Digital Mystique: Advanced Multi-Modal Approach to X-Men Character Transformation Using ComfyUI

Abstract

This research presents a novel approach to digital character transformation, specifically focusing on the conversion of human actors into the iconic blue-skinned Mystique character from the X-Men franchise. Traditional makeup application for this character is extensively time-consuming, requiring 6-12 hours with a team of six artists. This project explores an innovative solution utilizing advanced artificial intelligence techniques to automate this process through digital means. By leveraging a combination of state-of-the-art models including GPT Image, ICEdit, HiDream E1 for image transformation, and Wan 2.1 Fun Control for video processing, we demonstrate a comprehensive workflow implemented in ComfyUI. The methodology involves precise skin and feature detection, targeted color and texture transformation, and seamless video style transfer. Our results show that this approach significantly reduces the time and resources needed for character transformation while maintaining visual fidelity. This research contributes to the growing field of AI-assisted visual effects in media production, offering practical applications for film and television industries.

1. Introduction

The creation of visually compelling characters in film and television often demands extensive practical effects and makeup. The character of Mystique from the X-Men franchise represents one of the most labor-intensive makeup applications in modern cinema, requiring between 6 and 12 hours of application time by a team of six makeup artists This process not only consumes significant production resources but also places substantial physical and psychological stress on the performing actors, who must remain stationary for extended periods application.

With recent advancements in artificial intelligence and computer vision, there exists an opportunity to revolutionize this process through digital transformation techniques. This research explores the implementation of an automated digital makeup system capable of transforming regular human skin into Mystique's distinctive blue, scaled appearance in both images and videos.

The primary objectives of this research are:

1. To develop an efficient and accurate method for human skin detection and segmentation
2. To implement a digital transformation process that convincingly replicates Mystique's blue skin texture
3. To ensure the transformation maintains the original facial expressions and body movements
4. To extend the image-based transformation to video, preserving temporal consistency

This project is significant for several reasons. First, it demonstrates the practical application of AI in media production workflows. Second, it showcases the integration of multiple specialized AI models to solve a complex visual transformation task. Finally, it offers a potential solution that could significantly reduce production costs and actor discomfort in character creation for film and television.

2. Background Research

2.1 Traditional Makeup for Mystique Character

The traditional process of transforming actors into Mystique involves extensive prosthetics, full-body makeup application, and detailed hand-painting. For Jennifer Lawrence's portrayal in "X-Men: First Class" (2011), the makeup process took approximately 8 hours and required the actor to stand for long periods (Looper, 2018). The physical discomfort was so significant that for subsequent films, the production team altered the character's storyline to reduce the amount of blue-form screen time (Vanity Fair, 2019). This illustrates the very practical problems that digital solutions aim to address.

2.2 Digital Character Transformation

Digital character transformation has evolved significantly over the past decade. Early approaches relied heavily on manual rotoscoping and frame-by-frame adjustments (Failes, 2019). Contemporary methods leverage deep learning to automate much of

this process. Notable examples include DeepFakes for face swapping (Westerlund, 2019) and GAN-based texture transfer techniques (Karras et al., 2020). However, these approaches often lack the precision required for professional media production, particularly in maintaining facial expressions and handling varying lighting conditions.

2.3 AI Models for Image and Video Transformation

Recent developments in AI have produced models specifically designed for image and video manipulation. Multi-modal models such as GPT-4 with Vision, Google Gemini, and specialized models like ICEdit (Zhang et al., 2025) and HiDream E1 allow for language-guided image editing. For video transformation, models such as Wan 2.1 and HunyuanCustom offer capabilities for style transfer while maintaining temporal consistency.

These advancements create new possibilities for digital character transformation that were previously unattainable. This research explores the integration of these cutting-edge models within a practical workflow for Mystique character transformation.

3. Methodology

3.1 System Overview

This project implements a comprehensive pipeline for transforming human subjects into the Mystique character using ComfyUI as the primary framework for model integration and workflow management. The methodology evolved through multiple iterations, with each approach evaluated based on quality of transformation, preservation of identity features, and operational efficiency.

The final system architecture consists of four main components:

1. Human detection and segmentation
2. Image-based skin transformation
3. Composite image generation
4. Video transformation and temporal consistency preservation

3.2 Model Selection Process

The selection of appropriate models for each component was a critical aspect of this research. Initial experiments with inpainting approaches using Flux Fill revealed limitations in mask adherence and texture quality. This led to the exploration of

alternative approaches and the eventual adoption of language-guided image editing models.

After comparative analysis, we identified that:

- GPT Image demonstrated superior capabilities in skin modification while preserving identity features
- ICEdit showed the fastest processing speed among open-source models and excelled in hair and eye color transformation
- HiDream E1 provided high-quality instruction-based image editing but with higher computational requirements
- For video transformation, Wan 2.1 Fun Control emerged as the optimal choice due to its ability to perform style transfer without requiring custom LoRA training

3.3 Skin and Feature Detection

Accurate detection of skin, hair, and facial features is fundamental to the transformation process. We implemented this using specialized nodes in ComfyUI:

- Human Parts Mask Generator node for general body segmentation
- Easy-Use node's Human Segmentation component for precise skin and hair mask generation
- RMBG node's Face Segmentation for initial eye area detection (later supplemented with manual refinement)

The segmentation approach produces separate masks for skin, hair, and eyes, allowing for targeted transformations specific to each feature. This granular control is essential for accurately replicating Mystique's distinctive appearance, where skin becomes blue and scaly, hair remains red, and eyes transform to yellow.

3.4 Image Transformation Workflow

The image transformation process follows these steps:

1. Input image preprocessing and normalization
2. Generation of segmentation masks for skin, hair, and eyes
3. Application of ICEdit for initial transformation using text prompts that specify Mystique's features
4. Secondary transformation using GPT Image for enhanced skin texture and detail
5. Mask-based compositing of transformed elements

6. Post-processing for color harmony and final adjustments

For the transformation prompts, we used specific language describing Mystique's characteristics: "blue scaly skin texture like Mystique from X-Men" for skin areas, "bright red hair like Mystique from X-Men" for hair regions, and "yellow reptilian eyes like Mystique from X-Men" for eye areas.

3.5 Video Transformation Approach

Extending the transformation to video presented additional challenges, particularly in maintaining consistency across frames. After evaluating several options, we selected Wan 2.1 Fun Control as the primary video transformation model due to its ability to perform controlled style transfer without requiring custom training.

The video transformation workflow consists of:

1. Video preprocessing and frame extraction
2. Application of Wan 2.1 Fun Control with reference images of successfully transformed Mystique images
3. Parameter tuning to balance transformation strength and motion preservation
4. Frame recombination and post-processing

We noted that the Wan 2.1 Fun Control model performs optimally with videos featuring moderate movement. For sequences with extreme motion, additional frame-by-frame adjustments may be necessary to maintain consistent transformation quality.

4. Implementation Details

4.1 ComfyUI Implementation

ComfyUI served as the primary framework for implementing the transformation pipeline. This node-based interface allows for flexible integration of various AI models and processing components. The complete workflow was developed through multiple iterations, with the final version featuring optimized node connections and parameter settings.

Key advantages of using ComfyUI for this project include:

Visual programming interface that facilitates workflow development and adjustment

Native support for multiple AI models, including those used in this research

Efficient GPU memory management for processing high-resolution images and videos

Extensibility through custom nodes and Python scripting

4.2 Segmentation Implementation

The segmentation component of our workflow utilizes a combination of automatic and manual techniques:

```
# Pseudocode for segmentation process
def generate_masks(input_image):
    # Generate initial segmentation
    body_mask = human_parts_mask_generator(input_image)

    # Refine for specific features
    skin_mask = easy_use_human_segmentation(input_image, 'skin
    hair_mask = easy_use_human_segmentation(input_image, 'hair

    # Initial eye detection
    eye_mask_initial = rmbg_face_segmentation(input_image, 'ey

    # Manual refinement for eyes if needed
    eye_mask_final = manual_refinement(eye_mask_initial)

    return skin_mask, hair_mask, eye_mask_final
```

For eye mask generation, we found that current automatic segmentation methods often produce inconsistent results. Therefore, we supplemented the automatic process with manual refinement in Adobe Photoshop for optimal precision.

4.3 Model Integration

The integration of multiple AI models required careful orchestration to maintain processing efficiency and result quality. ICEdit was implemented through ComfyUI's native support, while GPT Image required external API integration.

For the ICEdit implementation, we utilized the following configuration:

```
# ICEdit configuration
{
```

```
    "model": "icedit-lora",
    "lora_weight": 0.8,
    "prompt": "Blue scaly skin like Mystique from X-Men, detai
    "negative_prompt": "smooth skin, human skin tone, unrealis
    "guidance_scale": 7.5,
    "steps": 25
}
```

For video transformation, the Wan 2.1 Fun Control model was configured with the following parameters:

```
# Wan 2.1 Fun Control configuration
{
    "model_path": "wan2.1_fun_control_full.safetensors",
    "reference_image": "mystique_reference.png",
    "strength": 0.85,
    "motion_strength": 0.7,
    "frames": 24,
    "fps": 24,
    "seed": 42
}
```

4.4 Composite Generation

The final image composition process combines the transformed elements using the generated masks. This step is crucial for creating a seamless integration of the transformed features with the original image.

```
# Pseudocode for composite generation
def create_composite(original_image, skin_transform, hair_tran
    composite = original_image.copy()

    # Apply transformations using masks
    composite = apply_masked_transformation(composite, skin_tr
    composite = apply_masked_transformation(composite, hair_tr
    composite = apply_masked_transformation(composite, eye_tra

    # Final color grading and adjustments
    composite = post_processing(composite)
```

```
    return composite
```

The final post-processing step includes color harmonization, detail enhancement, and any necessary manual adjustments to ensure the transformation appears natural and consistent.

5. Experimental Results

5.1 Model Comparative Analysis

Our experimental evaluation compared different approaches to the Mystique transformation task. Table 1 summarizes the performance of different models and techniques across key metrics.

Approach	Skin Transformation Quality	Identity Preservation	Processing Speed	Ease of Implementation
Flux Fill (Inpainting)	Low	Medium	Fast	High
ICEdit	Medium	High	Fast	Medium
HiDream E1	High	High	Slow	Medium
GPT Image	Very High	Very High	Medium	Low
Combined (ICEdit + GPT Image)	Very High	Very High	Medium	Low

Table 1: Comparative analysis of different transformation approaches

Initial experiments with inpainting approaches using Flux Fill revealed significant limitations. Despite accurate mask generation, the transformed skin areas frequently failed to match the mask boundaries and produced unconvincing textures. This led to the exploration of language-guided image editing models as an alternative approach.

The combination of ICEdit and GPT Image produced the most compelling results, leveraging ICEdit's efficiency for initial transformation and GPT Image's superior quality for refinement. This hybrid approach balanced quality and processing requirements effectively.

5.2 Video Transformation Evaluation

For video transformation, we evaluated several models based on their ability to maintain temporal consistency while achieving high-quality transformation:

Video Model	Transformation Quality	Temporal Consistency	Motion Handling	Implementation Complexity
Hunyuan	High	Medium	Medium	High (requires LoRA)
Wan 2.1	High	High	Medium	High (requires LoRA)
Wan 2.1 Fun Control	High	High	Good (moderate movement)	Medium (no LoRA required)
HunyuanCustom	Very High	Very High	Very High	Very High (hardware demands)
VACE	Very High	Very High	Very High	Very High (hardware demands)

Table 2: Comparative analysis of video transformation models

Wan 2.1 Fun Control emerged as the optimal choice for this project due to its balance of quality and implementation feasibility. Unlike other models that require custom LoRA training with extensive image datasets, Wan 2.1 Fun Control can perform style transfer directly using reference images. This significantly reduced development time and resource requirements.

The model performed exceptionally well with videos featuring moderate movement, though more extreme motions occasionally resulted in transformation inconsistencies. For production use, these limitations could be addressed through selective application and manual touchups of problematic frames.

5.3 Hardware and Performance Considerations

The hardware requirements for implementing this transformation pipeline vary significantly depending on the models used:

Component	Minimum Requirements	Recommended Requirements	Processing Time (1080p Image)
Image Segmentation	8GB VRAM	12GB VRAM	5-10 seconds
ICEdit Transformation	8GB VRAM	12GB VRAM	10-15 seconds
GPT Image (API)	N/A (Cloud-based)	N/A (Cloud-based)	15-30 seconds
Wan 2.1 Fun Control	16GB VRAM	24GB VRAM	30-60 seconds per frame

Table 3: Hardware requirements and performance metrics

While more advanced models like HunyuanCustom and VACE demonstrate superior capabilities, their substantial hardware requirements (32GB+ VRAM) make them impractical for many production environments. The selected approach balances quality with reasonable hardware demands, making it accessible to a wider range of users.

6. Discussion

6.1 Evaluation of Results

The implemented solution successfully achieves the core objectives of automating Mystique character transformation while maintaining identity features and producing visually compelling results. The combination of specialized segmentation techniques with multi-modal image editing models provides a robust approach to this challenging transformation task.

Key strengths of the implemented solution include:

- High-quality skin transformation with realistic texture and color
- Effective preservation of identity features and expressions
- Reasonable processing times suitable for production environments
- Flexible workflow that can be adapted to different input conditions

However, several limitations remain:

- Manual refinement is still required for optimal eye mask generation

Video transformation quality degrades with extreme movements

The process requires splitting between multiple specialized models rather than a single end-to-end solution

Reliance on external API services (GPT Image) introduces potential workflow dependencies

6.2 Comparison to Traditional Methods

Compared to traditional physical makeup application, the digital approach offers several significant advantages:

Aspect	**Traditional Makeup**	**Digital Transformation**
Time Required	6-12 hours	Minutes to hours (depending on footage length)
Personnel Requirements	Team of 6+ makeup artists	1-2 technical artists
Actor Comfort	High discomfort, limited movement	No additional discomfort
Consistency	Varies between applications	Consistent with controlled parameters
Flexibility for Changes	Requires complete reapplication	Parameters can be adjusted after filming

Table 4: Comparison between traditional and digital transformation methods

The digital approach represents a significant advancement in efficiency and actor comfort, while also providing greater creative flexibility in post-production. However, it should be noted that the digital approach is still evolving and may not match the tactile authenticity of physical makeup in all contexts, particularly for close-up shots where physical interaction with the environment is required.

6.3 Future Directions

The field of AI-based character transformation is rapidly evolving, with several promising developments on the horizon:

Emerging models like VACE and newer versions of HunyuanCustom show significant potential for improving video transformation quality and motion

handling

Integration of real-time transformation capabilities could enable on-set previsualization

Development of specialized fine-tuning techniques for character-specific transformations could improve quality and reduce processing requirements

Exploration of 3D-aware transformation methods could enhance consistency across different viewing angles

Future research should focus on developing more unified approaches that reduce the need for multiple specialized models and manual intervention. Additionally, optimizing these techniques for real-time or near-real-time performance would significantly expand their practical applications in production environments.

7. Conclusion

This research presents a comprehensive approach to digital character transformation, specifically focusing on the conversion of human actors into the iconic Mystique character from X-Men. By leveraging a combination of advanced AI models including ICEdit, GPT Image, and Wan 2.1 Fun Control, we have demonstrated a viable alternative to traditional makeup techniques that significantly reduces time, resource requirements, and actor discomfort.

The implemented solution successfully addresses the key challenges of this transformation task, including accurate skin and feature detection, realistic texture application, and temporal consistency in video sequences. The comparative analysis reveals that while no single model currently excels in all aspects of this task, a carefully orchestrated combination of specialized models can achieve compelling results.

The methodological approach and findings from this research contribute to the growing field of AI-assisted visual effects in media production. The techniques developed here have potential applications beyond the specific case of Mystique, extending to various character transformations and visual effects tasks in film and television production.

As AI models continue to evolve, we anticipate that the quality, efficiency, and accessibility of digital character transformation will improve further, potentially revolutionizing how certain visual effects are approached in media production. This research provides a foundation for these future developments while offering a practical solution to a specific and challenging transformation task.

References

Failes, I. (2019) 'The evolution of digital make-up', *Befores & Afters*, 15 March. Available at: https://beforesandafters.com/2019/03/15/the-evolution-of-digital-make-up/.

Alibaba-PAI (2025) Wan2.1-Fun-1.3B-Control. Available at: https://huggingface.co/alibaba-pai/Wan2.1-Fun-1.3B-Control (Accessed: 21 May 2025). Source code: https://github. com/Wan-Video/Wan2.1

Black Forest Labs (2025) Flux.1: A New Era of Creation [Software]. Available at: https://huggingface.co/black-forest-labs/FLUX.1-dev. Source code: https: //github.com/black-forest-labs/flux

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T. (2020) 'Analyzing and improving the image quality of StyleGAN', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8110-8119.

Westerlund, M. (2019) 'The emergence of deepfake technology: A review', *Technology Innovation Management Review*, 9(11), pp. 40-53.

Zhang, Z., Xie, J., Lu, Y., Yang, Z. and Yang, Y. (2025) In-Context Edit: Enabling Instructional Image Editing with In-Context Generation in Large Scale Diffusion Transformer. Available at: https://arxiv.org/abs/2504.20690 (Accessed: 29 Apr 2025). Source code: https://github.com/River-Zhang/ICEdit

CHEREF-Mehdi (2021) 'SkinDetection: Skin detection using HSV and YCbCr', *GitHub repository*. Available at: https://github.com/CHEREF-Mehdi/SkinDetection (Accessed: 9 May 2025).

ComfyUI (2025) *ComfyUI Documentation: Generate video, images, 3D, audio with AI*. Available at: https://www.comfy.org/ (Accessed: 8 May 2025).

Hu, T., Yu, Z., Zhou, Z., Liang, S., Zhou, Y., Lin, Q. and Lu, Q. (2025) HunyuanCustom: A Multimodal-Driven Architecture for Customized Video Generation. Available at: https://arxiv.org/abs/2505.04512 (Accessed: 8 May 2025). Source code: https://github.com/Tencent/ HunyuanCustom

Kong, W., Tian, Q., Zhang, Z., Min, R., Dai, Z., Zhou, J., Xiong, J., Li, X., Wu, B., Zhang, J. and others (2024) HunyuanVideo: A Systematic Framework For Large Video Generative Models. Available at: https://arxiv.org/abs/2412.03603 (Accessed: 11 Mar 2025). Source code: https://github.com/Tencent-Hunyuan/HunyuanVideo

Wang, Z., Liu, Y., Zhang, J., Tian, Q., Wang, X., Li, X., Dai, Z., Kong, W., Min, R., Wu, B. and others (2025) VACE: A Unified Framework for Video Generation and Editing. Available at: https://arxiv.org/abs/2503.07598). Source code:

https://github. com/ali-vilab/VACE

OpenAI (2025) Introducing 4o Image Generation. Available at: https://openai.com/index/introducing-4o-image-generation/