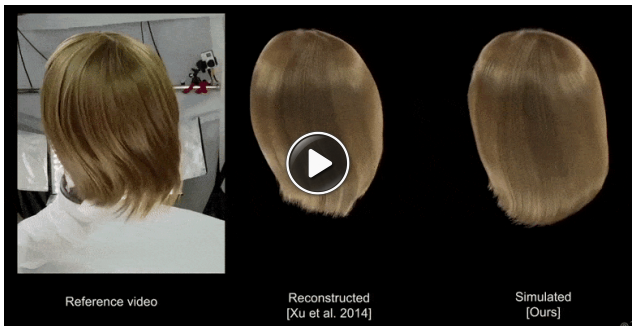
# Tackling ‘Bad Hair Days’ in Human Image Synthesis

Originally published at https://www.unite.ai/tackling-bad-hair-days-in-human-image-synthesis/


Since the golden age of Roman statuary, depicting human hair has been a thorny challenge. The average human head contains 100,000 strands, has varying refractive indices according to its color, and, beyond a certain length, will move and reform in ways that can only be simulated by complex physics models – to date, only applicable through ‘traditional’ CGI methodologies.




CLICK TO VIEW ANIMATED GIF

*From 2017 research by Disney, a physics-based model attempts to apply realistic movement to a fluid hair style in a CGI workflow.* Source: https://www.youtube.com/watch?v=-6iF3mufDW0

The problem is poorly addressed by modern popular deepfakes methods. For some years, the leading package DeepFaceLab has had a ‘full head’ model which can only capture rigid embodiments of short (usually male) hairstyles; and recently DFL stablemate FaceSwap (both packages are derived from the controversial 2017 DeepFakes source code) has offered an implementation of the BiseNet semantic segmentation model, allowing a user to include ears and hair in deepfake output.

Even when depicting very short hairstyles, the results tend to be very limited in quality, with full heads appearing superimposed on footage, rather than integrated into it.

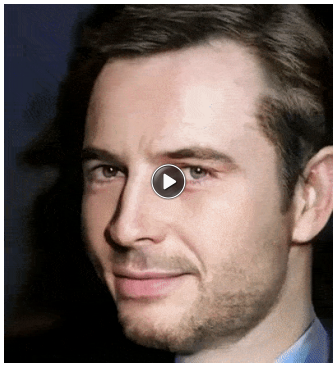
## GAN Hair

The two major competing approaches to human simulation are Neural Radiance Fields (NeRF), which can capture a scene from multiple viewpoints and encapsulate a 3D representation of these viewpoints in an explorable neural network; and Generative Adversarial Networks (GANs), which are notably more advanced in terms of human image synthesis (not least because NeRF only emerged in 2020).

NeRF’s inferred understanding of 3D geometry enables it to replicate a scene with great fidelity and consistency, even if it currently has little or no scope for the imposition of physics models – and, in fact, relatively limited scope for any kind of transformation on the gathered data that does not relate to changing the camera viewpoint. Currently, NeRF has very limited capabilities in terms of reproducing human hair movement.

GAN-based equivalents to NeRF start at an almost fatal disadvantage, since, unlike NeRF, the latent space of a GAN does not natively incorporate an understanding of 3D information. Therefore 3D-aware GAN facial image synthesis has become a hot pursuit in image generation research in recent years, with 2019’s InterFaceGAN one of the leading breakthroughs.

However, even InterFaceGAN's showcased and cherry-picked results demonstrate that neural hair consistency remains a tough challenge in terms of temporal consistency, for potential VFX workflows:



CLICK TO VIEW ANIMATED GIF

*'Sizzling' hair in a pose transformation from InterFaceGAN.*
Source: https://www.youtube.com/watch?v=uoftpl3Bj6w

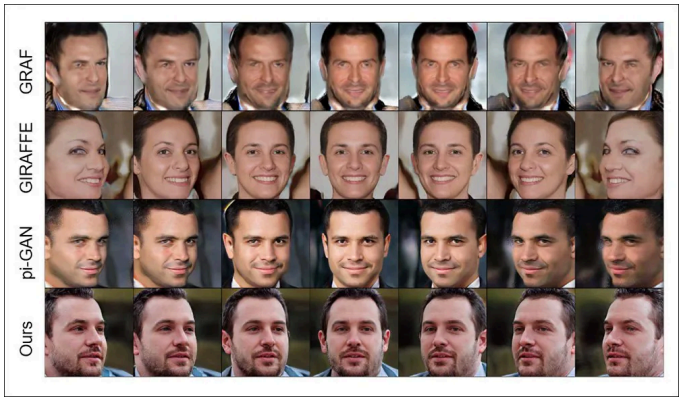
As it becomes more evident that consistent view generation via manipulation of the latent space alone may be an alchemy-like pursuit, an increasing number of papers are emerging that incorporate CGI-based 3D information into a GAN workflow as a stabilizing and normalizing constraint.

The CGI element may be represented by intermediate 3D primitives such as a Skinned Multi-Person Linear Model (SMPL), or by adopting 3D inference techniques in a manner similar to NeRF, where geometry is evaluated from the source images or video.

One new work along these lines, released this week, is *Multi-View Consistent Generative Adversarial Networks for 3D-aware Image Synthesis* (MVCGAN), a collaboration between ReLER, AAII, University of Technology Sydney, the DAMO Academy at Alibaba Group, and Zhejiang University.

*Plausible and robust novel facial poses generated by MVCGAN on images derived from the CELEBA-HQ dataset.* Source: https://arxiv.org/pdf/2204.06307.pdf

MVCGAN incorporates a generative radiance field network (GRAF) capable of providing geometric constraints in a Generative Adversarial Network, arguably achieving some of the most authentic posing capabilities of any similar GAN-based approach.




*Comparison between MVCGAN and prior methods GRAF, GIRAFFE, and pi-GAN.*

However, supplementary material for MVCGAN reveals that obtaining hair volume, disposition, placement and behavior consistency is a problem that's not easily tackled through constraints based on externally-imposed 3D geometry.

CLICK TO VIEW ANIMATED GIF

*From supplementary material not publicly released at the time of writing, we see that while facial pose synthesis from MVCGAN represents a notable advance on the current state of the art, temporal hair consistency remains a problem.*

Since 'straightforward' CGI workflows still find temporal hair reconstruction such a challenge, there's no reason to believe that conventional geometry-based approaches of this nature are going to bring consistent hair synthesis to the latent space anytime soon.

## Stabilizing Hair with Convolutional Neural Networks

However, a forthcoming paper from three researchers at the Chalmers Institute of Technology in Sweden may offer an additional advance in neural hair simulation.

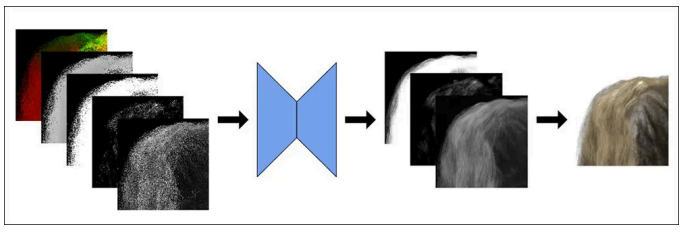

CLICK TO VIEW ANIMATED GIF

*On the left, the CNN-stabilized hair representation, on the right, the ground truth. See video embedded at end of article for better resolution and additional examples.* Source: https://www.youtube.com/watch?v=AvnJkwCmsT4

Titled *Real-Time Hair Filtering with Convolutional Neural Networks*, the paper will be published for the i3D symposium in early May.

The system comprises an autoencoder-based network capable of evaluating hair resolution, including self-shadowing and taking account of hair thickness, in real time, based on a limited number of stochastic samples seeded by OpenGL geometry.

The approach renders a limited number of samples with stochastic transparency and then trains a U-net to reconstruct the original image.



*Under MVCGAN, a CNN filters stochastically sampled color factors, highlights, tangents, depth and alphas, assembling the synthesized results into a composite image.*

The network is trained on PyTorch, converging over a period of six to twelve hours, depending on network volume and the number of input features. The trained parameters (weights) are then used in the real-time implementation of the system.

Training data is generated by rendering several hundred images for straight and wavy hairstyles, using random distances and poses, as well as diverse lighting conditions.

*Various examples of training input.*

Hair translucency across the samples is averaged from images rendered with stochastic transparency at supersampled resolution. The original high resolution data is downsampled to accommodate network and hardware limits, and later upsampled, in a typical autoencoder workflow.

The real-time inference application (the 'live' software that leverages the algorithm derived from the trained model) employs a mix of NVIDIA CUDA with cuDNN and OpenGL. The initial input features are dumped into OpenGL multisampled color buffers, and the result shunted to cuDNN tensors before processing in the CNN. Those tensors are then copied back to a 'live' OpenGL texture for imposition into the final image.

The real-time system operates on a NVIDIA RTX 2080, producing a resolution of 1024×1024 pixels.

Since hair color values are entirely disentangled in the final values obtained by the network, changing the hair color is a trivial task, though effects such as gradients and streaks remain a future challenge.

The authors have released the code used in the paper's evaluations at GitLab. Check out the supplementary video for MVCGAN below.

Real-Time Hair Filtering with Convolutional Neural Networks


## Conclusion

Navigating a the latent space of an autoencoder or GAN is still more akin to sailing than precision driving. Only in this very recent period are we beginning to see credible results for pose generation of 'simpler' geometry such as faces, in approaches such as NeRF, GANs, and non-deepfake (2017) autoencoder frameworks.

The significant architectural complexity of human hair, combined with the need to incorporate physics models and other traits for which current image synthesis approaches have no provision, indicates that hair synthesis is unlikely to remain an integrated component in general facial synthesis, but is going to require dedicated and separate networks of some sophistication – even if such networks may eventually become incorporated into wider and more complex facial synthesis frameworks.

*First published 15th April 2022.*

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**