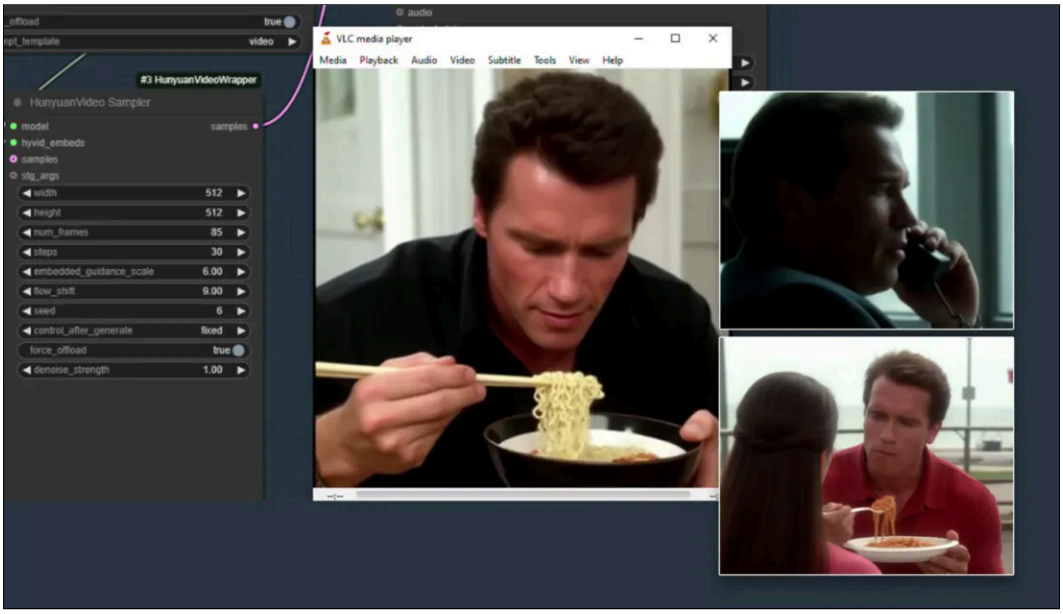
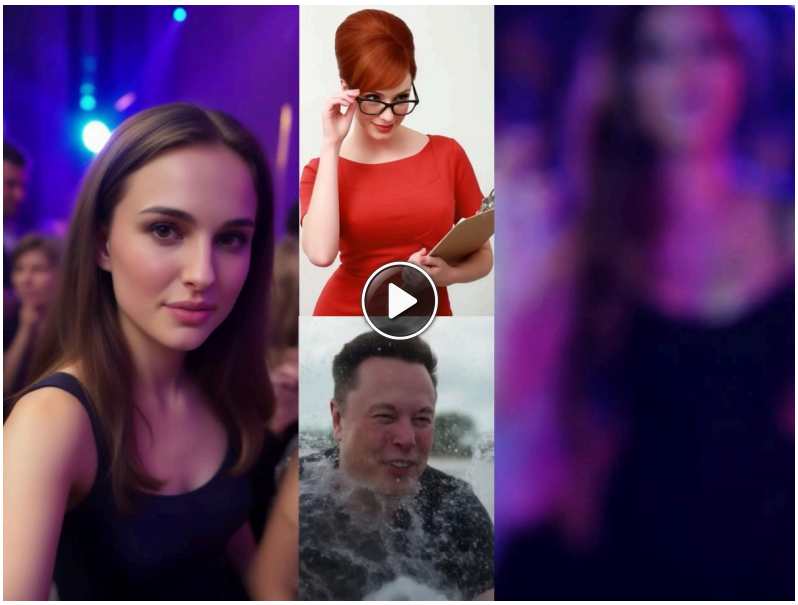
The Rise of Hunyuan Video Deepfakes

Originally published at https://www.unite.ai/the-rise-of-hunyuan-video-deepfakes/



Due to the nature of some of the material discussed here, this article will contain fewer reference links and illustrations than usual.

Something noteworthy is currently happening in the AI synthesis community, though its significance may take a while to become clear. Hobbyists are training generative AI video models to reproduce the likenesses of people, using video-based LoRAs on Tencent's recently released open source Hunyuan Video framework.*



CLICK TO PLAY VIDEO ON SITE

Click to play. *Diverse results from Hunyuan-based LoRA customizations freely available at the Civit community. By training low-rank adaptation models (LoRAs), issues with temporal stability, which have plagued AI video generation for two years, are significantly reduced.* Sources: civit.ai

In the video shown above, the likenesses of actresses Natalie Portman, Christina Hendricks and Scarlett Johansson, together with tech leader Elon Musk, have been trained into relatively small add-on files for the Hunyuan generative video system, which can be installed without content filters (such as NSFW filters) on a user's computer.

The creator of the Christina Hendricks LoRA shown above states that only 16 images from the *Mad Men* TV show were needed to develop the model (which is a mere 307mb download); multiple posts from the Stable Diffusion community at Reddit and Discord confirm that LoRAs of this kind do not require high amounts of training data, or high training times, in most cases.

CLICK TO PLAY VIDEO ON SITE

Click to play. *Arnold Schwarzenegger is brought to life in a Hunyuan video LoRA that can be downloaded at Civit. See https://www.youtube.com/watch?v=1D7B9g9rY68 for further Arnie examples, from AI enthusiast Bob Doyle.*

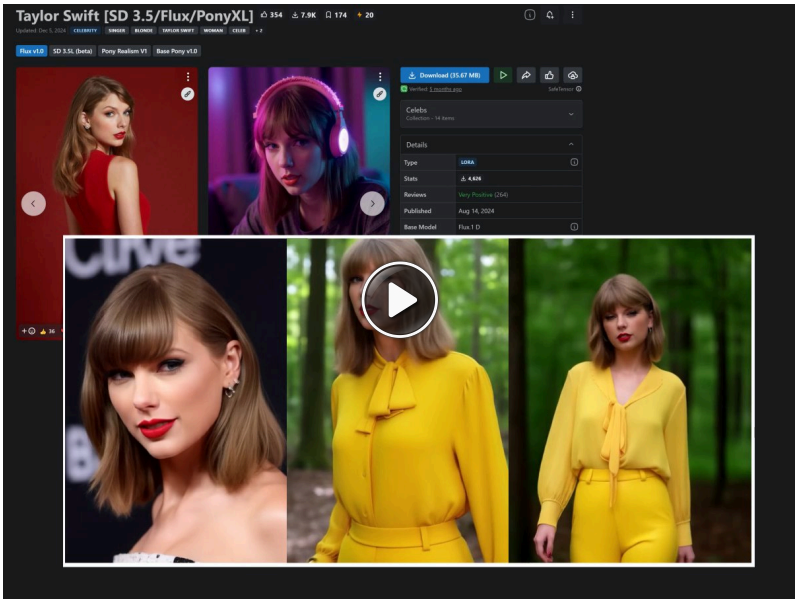
Hunyuan LoRAs can be trained on either static images or videos, though training on videos requires greater hardware resources and increased training time.

The Hunyuan Video model features 13 billion parameters, exceeding Sora's 12 billion parameters, and far exceeding the less-capable Hunyuan-DiT model released to open source in summer of 2024, which has only 1.5 billion parameters.

As was the case two and a half years ago with Stable Diffusion and LoRA (see examples of Stable Diffusion 1.5's 'native' celebrities here), the foundation model in question has a far more limited understanding of celebrity personalities, compared to the level of fidelity that can be obtained through 'ID-injected' LoRA implementations.

Effectively, a customized, personality-focused LoRA gets a 'free ride' on the significant synthesis capabilities of the base Hunyuan model, offering a notably more effective human synthesis than can be obtained either by 2017-era autoencoder deepfakes or by attempting to add movement to static images via systems such as the feted LivePortrait.

All the LoRAs depicted here can be downloaded freely from the highly popular Civit community, while the more abundant number of older custom-made 'static-image' LoRAs can also potentially create 'seed' images for the video creation process (i.e., image-to-video, a pending release for Hunyuan Video, though workarounds are possible, for the moment).



CLICK TO PLAY VIDEO ON SITE

Click to play. *Above, samples from a 'static' Flux LoRA; below, examples from a Hunyuan video LoRA featuring musician Taylor Swift. Both of these LoRAs are freely available at the Civit community.*

As I write, the Civit website offers 128 search results for 'Hunyuan'*. Nearly all of these are in some way NSFW models; 22 depict celebrities; 18 are designed to facilitate the generation of hardcore pornography; and only seven of them depict men rather than women.

So What's New?

Due to the evolving nature of the term *deepfake*, and limited public understanding of the (quite severe) limitations of AI human video synthesis frameworks to date, the significance of the Hunyuan LoRA is not easy to understand for a person casually following the generative AI scene. Let's review some of the key differences between Hunyuan LoRAs and prior approaches to identity-based AI video generation.

1: Unfettered Local Installation

The most important aspect of Hunyuan Video is the fact that it can be downloaded locally, and that it puts a very powerful and *uncensored* AI video generation system in the hands of the casual user, as well as the VFX community (to the extent that licenses may allow across geographical regions).

The last time this happened was the advent of the release to open source of the Stability.ai Stable Diffusion model in the summer of 2022. At that time, OpenAI's DALL-E2 had captured the public imagination, though DALLE-2 was a paid service with notable restrictions (which grew over time).

When Stable Diffusion became available, and Low-Rank Adaptation then made it possible to generate images of the identity of *any* person (celebrity or not), the huge locus of developer and consumer interest helped Stable Diffusion to eclipse the popularity of DALLE-2; though the latter was a more capable system out-of-the-box, its censorship routines were seen as onerous by many of its users, and customization was not possible.

Arguably, the same scenario now applies between Sora and Hunyuan – or, more accurately, between *Sora-grade* proprietary generative video systems, and open source rivals, of which Hunyuan is the first – but probably not the last (here, consider that Flux would eventually gain significant ground on Stable Diffusion).

Users who wish to create Hunyuan LoRA output, but who lack effectively beefy equipment, can, as ever, offload the GPU aspect of training to online compute services such as RunPod. This is not the same as creating AI videos at platforms such as Kaiber or Kling, since there is no semantic or image-based filtering (censoring) entailed in renting an online GPU to support an otherwise local workflow.

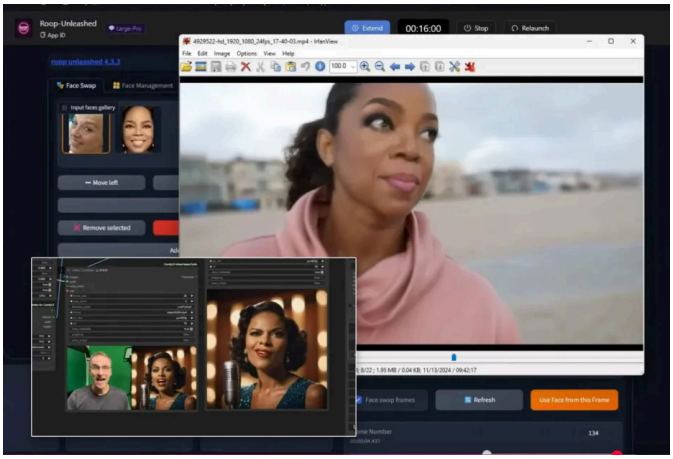
2: No Need for 'Host' Videos and High Effort

When deepfakes burst onto the scene at the end of 2017, the anonymously-posted code would evolve into the mainstream forks DeepFaceLab and FaceSwap (as well as the DeepFaceLive real-time deepfaking system).

This method required the painstaking curation of thousands of face images of each identity to be swapped; the less effort put into this stage, the less effective the model would be. Additionally, training times varied between 2-14 days, depending on available hardware, stressing even capable systems in the long term.

When the model was finally ready, it could only impose faces into existing video, and usually needed a 'target' (i.e., real) identity that was close in appearance to the superimposed identity.

More recently, ROOP, LivePortrait and numerous similar frameworks have provided similar functionality with far less effort, and often with superior results – but with no capacity to generate accurate full-body deepfakes – or any element other than faces.



Examples of ROOP Unleashed and LivePortrait (inset lower left), from Bob Doyle's content stream at YouTube. Sources: https://www.youtube.com/watch?v=i39xeYPBAAM and https://www.youtube.com/watch?v=QGatEItg2Ns

By contrast, Hunyuan LoRAs (and the similar systems that will inevitably follow) allow for unfettered creation of entire worlds, including full-body simulation of the user-trained LoRA identity.

3: Massively Improved Temporal Consistency

Temporal consistency has been the Holy Grail of diffusion video for several years now. The use of a LoRA, together with apposite prompts, gives a Hunyuan video generation a constant identity reference to adhere to. In theory (these are early days), one could train multiple LoRAs of a particular identity, each wearing specific clothing.

Under those auspices, the clothing too is less likely to 'mutate' throughout the course of a video generation (since the generative system bases the next frame on a very limited window of prior frames).

(Alternatively, as with image-based LoRA systems, one can simply apply multiple LoRAs, such as identity + costume LoRAs, to a single video generation)

4: Access to the 'Human Experiment'

As I recently observed, the proprietary and FAANG-level generative AI sector now appears to be so wary of potential criticism relating to the human synthesis capabilities of its projects, that actual *people* rarely appear in project pages for major announcements and releases. Instead, related publicity literature increasingly tends to show 'cute' and otherwise 'non-threatening' subjects in synthesized results.

With the advent of Hunyuan LoRAs, for the first time, the community has an opportunity to push the boundaries of LDM-based human video synthesis in a highly capable (rather than marginal) system, and to fully explore the subject that most interests the majority of us – people.

Implications

Since a search for 'Hunyuan' at the Civit community mostly shows celebrity LoRAs and 'hardcore' LoRAs, the central implication of the advent of Hunyuan LoRAs is that they will be used to create AI pornographic (or otherwise defamatory) videos of real people – celebs and unknowns alike.

For compliance purposes, the hobbyists who create Hunyuan LoRAs and who experiment with them on diverse Discord servers are careful to prohibit examples of real people from being posted. The reality is that even *image*-based deepfakes are now severely weaponized; and the prospect of adding truly realistic videos into the mix may finally justify the heightened fears that have been recurrent in the media over the last seven years, and which have prompted new regulations.

The Driving Force

As ever, porn remains the driving force for technology. Whatever our opinion of such usage, this relentless engine of impetus drives advances in the state-of-the-art that can ultimately benefit more mainstream adoption.

In this case, it is possible that the price will be higher than usual, since the open-sourcing of hyper-realistic video creation has obvious implications for criminal, political and ethical misuse.

One Reddit group (which I will not name here) dedicated to AI generation of NSFW video content has an associated, open Discord server where users are refining ComfyUI workflows for Hunyuan-based video porn generation. Daily, users post examples of NSFW clips – many of which can reasonably be termed 'extreme', or at least straining the restrictions stated in forum rules.

This community also maintains a substantial and well-developed GitHub repository featuring tools that can download and process pornographic videos, to provide training data for new models.

Since the most popular LoRA trainer, Kohya-ss, now supports Hunyuan LoRA training, the barriers to entry for unbounded generative video training are lowering daily, along with the hardware requirements for Hunyuan training and video generation.

The crucial aspect of dedicated training schemes for porn-based AI (rather than *identity*-based models, such as celebrities) is that a standard foundation model like Hunyuan is not specifically trained on NSFW output, and may therefore either perform poorly when asked to generate NSFW content, or fail to disentangle learned concepts and associations in a performative or convincing manner.

By developing fine-tuned NSFW foundation models and LoRAs, it will be increasingly possible to project trained identities into a dedicated 'porn' video domain; after all, this is only the video version of something that has already occurred for still images over the last two and a half years.

VFX

The huge increase in temporal consistency that Hunyuan Video LoRAs offer is an obvious boon to the AI visual effects industry, which leans very heavily on adapting open source software.

Though a Hunyuan Video LoRA approach generates an entire frame and environment, VFX companies have almost certainly begun to experiment with isolating the temporally-consistent human faces that can be obtained by this method, in order to superimpose or integrate faces into real-world source footage.

Like the hobbyist community, VFX companies must wait for Hunyuan Video's image-to-video and video-to-video functionality, which is potentially the most useful bridge between LoRA-driven, ID-based 'deepfake' content; or else improvise, and use the interval to probe the outer capabilities of the framework and of potential adaptations, and even proprietary in-house forks of Hunyuan Video.

Though the license terms for Hunyuan Video technically allow the depiction of real individuals so long as permission is given, they prohibit its use in the EU, United Kingdom, and in South Korea. On the 'stays in Vegas' principle, this does not necessarily mean that Hunyuan Video will not be used in these regions; however, the prospect of external data audits, to enforce a growing regulations around generative AI, could make such illicit usage risky.

One other potentially ambiguous area of the license terms states:

> *'If, on the Tencent Hunyuan version release date, the monthly active users of all products or services made available by or for Licensee is greater than 100 million monthly active users in the preceding calendar month, You must request a license from Tencent, which Tencent may grant to You in its sole discretion, and You are not authorized to exercise any of the rights under this Agreement unless or until Tencent otherwise expressly grants You such rights.'*

This clause is clearly aimed at the multitude of companies that are likely to 'middleman' Hunyuan Video for a relatively tech-illiterate body of users, and who will be required to cut Tencent into the action, above a certain ceiling of users.

Whether or not the broad phrasing could also cover *indirect* usage (i.e., via the provision of Hunyuan-enabled visual effects output in popular movies and TV) may need clarification.

Conclusion

Since deepfake video has existed for a long time, it would be easy to underestimate the significance of Hunyuan Video LoRA as an approach to identity synthesis, and deepfaking; and to assume that the developments currently manifesting at the Civit community, and at related Discords and subreddits, represent a mere incremental nudge towards truly controllable human video synthesis.

More likely is that the current efforts represent only a fraction of Hunyuan Video's potential to create completely convincing full-body and full-environment deepfakes; once the image-to-video component is released (rumored to be occurring this month), a far more granular level of generative power will become available to both the hobbyist and professional communities.

When Stability.ai released Stable Diffusion in 2022, many observers could not determine why the company would just give away what was, at the time, such a valuable and powerful generative system. With Hunyuan Video, the profit motive is built directly into the license – albeit that it may prove difficult for Tencent to determine when a company triggers the profit-sharing scheme.

In any case, the result is the same as it was in 2022: dedicated development communities have formed immediately and with intense fervor around the release. Some of the roads that these efforts will take in the next 12 months are surely set to prompt new headlines.

** Up to 136 by the time of publication.*

First published Tuesday, January 7, 2025

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More