IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Optimizing the VFX Pipeline: A Comparative Study of AI-Driven vs. Manual Workflows in Rotoscoping and Inpainting

Mr. Tushar N. Chaudhari
Student, Department of Computer Science
Vidyavardhini's College of Engineering and Technology, Vasai, India

Abstract: *"The Visual Effects (VFX) industry is currently facing a massive surge in content demand, making traditional, labor-intensive processes like rotoscoping a major production bottleneck. This research explores the integration of Artificial Intelligence (AI) to automate the rotoscoping workflow within a standard VFX pipeline. By conducting a comparative analysis between manual spline-based rotoscoping and AI-driven semantic segmentation models (such as Segment Anything Model), this paper evaluates the efficiency, edge accuracy, and temporal consistency of both methods. Experimental results indicate that AI-powered automation can reduce the initial masking time by approximately 70-80%, though human intervention remains necessary for fine-tuning complex motion. This study concludes that while AI cannot yet fully replace manual artistry, its role as a pre-processing tool significantly optimizes turnaround times for modern post-production houses, providing a scalable solution for high-volume VFX tasks."*

Keywords: VFX Pipeline, Rotoscoping, AI Automation, Digital Inpainting, Machine Learning, Computer Vision

I. INTRODUCTION

1.1 Background of the Study

Visual Effects (VFX) have transitioned from a luxury in filmmaking to a fundamental necessity. In the modern digital era, almost every frame of a cinematic production undergoes some form of digital manipulation. This is especially true in massive film industry around the globe, where the volume of content creation for OTT platforms and cinema has increased exponentially. However, the underlying technical processes specifically Rotoscoping and Inpainting remain remarkably labor-intensive.

1.2 The Evolution of VFX Pipelines

Traditionally, a VFX "Show Setup" involves a linear pipeline: Ingestion, Tracking, Rotoscoping, Paint/Inpainting, and finally Compositing. Rotoscoping, the process of creating a matte or mask for an object in a moving image, is the foundation of most VFX shots. Digital Inpainting, on the other hand, involves removing unwanted elements (like tracking markers, wires, or even people) and reconstructing the background. For decades, these tasks were performed frame-by-frame by junior artists, a process that is both prone to human error and financially draining for studios.

1.3 The Advent of AI in Post-Production

The field of Computer Science, specifically Artificial Intelligence (AI) and Computer Vision (CV), has recently introduced revolutionary models capable of understanding image depth and semantic segments. Models like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) are now being integrated into

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

VFX software. This shift promises to transform the "Traditional Pipeline" into an "AI-Enhanced Pipeline," significantly reducing the time spent on "low-level" tasks.

II. LITERATURE REVIEW

2.1 Overview of Traditional Rotoscoping Techniques

Rotoscoping has its roots in early animation, but in the digital era, it transitioned to spline-based manipulation. According to early industry standards established by tools like Adobe After Effects, Nuke and Silhouette FX, the process involves creating closed-loop Bézier curves around an object. The technical challenge lies in "Temporal Consistency" ensuring that the mask does not "jitter" or "chatter" across multiple frames. Traditional algorithms relied on basic point-tracking (Feature Tracking) which often fails during motion blur or occlusions, necessitating manual frame-by-frame adjustment by a technician.

2.2 Digital Inpainting and Pixel Reconstruction

Digital Inpainting, or "Image Completion," is the process of filling missing pixels in a way that is visually undetectable. Early computational methods, such as the PatchMatch algorithm (Barnes et al., 2009), focused on finding similar patches within the same image to fill holes. While effective for static images, these methods struggled with video, where lighting changes and camera movement require "Temporal Inpainting" where the algorithm must look at frames before and after the current one to find the missing data.

2.3 The Shift to Deep Learning and CNNs

The introduction of Convolutional Neural Networks (CNNs) shifted the focus from pixel-matching to semantic understanding. Research by Long et al. (2015) on Fully Convolutional Networks (FCNs) for semantic segmentation paved the way for automated masking. Instead of tracing an edge, the AI "understands" what a "human" or a "car" looks like and generates a probability map, which is then converted into a binary mask. This significantly reduces the manual workload but introduces a new problem: the AI often ignores fine details like hair or motion-blurred edges.

2.4 Segment Anything Model (SAM) and Zero-Shot Learning

A major breakthrough occurred with Meta AI's Segment Anything Model (SAM) in 2023. Unlike previous models that required training on specific objects, SAM is a "Foundation Model" capable of zero-shot generalization. In a VFX pipeline, this means a "Show Setup" can be automated by providing a single "prompt" (a click or a box) around an object, and the AI handles the segmentation for the entire sequence. This represents the current "state-of-the-art" in AI-driven VFX optimization.

2.5 Generative Adversarial Networks (GANs) for Video Inpainting

For Inpainting, GANs have become the standard. A GAN consists of two networks: a Generator that tries to fill the hole, and a Discriminator that tries to detect if the fill is "fake." This competition results in highly realistic textures. Modern architectures like ProPainter use "Recurrent Flow-Guided Transformers" to ensure that the reconstructed background remains stable as the camera moves, solving the "flicker" issues found in earlier automated methods.

2.6 Identified Research Gap

While there is significant academic research on individual AI models (SAM, GANs, Transformers), there is a lack of comparative data specifically focused on the VFX production environment in Industry. Most studies focus on accuracy metrics (like mIoU) but ignore the "Time-to-Delivery" metric, which is critical for local studios. This paper aims to fill that gap by providing a direct comparison of "Man-Hours" versus "Machine-Hours.".

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

III.PROBLEM DEFINITION

The modern VFX production cycle faces a critical contradiction: while camera technology and resolution (4K/8K) are advancing rapidly, the foundational tasks of Rotoscoping and Inpainting remain tethered to manual, frame-by-frame manipulation. The core problems addressed in this research are defined as follows:

3.1 Technical Bottleneck (Temporal Consistency):

Manual rotoscoping often results in "jitter" due to human fatigue, while basic automated tools fail to maintain edge integrity during high-speed motion or occlusions (when one object passes behind another).

3.2 Scalability Issues:

As most vfx studios move toward high-volume OTT content (Netflix, Amazon Prime), the "Man-to-Shot" ratio is becoming unsustainable. Traditional workflows require one artist per complex shot, leading to massive overheads and delayed deliveries.

3.3 Data Inefficiency:

Traditional software does not "learn" from previous frames. If an artist rotoscopes frame 1, the software doesn't automatically understand the movement of frame 100 without manual keyframing.

3.4 Reconstruction Artifacts:

In digital inpainting, removing large objects often leaves "ghosting" or "blurring" artifacts because traditional algorithms cannot intelligently predict the background texture based on the surrounding environment.

IV. OBJECTIVE AND SCOPE

4.1 Primary Objective

The primary objective of this research is to design and evaluate an optimized VFX pipeline that integrates AI-driven models to automate the tasks of Rotoscoping and Inpainting. The study seeks to prove that a "Hybrid Workflow" (AI-assisted with minimal human supervision) is superior to the traditional "Manual Workflow" in terms of time, cost, and technical accuracy.

4.2 Specific Goals

To implement the Segment Anything Model (SAM) for automated mask generation and compare its speed against manual spline-tracing. To evaluate the effectiveness of Deep Learning-based Inpainting (specifically using GAN-based architectures) in reconstructing complex backgrounds. To quantify the "Optimization Percentage" in a standard "Show Setup."

4.3 Scope of the Study

Target Domain: The study is focused on the post-production industry, like vfx around the globe.
Technical Boundary: The research is limited to 2D and 2.5D image processing. It does not cover 3D matchmoving or liquid simulations.
Data Set: The experiment will utilize a 5-10 second video sequence (high motion) to test the limits of both manual and AI methods.
Software Boundary: The comparison will be held between industry-standard manual tools (like Adobe After Effects, Nuke or Silhouette FX) and Python-based AI implementations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

V. RESEARCH METHODOLOGY

5.1 Research Design

This study employs an Experimental and Comparative Research Design. Two distinct pipelines the "Manual Baseline" and the "AI-Enhanced Pipeline" are executed on the same raw video dataset to ensure a fair comparison. The research focuses on two primary variables: Processing Time (Independent Variable) and Visual Accuracy (Dependent Variable).

5.2 Technical Specifications

To ensure reproducibility, the experiment is conducted in a controlled environment:

Hardware Layer:

CPU: Intel Core i7-12700K (12 Cores)
GPU: NVIDIA GeForce RTX 3060 with 12GB VRAM (Crucial for CUDAbased
AI processing)
RAM: 32GB DDR4

Software Layer:

Operating System: Ubuntu 22.04 LTS / Windows 11
Development Environment: Python 3.10, PyTorch 2.0
VFX Tools: Adobe After Effects (Manual Control)/Nuke/Silhouette FX, Blender (VFX Environment)
AI Frameworks: Meta's Segment Anything Model (SAM), ProPainter/LaMa libraries.

5.3 System Architecture (The AI Pipeline)

The proposed optimized pipeline follows a four-stage architecture:
1. Ingestion & Pre-processing: Converting raw footage into image sequences (PNG/EXR) and normalizing resolution to 1080p.
2. The Segmentation Module (Rotoscoping): The user provides a "Point Prompt" on the first frame. The SAM Encoder generates a feature embedding. A Mask Tracker (e.g., XMem) propagates the mask across subsequent frames to maintain temporal consistency.
3. The Inpainting Module (Object Removal): The generated masks are fed into a Generative Adversarial Network (GAN). The algorithm identifies "hole" regions and performs "Flow-Guided Feature Propagation" to fill pixels from neighbouring frames.
4. Export & Post-Processing: Re-compositing the cleaned frames and applying a Gaussian Blur filter to the edges of the masks to mimic natural lens fall-off.

5.4 Experimental Procedure

Phase A (Manual): An artist uses the 'Pen Tool' in After Effects. Keyframes are placed every 5 frames. Tracking markers are removed using the 'Clone Stamp' tool.
Phase B (AI-Driven): A Python script is executed to run the SAM-Inpainting pipeline. The script automates the mask generation for 150 frames without manual keyframing.

5.5 Evaluation Metrics

The findings will be analyzed based on:
Temporal Stability: Measuring "jitter" in the mask edges.
Throughput: Frames Processed Per Hour (FPPH).
Computational Overhead: Monitoring VRAM and CPU utilization.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

VI. ANALYSIS & FINDINGS

6.1 Quantitative Analysis:

Performance Metrics The primary focus of this study was to quantify the optimization achieved through AI integration. The experiment was conducted on a 5-second video clip at 30 FPS (Total 150 frames). The results are summarized in Table 1.

Table 1: Comparative Analysis of Production Time

Task Description	Manual Workflow (Mins)	AI-Driven Workflow (Mins)	Optimization (%)
Show Setup & Ingestion	10	05	50%
Rotoscoping (150 Frames)	140	12	91.4%
Inpainting (Object Removal)	90	18	80%
Final Rendering & QC	20	10	50%
Total Production Time	260 Mins (4.3 hrs)	45 Mins (0.75 hrs)	82.7%

6.2 Efficiency and Throughput

The data indicates a massive surge in Throughput. While the manual artist processed approximately 1.1 frames per minute, the AI-driven pipeline processed 12.5 frames per minute. For a large-scale studios handling 1,000+ shots, this optimization represents a significant reduction in the required workforce for "low-level" tasks, allowing senior artists to focus on high-level compositing.

6.3 Qualitative Analysis: Visual Fidelity

Beyond speed, the research analyzed the quality of the output:

Edge Integrity:

Manual rotoscoping provided the sharpest edges but suffered from "human jitter" in complex motion. AI-generated masks (via SAM) were highly consistent but occasionally required a "Refine Edge" pass for fine details like hair or fur.

Temporal Consistency:

AI-driven inpainting using GANs (ProPainter) showed superior background reconstruction. Unlike manual cloning, which often leaves "smudge" artifacts, the AI successfully predicted background textures by analyzing temporal data from preceding and succeeding frames.

6.4 Computational Overhead

The findings show that the AI workflow is highly dependent on GPU VRAM. During the Inpainting phase, VRAM usage peaked at 8.2 GB. This suggests that while the workflow is fast, it requires an initial investment in high-end hardware (NVIDIA RTX series), which is a critical consideration for small-scale VFX houses in India.

VII. LIMITATIONS & FUTURE SCOPE

7.1 Limitations of the Research

Despite the high optimization percentage, certain limitations were observed:

Transparency & Semi-Transparency:

AI models still struggle with semitransparent objects like smoke, glass, or fine motion blur.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Hardware Dependency:
The pipeline's speed is strictly tied to CUDA core count and VRAM capacity.

Complex Occlusions:
When an object is completely hidden for more than 30 frames,the AI occasionally loses the "track," requiring a manual restart of the prompt.

7.2 Future Scope

The future of this research lies in Real-time AI Integration.
Cloud-based Rendering:
Moving the AI pipeline to AWS or Google Cloud to allow multi-user collaboration.
Neural Radiance Fields (NeRF):
Exploring how 3D reconstruction can replace 2D inpainting for static set extensions.
Mobile VFX:
Optimizing these models to run on edge devices, allowing directors to see "pre-visualized" AI results on set in real-time.

VIII. CONCLUSION AND THE REFERENCES

The research successfully demonstrates that integrating Artificial Intelligence into the VFX pipeline is no longer a theoretical concept but a practical necessity for industry optimization. Through the comparative study of manual versus AI-driven workflows, it is evident that tasks such as Rotoscoping and Inpainting can be optimized by over 80%
using modern Deep Learning architectures like SAM and GANs.
This shift represents a move toward "Smart Production." While AI does not yet eliminate the need for skilled artists particularly for high precision tasks involving complex transparency it effectively removes the "production bottleneck" of repetitive labor. This allows for a more scalable, cost-effective, and rapid "Show Setup." Ultimately, the future of VFX lies in a hybrid model where human creativity is augmented by algorithmic efficiency, ensuring that the next generation of digital media is produced with both speed and technical excellence.

REFERENCES

[1]. Kirillov, A., et al. (2023). "Segment Anything." arXiv preprint arXiv:2304.02643. (The foundational paper for SAM).
[2]. Barnes, C., et al. (2009). "PatchMatch: A randomized correspondence algorithm for structural image editing." ACM Transactions on Graphics (ToG).
[3]. Goodfellow, I., et al. (2014). "Generative Adversarial Nets." Advances in Neural Information Processing Systems.
[4]. Zhou, B., et al. (2023). "ProPainter: Improving Propagation and Transformer for Video Inpainting." IEEE International Conference on Computer Vision (ICCV).
[5]. Long, J., Shelhamer, E., & Darrell, T. (2015). "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition.
[6]. Foundry (2022). "Machine Learning in VFX: The Power of CopyCat." Industry White Paper.
[7]. Adobe Research (2023). "Advancements in Roto Brush 3.0: Integrating AI for temporal consistency."
[8]. Vinyals, O., et al. (2019). "Deep Learning for Video Sequence Analysis." Nature Machine Intelligence.