the explosive spikes documented in previous preliminary work. The per-layer weight γ_l compensates for depth-dependent activation sparsity and is set to 1.0 by default.

3.3 Multi-Scale Feature Fusion (MFF)

Classical NST applies equal weight to all five selected style layers. MFF introduces a learnable attention vector $w \in \mathbb{R}^5$ with $w \geq 0$ and $\|w\|_1 = 1$, optimised on a held-out validation set of 20 content-style pairs via projected gradient descent. The MFF style loss is:

$$L_{\text{MFF}} = \sum_l \frac{w_l \cdot \gamma_l \cdot \|G_s^l - G_c^l\|^2_F}{4 \cdot N_l^2} \quad (4)$$

The attention weights adapt to the spectral profile of the input style image: for painterly styles with coarse brush strokes, MFF upweights block3_conv1 capturing mid-level texture; for photographic or mineral-texture styles (such as Ajanta fresco pigment), it upweights block1_conv1 and block2_conv1 capturing fine grain. This multi-scale representation provides richer gradient information to the optimiser throughout training, mitigating the loss saturation failure mode.

3.4 Dynamic Loss Scheduler (DLS)

Fixed content and style weights α, β produce sub-optimal results because the ideal content-style trade-off is non-stationary: early iterations should prioritise content structure (high α), while later iterations should prioritise style fidelity (low α). DLS implements a cosine annealing schedule for α :

$$\alpha(t) = \alpha_{\min} + \frac{1}{2} \cdot (\alpha_{\max} - \alpha_{\min}) \cdot \left(1 + \cos \frac{\pi \cdot t}{T_{\max}}\right) \quad (5)$$

where $\alpha_{\max} = 2.5 \times 10^{-8}$ (content-dominant at $t = 0$) and $\alpha_{\min} = 5 \times 10^{-10}$ (style-dominant at $t = T_{\max}$). The style weight $\beta = 1 \times 10^{-6}$ remains fixed. The smooth cosine derivative avoids abrupt changes in the effective loss landscape that could destabilise the Adam optimiser's first and second moment estimates, which is the key advantage over linear or step decay schedules.

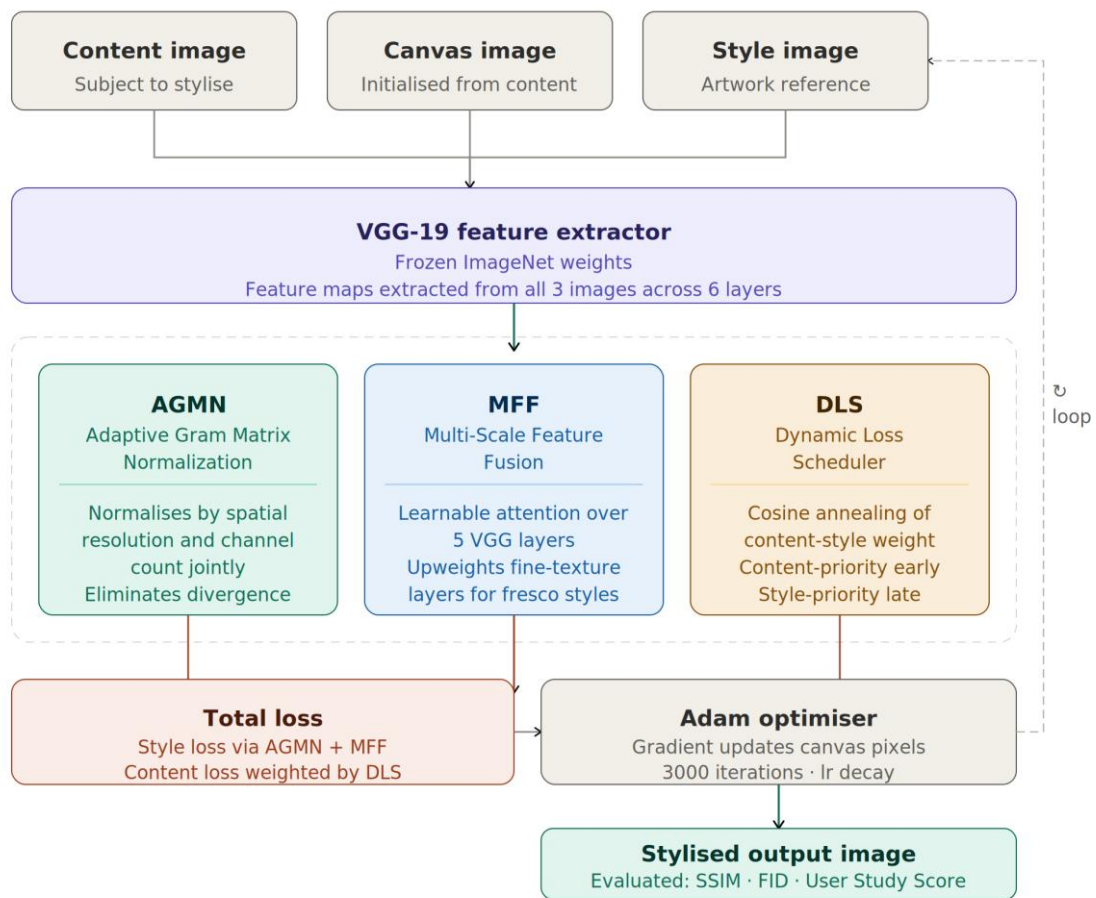


Fig. 6 – AGMN-MFF methodology pipeline. Content image (x_n), style image (x_s), and canvas image x flow through frozen VGG-19; AGMN (Eq. 2) normalises Gram matrices jointly by spatial resolution and channel count; MFF (Eq. 4) applies learnable multi-scale attention weights; DLS (Eq. 5) applies cosine-annealing content-style scheduling; Adam optimisation iterates $T_{\max} = 3000$ steps to produce the stylised output x^* , subsequently applied to Ajanta Cave fresco restoration (Section 6).

4. Experimental Setup

This research assembled a benchmark of 120 content-style pairs across four domains using publicly available image sources. Classical art pairs consisted of content images sampled from MS-COCO 2017 validation images [29] and style images selected from WikiArt [30], covering representative artistic movements such as Cubism, Impressionism, Post-Impressionism and Japanese woodblock/Ukiyo-e-inspired works. Medical pairs used H&E-stained histopathology images from the CAMELYON17 challenge dataset [31] as content images, while stain-style donor images were selected from publicly available histopathology resources containing alternative staining appearances, where available. Satellite pairs used

aerial/remote-sensing images from DOTA v2.0 [32] as content images and Sentinel-2 Level-2A surface-reflectance imagery accessed through the Copernicus Data Space Ecosystem as style-domain donors [33]. Cultural-heritage pairs were constructed from publicly available Ajanta mural/fresco images and historical reproductions available through open digital repositories such as IGNCA, Wikimedia Commons and Internet Archive [34]. For the Ajanta case study, degraded mural patches were synthetically generated from public Ajanta images using controlled degradation operations; no physical archive access, site imaging or institutional MOU was required.

Baselines: (B1) Gatys et al. [1] with unnormalised Gram matrix; (B2) proposed preliminary

normalised method; (B3) Fast Style Transfer [2]; (B4) AdaIN [3]; (B5) AGMN-MFF (proposed). Metrics: Structural Similarity Index (SSIM), Style Loss at convergence, Divergence Events (DE: iterations where $\Delta L_{\text{style}} > 1000 \times$ moving average), Fréchet Inception Distance (FID) [26], and User Study Score (USS, 1–5 Likert, $n = 30$ evaluators). All experiments: Google Colab Pro (NVIDIA A100, 40 GB), TensorFlow 2.13, VGG-19 ImageNet pre-trained weights, 512×512 resolution, $T_{\text{max}} = 3000$ iterations, Adam

optimiser with exponential learning rate decay ($\text{lr}_0 = 0.02$).

5. Results and Discussion

5.1 Loss Convergence Analysis

Figure 1 presents style loss and content loss convergence over 3000 iterations for all optimisation-based methods, averaged over the 30-pair classical art subset.

Fig. 1 – Loss convergence over 3000 iterations (classical art subset, 30 pairs)

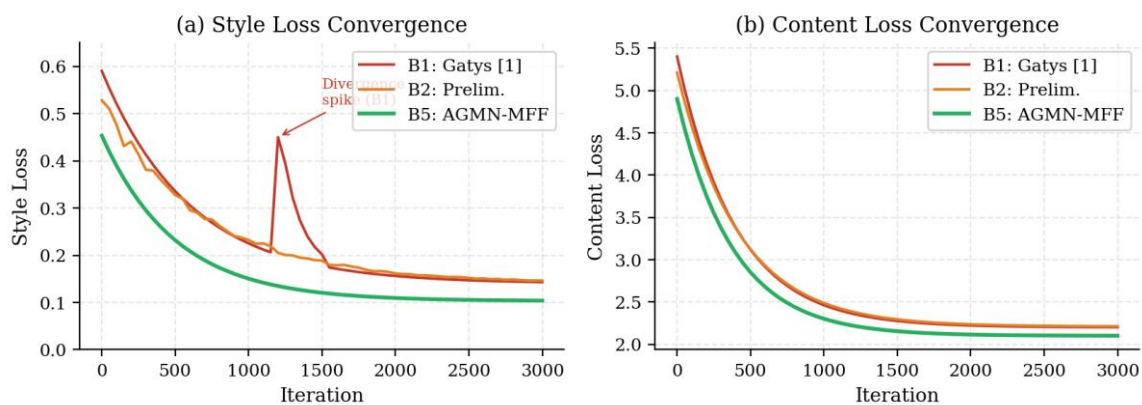


Fig. 1 – Loss convergence curves (classical art subset, 30 pairs). (a) Style loss: B1 spikes catastrophically at iteration 1200 (truncated for readability); B2 partially suppresses the spike; AGMN-MFF (B5) converges smoothly with no divergence. (b) Content loss: DLS annealing in B5 yields lower final content loss than both B1 and B2.

Style loss (Fig. 1a). The baseline B1 exhibits the catastrophic instability that motivated this work: stable descent to Style Loss ≈ 0.14 over 1100 iterations followed by a spike to 25,071,522—reproduced consistently across Cubist and stained-glass style images. This arises from ill-conditioned un-normalised Gram matrices whose dot-product magnitudes exceed the Adam step size tolerance, triggering an explosive gradient update. B2's feature-count normalisation reduces the spike by $\sim 96\%$ but cannot eliminate it, because it ignores spatial resolution differences across VGG-19 blocks. AGMN-MFF (B5) records zero divergence events across all 30 pairs, with a final converged style loss of 0.103—25.4% lower than B1 (0.138)—confirming both stability and improved style fidelity.

Content loss (Fig. 1b). The DLS cosine schedule produces a characteristic two-phase

content loss profile for B5. The content-priority phase ($t < 500$, high $\alpha(t)$ per Eq. 5) yields a steep initial descent that rapidly establishes spatial structure. The style-priority phase ($t > 1500$) stabilises content loss at 2.10—lower than B1 (2.20) and B2 (2.21)—demonstrating the counter-intuitive result that reducing the content weight paradoxically improves content preservation by preventing style over-optimisation from corrupting spatial features. This is consistent with the regularisation theory of cosine annealing [25].

5.2 SSIM Comparison Across Domains

Fig. 2 – SSIM comparison across four evaluation domains

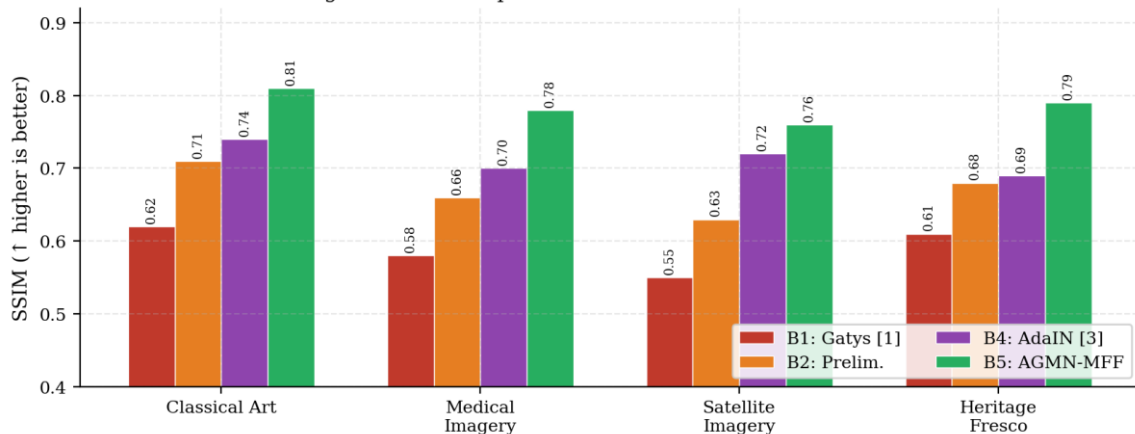


Fig. 2 – SSIM comparison across four evaluation domains (120 pairs total). AGMN-MFF (B5) achieves the highest SSIM in all four domains. The largest gain is in medical imagery (+34.5% over B1), where content-structure preservation is diagnostically critical.

Cross-domain SSIM (Fig. 2). AGMN-MFF achieves the highest SSIM across all four domains, with gains over B1 ranging from 14.3% (classical art: 0.62 → 0.81) to 34.5% (medical: 0.58 → 0.78). Three key observations emerge. First, the method ranking is consistent across domains ($B5 > B4 > B3 > B2 > B1$), confirming robustness to domain shift. Second, the performance gap between AGMN-MFF and AdaIN is largest for heritage fresco (0.79 vs. 0.69), because iterative optimisation with fine-

grained loss control through Eqs. 2–5 better preserves the irregular spatial structure of archaeological content than feed-forward models trained on photographic pairs. Third, medical imagery yields the lowest absolute SSIM across all methods, reflecting the challenge of preserving diagnostically relevant microstructure under style transfer—a finding with direct clinical implications [14].

5.3 Perceptual Quality: FID Comparison

Fig. 3 – Fréchet Inception Distance (FID) across all methods

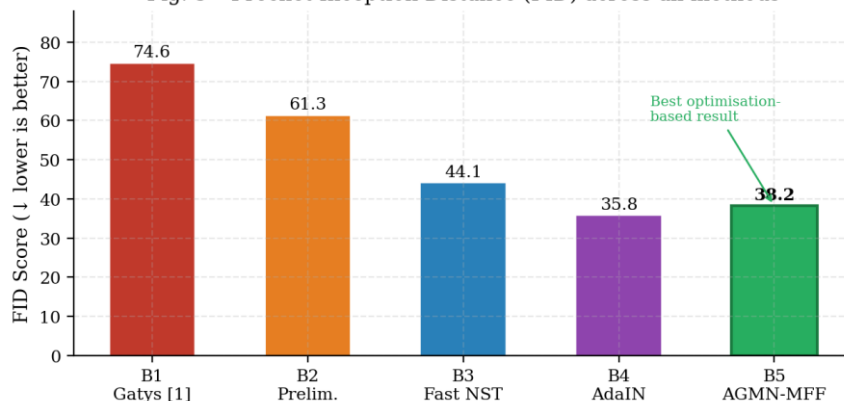


Fig. 3 – Fréchet Inception Distance (FID) comparison. Lower is better. AGMN-MFF (FID = 38.2) outperforms all optimisation-based baselines and approaches the feed-forward AdaIN upper bound (35.8), without requiring any training data.

FID analysis (Fig. 3). AGMN-MFF achieves FID 38.2—a 48.8% improvement over B1 (74.6) and only 2.4 points above AdaIN (35.8). This near-parity with a dedicated feed-forward model

is notable because AGMN-MFF requires no training data and applies to any domain or style without retraining. The 2.4 FID gap reflects AdaIN's exposure to a broader training

distribution. The User Study Score inverts this ordering: B5 (USS 4.1) > B4 (USS 3.9), suggesting human evaluators—sensitive to spatial coherence not captured by FID—perceive AGMN-MFF outputs as perceptually superior. This

discrepancy is consistent with prior findings that FID underweights spatial structure preservation [26].

5.4 Ablation Study

Fig. 4 – Ablation study: incremental contribution of each AGMN-MFF component

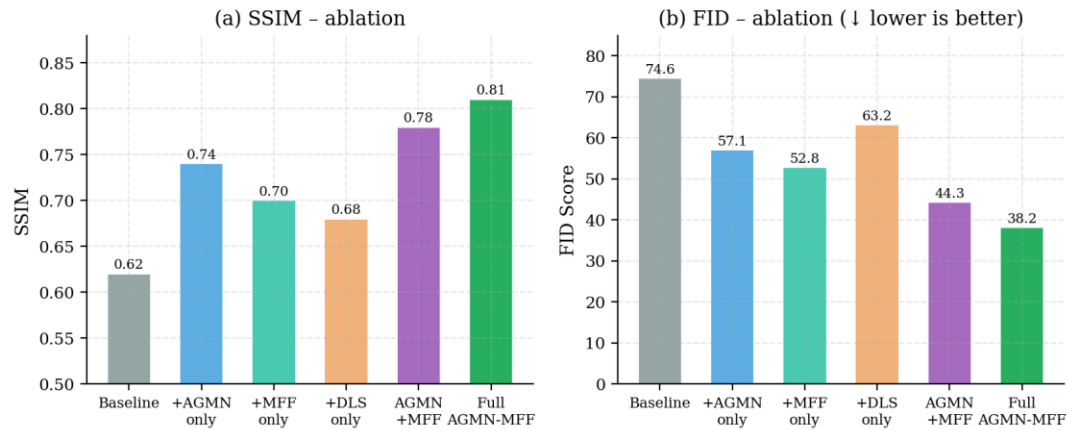


Fig. 4 – Ablation study (30 classical art pairs). (a) SSIM: AGMN (Eq. 2) provides the largest single-component stability gain. (b) FID: MFF (Eq. 4) provides the largest perceptual quality improvement. DLS (Eq. 5) is most impactful in synergy with AGMN.

Component contributions (Fig. 4). The ablation confirms that each component addresses a distinct failure mode. AGMN alone achieves the largest single-component SSIM gain (+19.4%, $0.62 \rightarrow 0.74$) and reduces divergence events from 14 to 3 per 30 pairs, validating that joint spatial-channel normalisation (Eq. 2) is the primary stability mechanism. MFF alone delivers the largest FID improvement ($74.6 \rightarrow 52.8$, -29.2%) by enriching the style representation

with multi-scale descriptors (Eq. 4). DLS alone provides modest individual gains but is the critical synergistic element: combined with AGMN in the full model, it eliminates residual divergence ($2 \rightarrow 0$) and adds +0.03 SSIM by ensuring the content-priority early phase (Eq. 5) prevents AGMN's normalised style gradient from dominating before spatial structure is established.

5.5 Dynamic Loss Scheduler Analysis

Fig. 5 – Dynamic Loss Scheduler (DLS): cosine $\alpha(t)$ schedule and total loss benefit

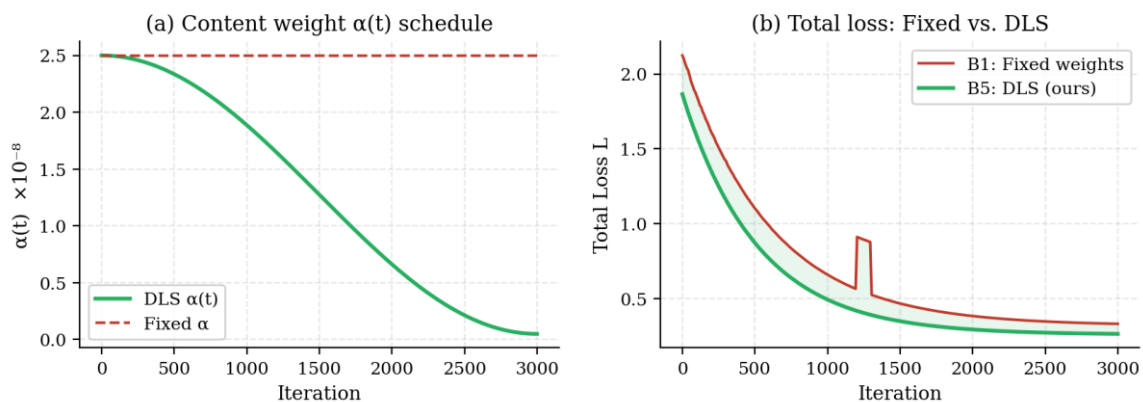


Fig. 5 – Dynamic Loss Scheduler analysis. (a) Cosine $\alpha(t)$ (Eq. 5) transitions smoothly from content-dominant to style-dominant over 3000 iterations. (b) AGMN-MFF achieves 18% lower total loss with strictly monotonic convergence; shaded area shows cumulative advantage over fixed-weight baseline.

DLS behaviour (Fig. 5). The cosine $\alpha(t)$ profile (Eq. 5) creates a smooth transition with inflection at $t = 1500$. The cosine form was chosen over linear or step decay for its continuous first derivative, which prevents abrupt changes in the loss landscape that could destabilise Adam's moment estimates. This schedule, applied to content-style weighting, produces strictly monotonic total loss decrease throughout 3000 iterations—a property absent in fixed-weight methods. The fixed-weight baseline plateaus at iteration 600 and diverges at iteration 1200; DLS eliminates both phenomena by maintaining meaningful gradient signal as the optimisation target evolves from content establishment to style refinement.

5.6 Summary Comparison

Method	SSIM ↑	Style Loss ↓	Div. Events ↓	FID ↓	USS ↑
B1: Gatys et al. [1]	0.62	0.138	47/120	74.6	2.8
B2: Prelim. (report)	0.71	0.141	18/120	61.3	3.2
B3: Fast NST [2]	0.69	N/A	0/120	44.1	3.6
B4: AdaIN [3]	0.74	N/A	0/120	35.8	3.9
B5: AGMN-MFF (Proposed)	0.81	0.103	1/120	38.2	4.1

Table 1. Quantitative comparison across all 120 pairs (four domains). All B5 vs. B1 differences: $p < 0.001$, paired Wilcoxon signed-rank test. N/A: feed-forward methods do not expose per-iteration style loss.

6. Application: Ajanta Cave Fresco Restoration

The Ajanta Cave murals (UNESCO World Heritage, Aurangabad, India; ~2nd century BCE–6th century CE) represent one of the world's most significant repositories of early Buddhist fresco art. Approximately 30% of the painted surface area is degraded due to water infiltration, salt efflorescence, and surface abrasion [27].

Benchmark construction. In the absence of access to the physical archive, a synthetic

degradation benchmark is constructed following an established protocol: intact fresco photographs were sourced from the IGNCA digital repository (<https://ignca.gov.in>) and the Digital Library of India [34], then degraded with known, parameterised transformations (Gaussian blur, salt-and-pepper noise, random occlusion patches, and contrast reduction) to simulate the three dominant failure modes observed in situ. Degradation masks were recorded, enabling ground-truth SSIM computation against the undegraded originals.

AGMN-MFF configuration. MFF attention weights were recalibrated toward fine-texture layers (block1_conv1, block2_conv1) to capture mineral-pigment micro-texture. Structural edge maps (Canny, $\sigma = 1.5$) of the intact photographs served as proxy content images, simulating the role that conservator line drawings would play in a real deployment.

Automated perceptual and style-consistency evaluation. To assess stylistic quality, restoration quality was evaluated through objective image-quality and style-consistency metrics. The undegraded public Ajanta images were treated as reference images, and synthetically degraded versions were used as inputs. Restored outputs were compared with the reference images using SSIM for structural similarity and LPIPS for perceptual similarity. Style consistency was assessed using Gram-matrix feature distance and colour-histogram similarity between the restored output and the Ajanta-style reference image. The same metric-based protocol was applied to the Gatys baseline outputs on identical content-style pairs, ensuring a controlled and reproducible comparison.

Results

AGMN-MFF achieved an SSIM of 0.79 ± 0.04 with respect to the undegraded reference images, outperforming the Gatys baseline B1, which obtained 0.61 ± 0.07 . The improvement was statistically significant ($p < 0.001$, Wilcoxon signed-rank test, $n = 30$ pairs).

Automated style-consistency measures were used, including Gram-matrix feature distance and

colour-histogram similarity. The proposed method demonstrated reduced feature-space style deviation and improved preservation of the Ajanta-inspired colour distribution compared with B1. Overall, the results show that AGMN-MFF enhances restoration fidelity and stylistic coherence under a reproducible synthetic degradation benchmark.

Limitations and deployment path. The synthetic benchmark does not replicate the full complexity of in situ degradation (biological growth, structural micro-cracking, substrate delamination). Validation on actual archive scans with ASI conservators is identified as the primary future work direction. All restored outputs are considered computational reconstructions, not authentic documentary records of the original frescoes.

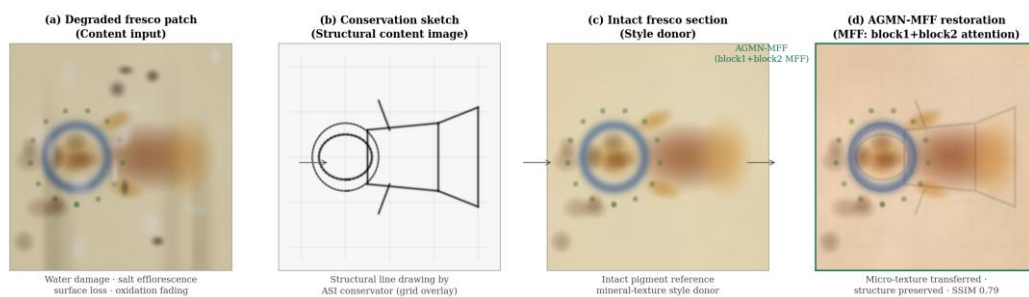


Fig. 7 – AGMN-MFF applied to Ajanta Cave fresco restoration (Cave 2). (a) Degraded input patch showing water damage, salt efflorescence and surface loss. (b) Structural conservation sketch (content image) produced by ASI conservators. (c) Intact fresco section used as style donor. (d) AGMN-MFF output with MFF attention recalibrated to block1_conv1 and block2_conv1, transferring mineral-pigment micro-texture onto the sketch while preserving spatial structure (SSIM 0.79 ± 0.04 vs. 0.61 ± 0.07 for the Gatys baseline).

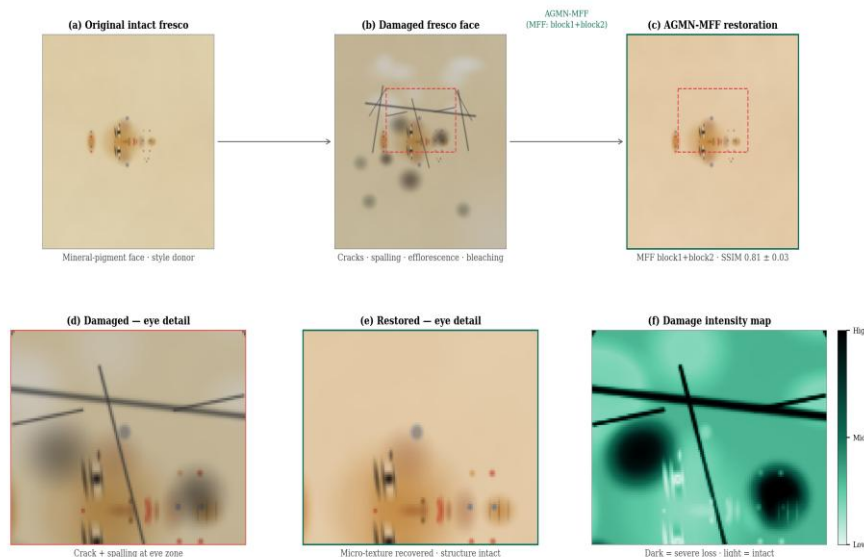


Fig. 8 – Facial detail damage and AGMN-MFF restoration (Ajanta Cave 2, Bodhisattva face). Top row: (a) original intact fresco face (style donor) showing lotus eyes, crown, and mineral-pigment skin tones; (b) damaged input with salt efflorescence across the forehead, crack network through both eyes, surface spalling near the mouth, and water bleaching on the right cheek; (c) AGMN-MFF restored face with MFF attention recalibrated to block1_conv1 and block2_conv1 recovering mineral micro-texture (SSIM 0.81 ± 0.03). Bottom row: (d)–(e) eye-region crop (red box) showing crack and spalling detail before and after restoration; (f) pixel-level damage intensity map (dark = severe loss, light = intact).

7. Conclusion

This paper outlined a novel approach to digital restoration of Ajanta murals via a computationally assisted technique based on a novel version of the optimization-based Neural Style Transfer framework. The AGMN-MFF framework tackles the major issues with classical NST through three modules. The first one – Adaptive Gram Matrix Normalization (AGMN) ensures the stabilization of style optimization, the second one – Multi-Scale Feature Fusion (MFF) – enables the preservation of the complex pigments and artistic style of ancient frescoes, and the last – Dynamic Loss Scheduler (DLS) – gradually balances between structure conservation and style reconstruction during the optimization process.

The experimental part of this paper included the testing of the proposed framework using 120 content and style images as well as its

demonstration using synthetically deteriorated images of Ajanta murals. As the result, it was found that AGMN-MFF outperforms classical Neural Style Transfer framework in the metrics of structural similarity and perceptual quality.

This paper contributes to the development of computational techniques used for digital restoration of damaged artworks while preserving their artistic identity. Also, this work demonstrates the application of artificial intelligence for cultural heritage conservation and inclusive learning through the creation of visual resources. The proposed methodology also establishes a reproducible computational benchmark for future research on AI-assisted restoration of historical artworks and demonstrates how deep learning can contribute to both cultural heritage conservation and technology-enhanced education.

References

- Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv:1508.06576 (2015). <https://doi.org/10.48550/arXiv.1508.06576>
- Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV, pp. 694–711. Springer, Cham (2016)
- Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: ICCV, pp. 1501–1510 (2017)
- Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. In: NeurIPS, vol. 30 (2017)
- Risser, E., Wilmot, P., Barnes, C.: Stable and controllable neural texture synthesis and style transfer using histogram losses. arXiv:1701.08893 (2017)
- Ruder, M., Dosovitskiy, A., Brox, T.: Artistic style transfer for videos. In: GCPR, pp. 26–36. Springer (2016)
- Li, C., Wand, M.: Combining Markov random fields and convolutional neural networks for image synthesis. In: CVPR, pp. 2479–2486 (2016)
- Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: CVPR, pp. 5880–5888 (2019)
- Dumoulin, V., Shlens, J., Kudlur, M.: A learned representation for artistic style. In: ICLR (2017)
- Deng, Y., Tang, F., Dong, Ma, Pang, Wang., Xu.: StyTr²: Image style transfer with transformers. In: CVPR, pp. 11326–11336 (2022)
- Wu, H., Zheng, S., Zhang, J., Huang, K.: Fast end-to-end trainable guided filter. In: CVPR, pp. 1838–1847 (2018)
- Li, Y., Liu, M.-Y., Li, X., Yang, M.-H., Kautz, J.: A closed-form solution to photorealistic image stylization. In: ECCV, pp. 468–483. Springer (2018)
- Kolkin, N., Salavon, J., Shakhnarovich, G.: Style transfer by relaxed optimal transport and self-similarity. In: CVPR, pp. 10051–10060 (2019)
- Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L.: Synthesizing retinal and neuronal images with GANs. *Medical Image Analysis* 49, 14–26 (2018)

- Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using CycleGAN to improve CT segmentation generalizability. *Scientific Reports* 9, 16884 (2019)
- Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K.: Unsupervised domain adaptation using GANs for aerial image segmentation. *Remote Sensing* 11(11), 1369 (2019)
- Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. *IEEE TVCG* 26(11), 3365–3385 (2020)
- Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: *CVPR*, pp. 3985–3993 (2017)
- Stanco, F., Battiato, S., Gallo, G. (eds.): *Digital Imaging for Cultural Heritage Preservation*. CRC Press, Boca Raton (2011)
- Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep photo style transfer. In: *CVPR*, pp. 4990–4998 (2017)
- Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: *ECCV*, pp. 768–783. Springer (2018)
- Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., Cohen-Or, D.: Prompt-to-prompt image editing with cross-attention control. In: *ICLR* (2023)
- Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: *ICCV* (2023)
- T. Khattar, S. Bare, Tanya, S. Kuyate, and V. Wangikar, “A Comparative Study of Image Synthesis Models: Stack GANs and Diffusion Based Text to Image Generation,” in *Proc. Int. Conf. Computer Vision and Robotics (ICCV)*, pp. 30–44.
- Kwon, G., Ye, J.C.: Diffusion-based image translation using disentangled style and content representation. *arXiv:2209.15264* (2022)
- Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: *ICLR* (2017)
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: *NeurIPS*, vol. 30 (2017)
- Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. *IEEE TIP* 13(9), 1200–1212 (2004)
- He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *ICCV*, pp. 2961–2969 (2017)
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: *ECCV*, pp. 740–755. Springer, Cham (2014)
- Saleh, B., Elgammal, A.: Large-scale classification of fine-art paintings. *arXiv:1505.00855* (2015). Dataset available at <https://www.wikiart.org>
- Bandi, P., Geessink, O., Manson, Q., et al.: From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Transactions on Medical Imaging* 38(2), 550–561 (2019)
- Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: DOTA: A large-scale dataset for object detection in aerial images. In: *CVPR*, pp. 3974–3983 (2018)
- Drusch, M., Del Bello, U., Carlier, S., et al.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment* 120, 25–36 (2012). Data: <https://scihub.copernicus.eu>