The Limited Future of Deepfakes

By Martin Anderson

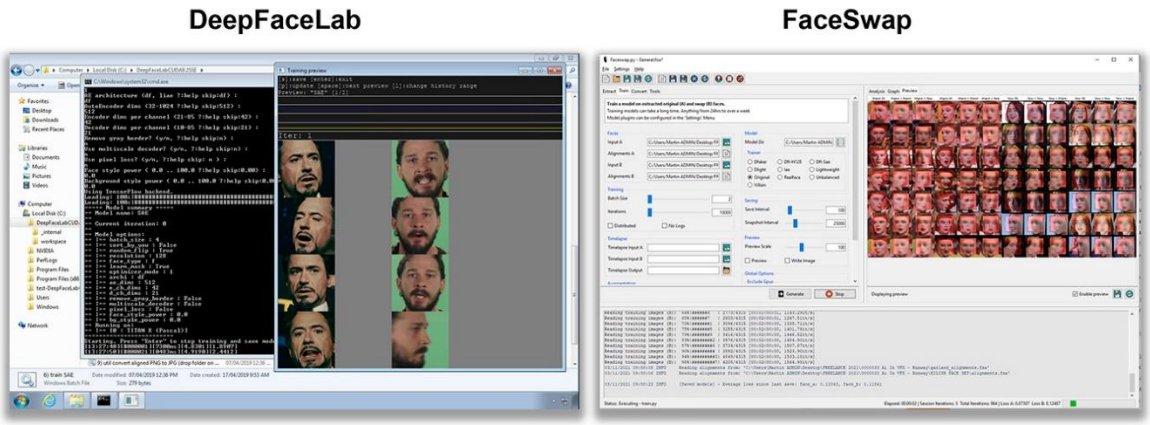
First published March 10th, 2021 at:
https://rossdawson.com/futurist/implications-of-ai/the-limited-future-of-deepfakes/
Web-archived version

The last time I reviewed the state of the art in AI VFX, deepfakes technology was barely a year into existence. The phenomenon of AI-powered face-swapping seemed set to corrupt, deceive and threaten society in terrible ways and across multiple sectors. It was predicted, even at state level, that deepfake output would eventually become indistinguishable from reality.

At a few years' sober distance, it could now be argued that deepfake videos– at least as far as the term refers to the porn-centric DeepFaceLab[1] and the slightly more sober FaceSwap project[2] – are set for as truncated an entry in the history of image synthesis as the fax machine holds in the history of communications: they're showing the way, but they're not fit to finish the journey. Hyper-realistic face swapping may well be coming – but it's probably coming from elsewhere.

Off-The-Shelf Fantasies

That's not to say that the consumer-facing architectures of DFL and FaceSwap couldn't be deconstructed, adapted and re-tooled for professional VFX purposes, or their core technologies re-imagined in more capable work-flows; but rather that these distributions (both of which are forks of the original and now-abandoned 2017 GitHub code) are off-the-shelf solutions aimed at hobbyists with access to gaming-level, rather than production-level GPU and pipeline resources; and that for these and other reasons (as we'll see), the core code cannot meet the general public expectation that deepfakes' quality will improve exponentially in time.



DeepFaceLab (left) seems to currently be the popular choice in production environments, though it features a wizard-driven approach via BAT files. By contrast, FaceSwap (right) has an integrated GUI.

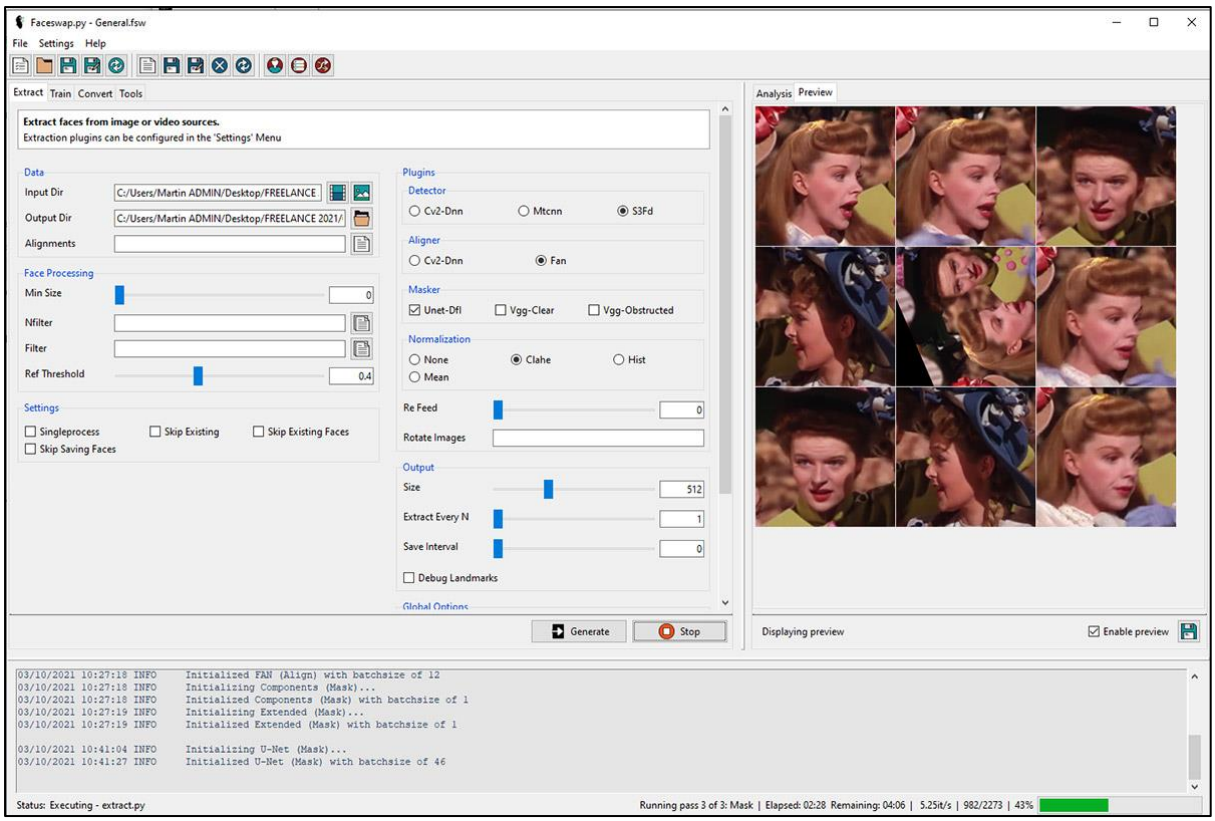
Later, we'll hear from a leading VFX industry researcher about some of the technical bottlenecks and architectural constraints that prevent current deepfake approaches from achieving greater realism. We'll also take a look at some initiatives that seek to extend or re-imagine new approaches to face simulation.

First, to understand the limitations of deepfake technology, we need to have a basic understanding of the process – so let's take a look at that.

The Deepfake Process

A deepfake from March 2021 by TikTok user NextFace, which inserts modern singer/songwriter Billie Eilish into the 1944 Judy Garland musical 'Meet Me In St. Louis'
(https://www.tiktok.com/@nextface/video/6936881472313806085)

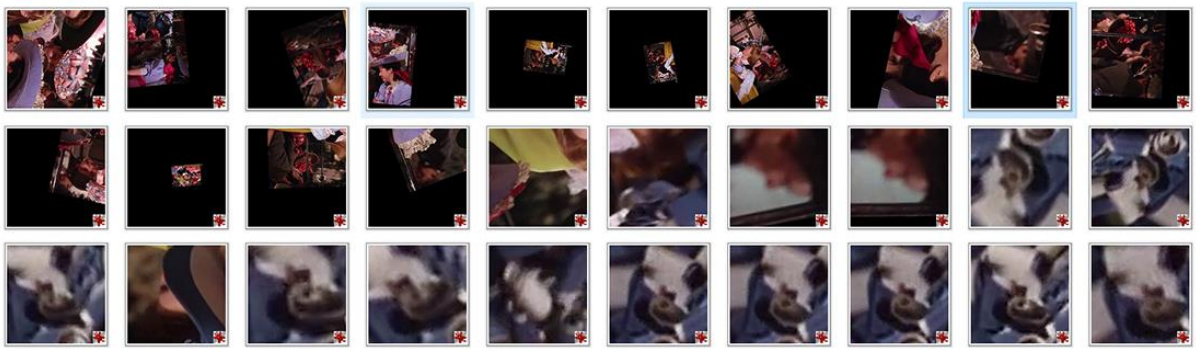
In a deepfakes workflow[3], a face-set is extracted from video clips for both the source subject (the person you wish to replace) and the target subject (the person you wish to eventually appear substituted into the video footage). The average number of images in the resulting face-set is 5-10,000.



The deepfake software, in this case FaceSwap, traverses video clips looking for faces – any faces – and saves those it finds into a face-set.

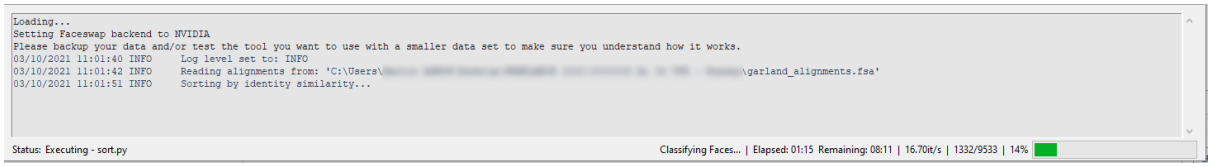
Though the software can be guided to look for a particular face based on a reference image, this functionality does not work well enough to be useful. Instead the user must weed out not only the non-target faces found by the facial recognition algorithm, but also the inevitable false positives that the program dumps into the dataset.

The clip from the 1944 Judy Garland movie contains many people, and here the software has dumped them all into the face-set indiscriminately.



The facial identification algorithm can be set to varying levels of sensitivity, but in nearly all cases it gets fooled by non-human patterns, textures and shapes, leading to many total mismatches during the extraction process.

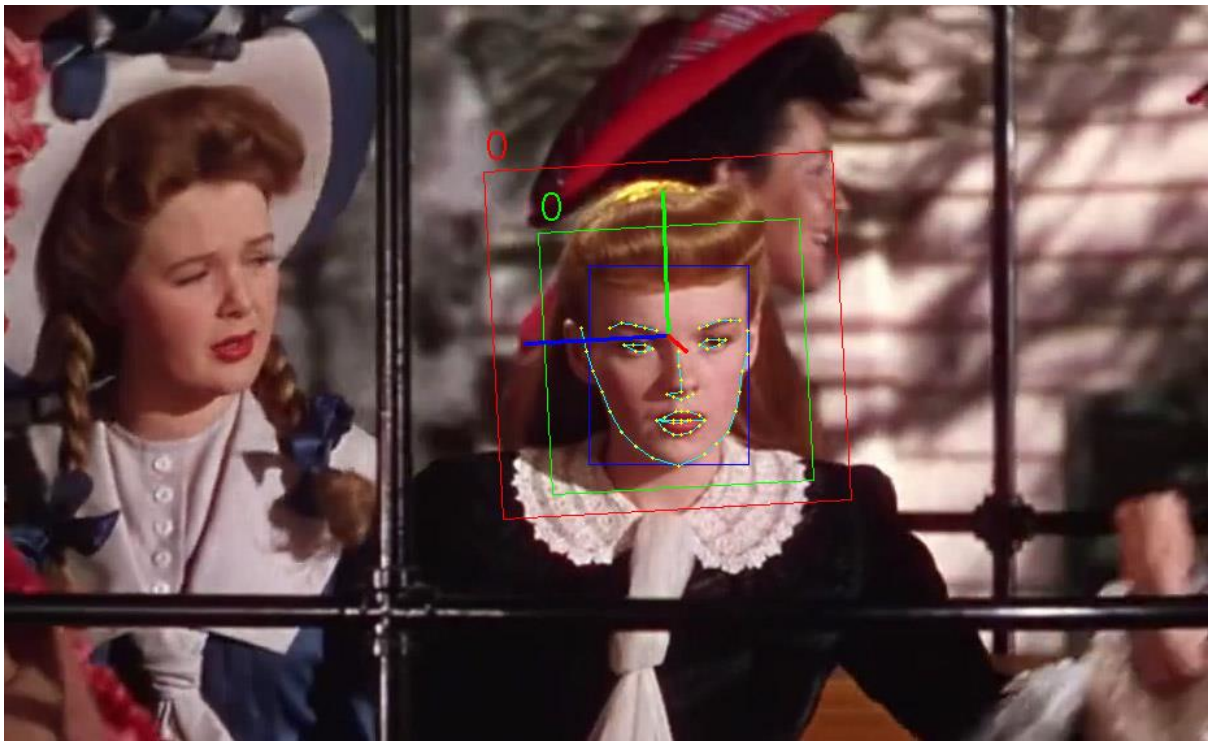
The software can sort the results by face, testing the similarity of the facial landmarks until the face candidates are grouped quite effectively, enabling the quick deletion of non-relevant matches.



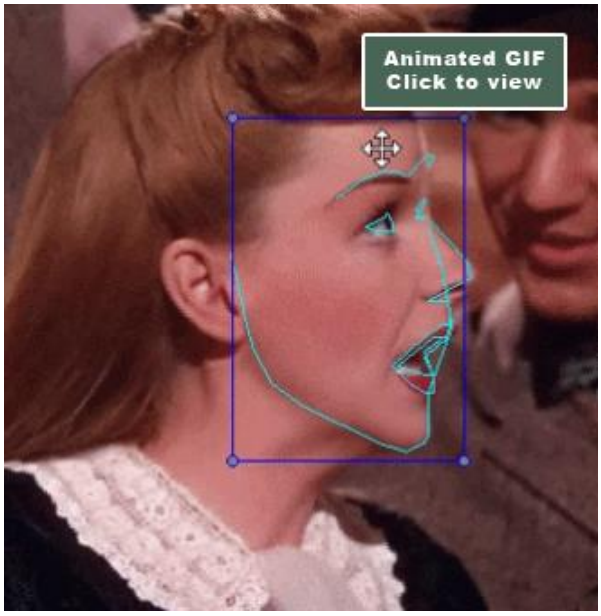
Python sorts the jumble of faces by identity similarity.

With the faces sorted by identity, it's easier to remove non-relevant matches.

The program uses the FAN facial landmark detection system[4] developed by research scientist Adrian Bulat[4].

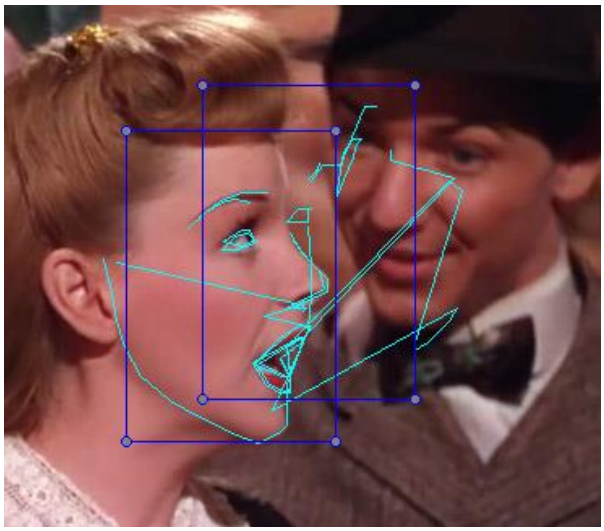


The alignments will frequently need adjusting for profile shots, or for acute angles.



FAN align usually performs poorly on profile shots and acute angles. Therefore it's necessary to intervene on all affected frames manually, which can be labor-intensive for longer clips.

Nearby faces are also an issue:



When FAN align identifies two faces close to each other, it frequently 'hedges its bets' about whether the mouth of one might belong to one face or another, leading to grotesque alignments.

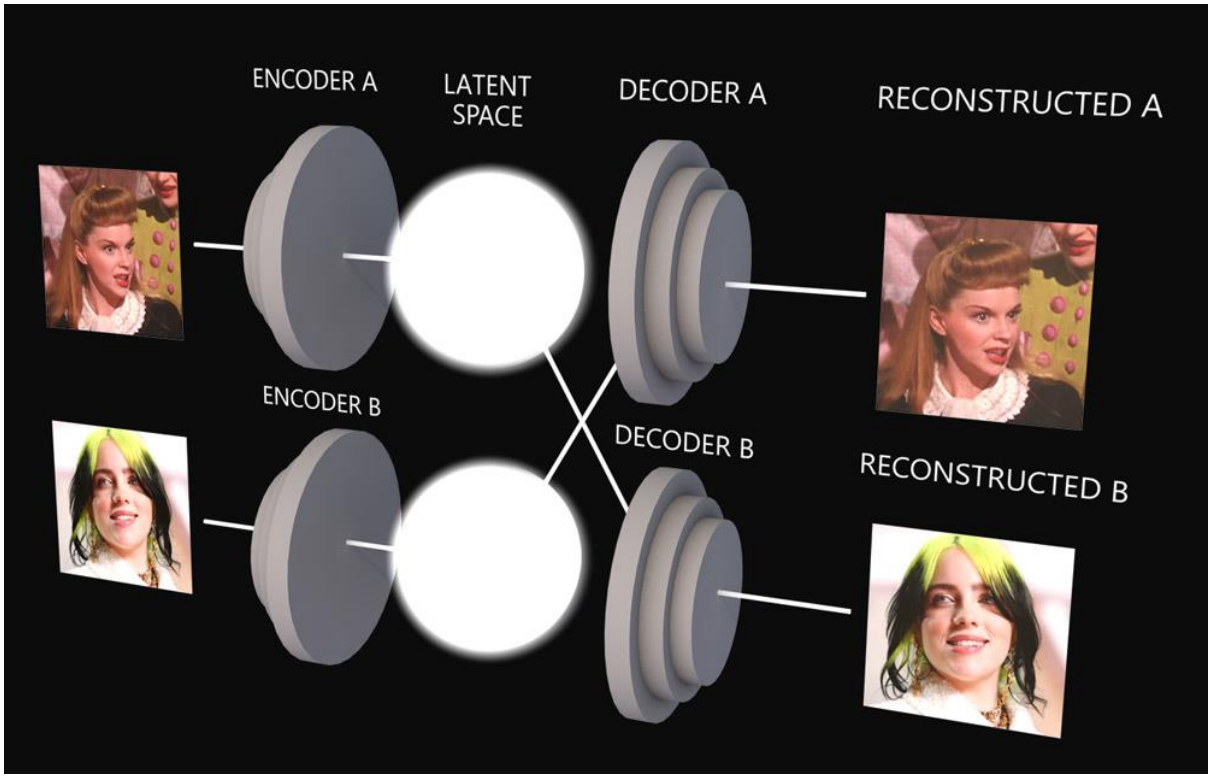
The masking feature follows the path of the alignment landmarks, but does not account for obstructing objects. Therefore masking/occlusion algorithms have been manually trained over time by the FaceSwap community and integrated into the package (whereas DFL features trainable masks, which are effective, but much less user-friendly).

Where the occlusion masking fails, manual clean-up of each frame is necessary.

After this, the curated source and target face-sets are trained on an encoder/decoder[5] (not GAN[6]) model.

During the training, each face-set's pixels are abstracted into mathematical parameters and facets in the 'latent space' of a shared autoencoder, which proceeds, at length — sometimes for weeks — to take apart the facial identities and learn how to reconstruct them[7], constantly grading its own success on the latest attempt, and learning from the scores (i.e. the estimated loss rates) that it assigns to the results of these attempts.

Unlike the shared autoencoder space, the decoder units (on the right in the image below) are dedicated to only *one* of the two specific identities; but during training the decoders have inherited the learned transformational mappings of both identities, and can be rewired to use information from identity A in order to generate images of identity B.



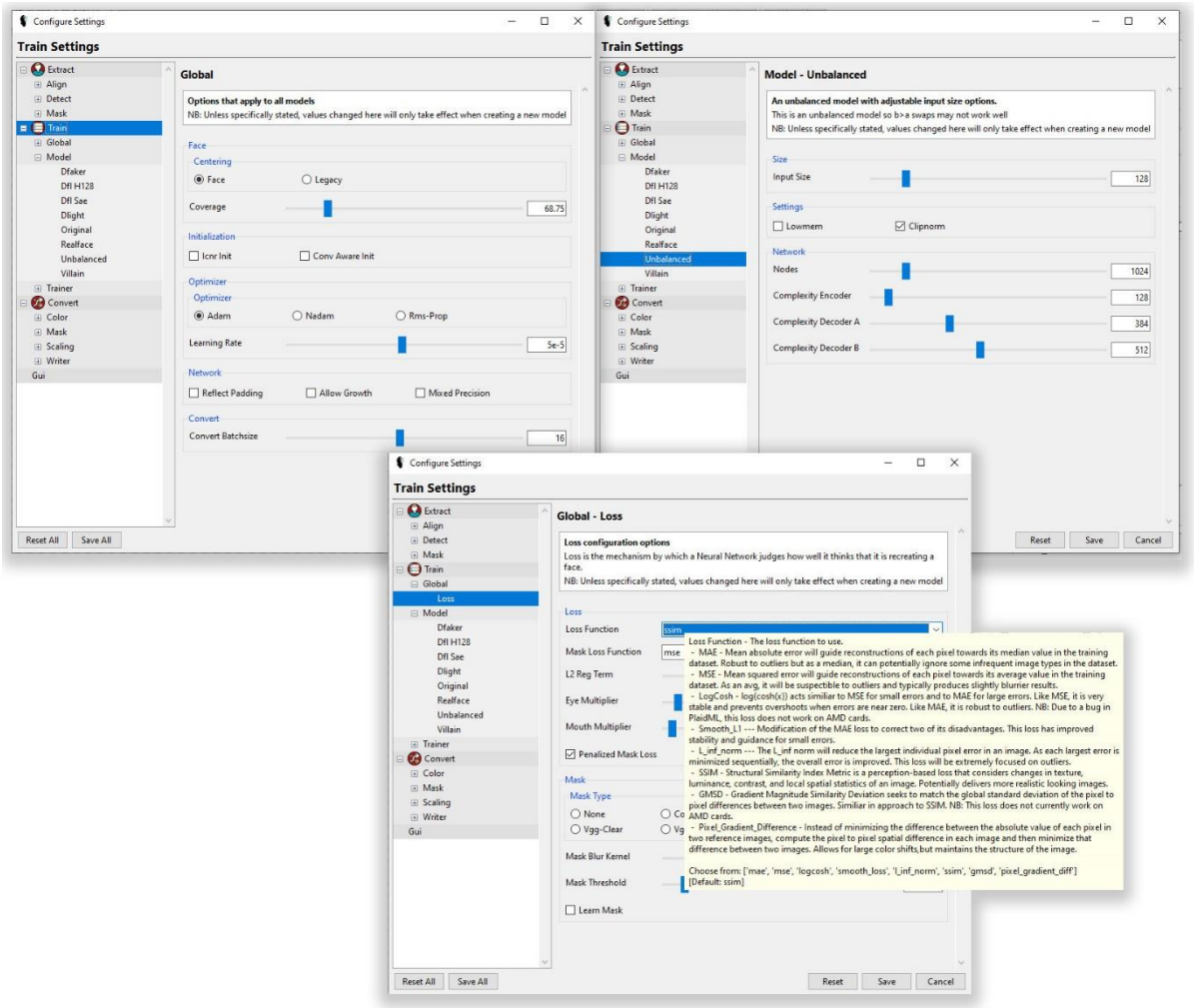
Note the way that the identity information can 'cross over', to the right of the latent space.

The initial ten minutes of training, as the model first learns the basic lineaments of the two subjects, and begins the long task of calculating better computed paths between the features of each subject.

Though all of the available model variations for FOSS deepfake software are derived (however distantly) from the original 2017 code[8], some model types produce higher resolution output, but require more available GPU VRAM.

It is possible to use more than one GPU simultaneously for training, but with diminishing returns, because, according to the developers[9], the only benefit of parallel multi-GPU training is increased batch sizes (the number of images processed at any one time in the latent space of the model) – and at very high batch sizes, the developers contend, the model becomes worse at learning details, sabotaging the intent.

Some of the more sophisticated models have dozens of potential parameters, but experimenting with them is time-consuming, and the results will likely not be consistent across different datasets and projects.



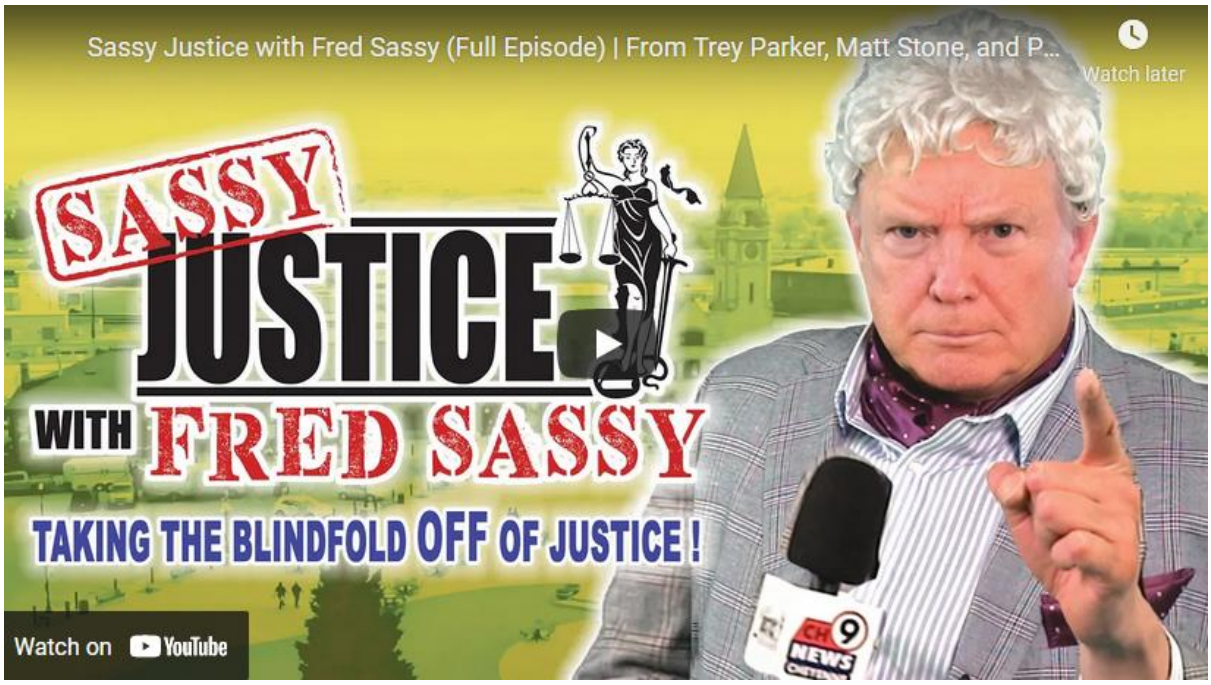
Minor changes in any of the settings can add quality, reduce training time — or result in days of lost effort, since experimentation is so time-consuming, and good results from one setting may not be transferable to the demands of different face-sets in

As the 'predicted accuracy' loss rates descend during training, the swaps become more authentic. However, loss rates don't directly signify success, and the user must evaluate for themselves, from the quality of the previews and through periodic test-swaps, whether the model is sufficiently trained.

At that point, the trained algorithms and the alignments and masks that were produced at the start can be used, at last, to superimpose a realistic 'target' face onto the source video clip (see earlier image of the Garland/Eilish swap).

Deepfakes Go to Hollywood

The most common haunts for non-porn deepfakes are ads[10], art projects[11], social satire[12], and indie curiosities[13]– as well as the Sassy Justice channel from South Park creators Trey Parker, and Matt Stone (and actor Peter Serafinowicz), which mocked Donald Trump in a series of well-rendered deepfake skits immediately prior to the 2020 US election[14].

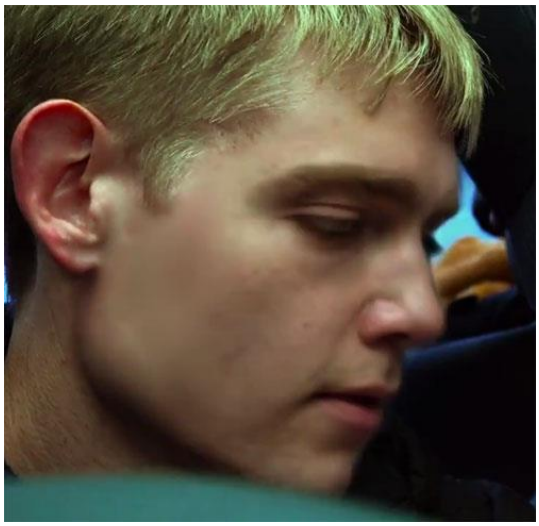


In regard to mainstream releases for FOSS deepfake output, Framestore's chief creative officer Tim Webber told[15] the FT in 2019 that deepfake software provided 'a few seconds' of a de-aged Bill Nighy in that year's *Pokémon Detective Pikachu*, though this (apparently) official debut of FOSS deepfake software in movie production went largely unnoticed – arguably a compliment to the technique.

Or not. "It required some fixes," Framestore Executive Creative Director William Bartlett said to[16] VFXVoice, "but deep fakes suited this need because of the nature of what would ultimately be presented on screen."

Additionally, *Terminator: Dark Fate* director Tim Miller was reported[17] to feel that some of the CG actor renderings for the movie (released four months after *Pokémon Detective Pikachu*) needed 'something extra', and directed the VFX department to use deepfake software on some existing CGI head replacements.

In 2020 a more extensive use of open source deepfake software emerged, when documentarian David France sought volunteers to 'donate' their faces as proxies for the at-risk subjects interviewed for *Welcome to Chechnya.*



Welcome To Chechnya *(2020) used a home-spun deepfake algorithm to protect the identities of the at-risk subjects featured. Seven 'face donors' volunteered their likenesses for the project.*

Ryan Laney, the visual effects producer for the documentary, describes[18] the custom framework that was developed for the project as 'an encryption key for the subjects identities', and consequently will not release extensive details of it.

I'm Not Quite Ready for My Close-up

From the moment that deepfake technology came to the attention of the world[19] via a (now shuttered[20]) Reddit group late in 2017, any headlines related to it that were not about porn[21] or politics[22] centered on the comparative shortcomings of Hollywood's own attempts at face simulation[23], and lionized the latest and greatest effort from a new tranche of celebrity deepfakers, such as Ctrl Shift Face[24] and Belgian VFX artist and deepfaker Christopher Ume[25].

At the time of writing, one of the most recent deepfake sensations was a collaboration[26] between Ume and Tom Cruise impersonator Miles Fisher, which exceeded 11 million views on TikTok, among other platforms.

During the video's publicity run, NBC cited[27] an image forensics professor at Berkeley as saying that the fake Cruise video is 'a big step forward in the development of this technology'.

Deepfake 'Tells'

However, notwithstanding what is clearly a phenomenal data-gathering, curation and compositional effort on the part of Ume, the DFL/FaceSwap workflow still shows 'the usual suspects', at this level of scrutiny, due to limitations that are beyond *any* user's control:

(https://www.youtube.com/watch?v=wq-kmFCrF5Q)

In the first image, there's a kind of demilitarized zone between the hairline and forehead, where the skin is unnaturally smoothed out. The wrinkles themselves are over-simplified and sketchy, and, in the second picture, the masking algorithm has failed to perform occlusion on the sunglasses as they come off 'Cruise's' face.

In the third image, the lineaments of the alpha channel for the face fails at one of deepfakes' toughest hurdles – the jawline. It also breaks continuity around the cheekbone area.

Where the problem is textural realism (rather than compositing issues), what can be done to improve output? Besides getting lucky in the current VRAM famine[28], and throwing more Tensor cores at the problem, actually not much.

Practical Limitations of Deepfake Architecture

"To make deepfakes work for production," says Dan Ring, "training times need to be significantly faster to meet delivery deadlines. But it's difficult to improve the performance of deepfakes architecture by throwing resources at it."

Based in Dublin, Ring is Head of Research for UK visual effects software company Foundry, with a deep interest in the potential of machine learning, and some experience of trialing FOSS deepfake software to establish its limits and potential.

"Hooking up a bank of Tesla V100s for a deepfakes set-up may get you faster training," he says, "but as the resolution of the input imagery and complexity of the model increases, you start to hit new bottlenecks.

"With that kind of parallel processing across GPUs in multiple machines, it's like sending ten smart people to solve a very hard problem and only letting them meet once a week. Although there's been some impressive work in this [regard], currently there's just no practical orchestration mechanism to make that work for production."

Ring says that ideally one would need a single latent space operating on a GPU that is faster and has higher levels of VRAM than anything that's on the market now, in order to push the boundaries of what's currently possible in the standard open source deepfake packages.

"This means that to reach the quality level needed for VFX today, you have to be clever, making choices about what kind of angles and expressions to prioritize."

Ring also notes that that there isn't a 1:1 relationship between the resolution of the input images and the plausibility of the output image.

"At some point," he says, "you become limited by the model size, complexity or architecture. This means additional post-processing is needed to raise the fidelity, either by an artist's hand, or more likely by a secondary network trained to remove errors in the performance transfer and enhance facial detail."

Digital Domain's Charlatan System

Machine learning is reported to play a significant role in Digital Domain's mysterious Charlatan system, which was employed to bring legendary NFL coach Vince Lombardi back to life[29] for a spot at Super Bowl LV, and to add decades to David Beckham for an appeal to solve the world's malaria crisis.

It's for the viewer to decide if the proprietary, and reportedly very labor-intensive tool does any better than the popular open source frameworks:

Disney's High Resolution Neural Face Swapping

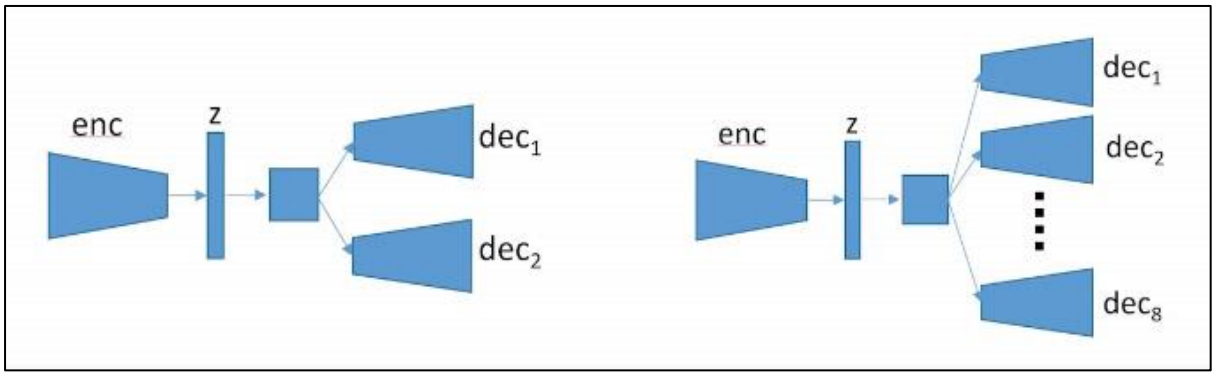
In research released in the summer of 2020, Disney Research Labs seems to have overcome the quality limits on the FOSS packages by abstracting the core principles of deepfake architecture into an original new system, and adding progressive training.

(https://www.youtube.com/watch?v=yji0t6KS7Qo)

The system can train more than two facial identities simultaneously, and is capable of delivering megapixel output, with radically improved temporal stability.

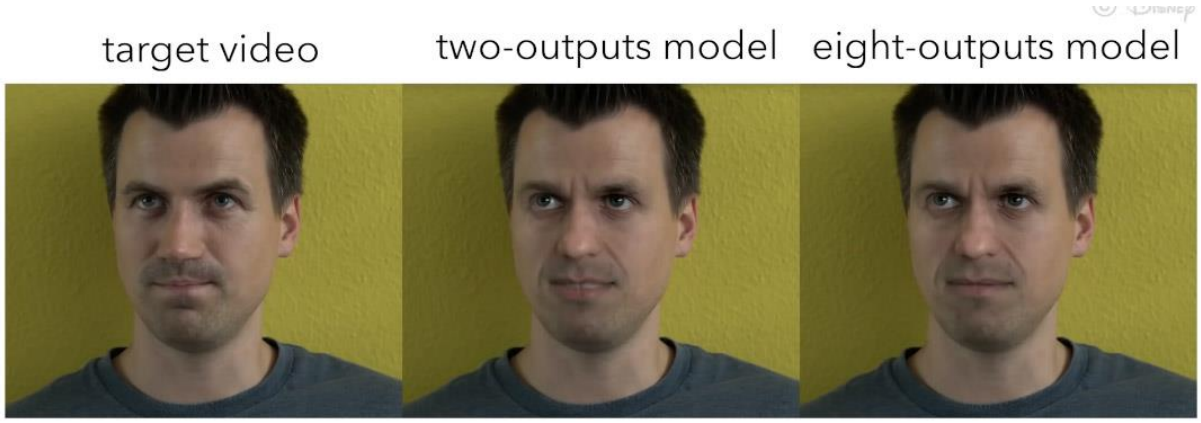
The new architecture, called High Resolution Neural Face Swapping (HRNFS) centers on a domain-transfer approach, wherein images from any number of identities in a single database are embedded into the shared latent space via a common encoder, and mapped back into pixel space using the decoder that's particular to whichever identity is desired for the swap.

The visualized design of the network inspired Disney's researchers to call this a 'comb network', wherein each identity-locked decoder is one of the 'teeth'.



(https://www.youtube.com/watch?v=yji0t6KS7Qo)

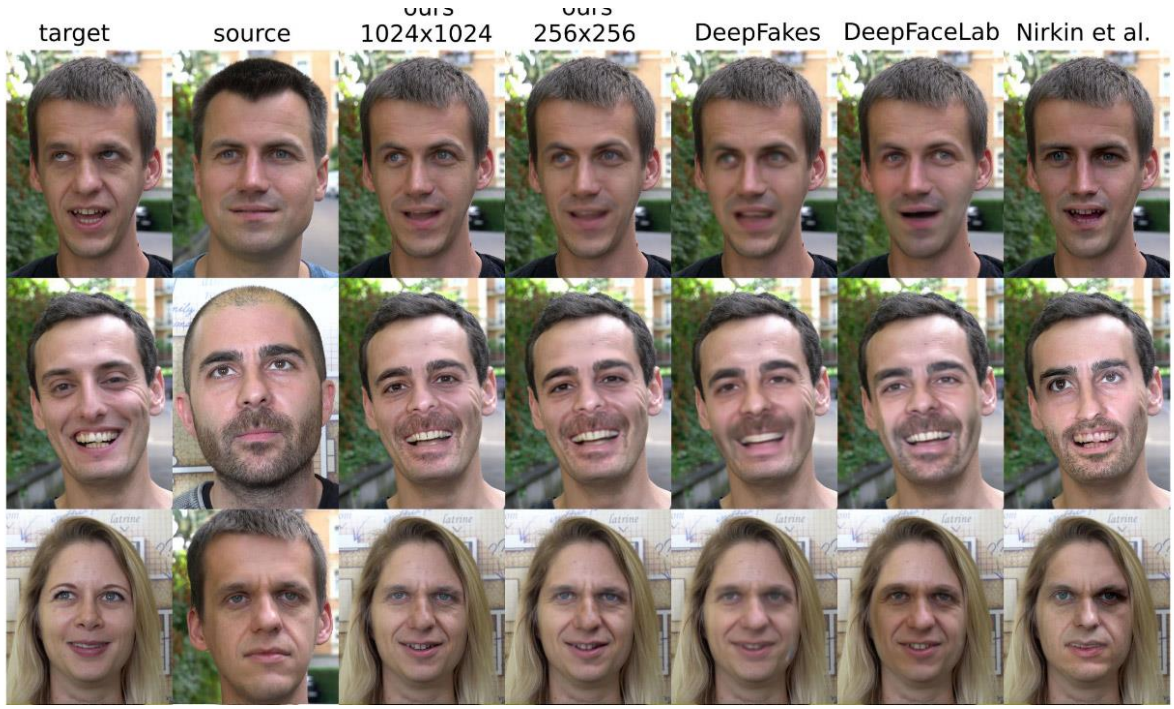
The fact that the system can train multiple identities without pairing them to any of the others makes for a drastic reduction in training time, according to the researchers.



Swaps are much better from a Disney model trained on eight identities (right) rather than two (middle).

The reported training experiments uses the Adam optimizer[30] training a 1024×1024 network for two identities. The training took three days on a single GeForce 1080Ti GPU. The 1080ti is considered a 'mid-level' consumer card in the deepfake community, which is more used to measuring training time in weeks, and which can train only two identities at a time, rather than the multiple identities that Disney's network can.

It could be argued that Disney's innovation is equivalent to creating an 'iron horse' rather than inventing the combustion engine, since it improves upon existing conceptual pipelines instead of clearing the map for a novel solution to the problem. The comparison tests published in the research paper show some of the same problems detailed earlier, such as unconvincing wrinkles in older face subjects.



Comparison of Disney's new architecture to the original Deepfakes code, DFL, and the method outlined in 2017 in a study from the Institute For Robotics And Intelligent Systems at USC.

We'll have to wait for production deployment (i.e., use in a movie) to see whether or not HRNFS represents a quantum leap in AI-based identity synthesis, or just a more efficient way to race towards the stubborn limits of deepfakes' data architecture.

However, the use of multiple trained identities to improve swap-quality for a single identity is a major advance, within those constraints.

Deepfake Detection to Improve Traditional VFX Methods?

'When amateur deepfakes on YouTube rival movie VFX face-replacements, the directors and studios start to ask about AI tech!'

So says the academy award-winning visual effects supervisor Scott Stokdyk, also the VFX lead on Tom Hanks' AI outing *BIOS* (2021).

'For now,' he told me. 'it seems that the deep fake work cannot get to 100% quality consistently on every shot, and that there is no current way with that methodology to get from 90% to 100%. It can be part of the toolset, but shouldn't be relied on for an entire project that has many different scenarios.'

But Stokdyk also envisages a rather different application of DeepFake technology than the cinematic proliferation we've been waiting for in vain the last three years.

'One area that interests me,' he said. 'is that the same technology that is starting to be used to detect deep fakes should also be repurposed to detect more general synthetic qualities in VFX images.

'These machine learning algorithms can look at image histograms or different representations of images (i.e. in the frequency or HSV domains) and should be able to nudge the images toward [being] more photoreal. It is still difficult in many situations to have all-CG VFX look real, and I think machine learning will help with this problem.'

Conclusion

Not every amazing new technology develops consistently, and some don't really develop at all: the combustion engine; the incandescent light bulb; the refrigerator; and the toilet, to name but a few of those innovations that made only a single evolutionary leap.

The current architecture of popular deepfake software (and in most professional use cases, that's DeepFaceLab, even if VFX producers play the fact down, perhaps to avoid being associated with deepfake porn) has changed very little in the three years that it's been fueling catastrophic headlines. What *has* changed is that deepfake practitioners have had years to thoroughly scope the limits of the process, and to develop tortuous workarounds for the endless blind alleys of the software. Likewise the VFX industry.

So is something better brewing in the labs of the world's leading visual effects houses? All the VFX producers I talked to in the course of writing this article (not all of which are named here) who were prepared to admit which software they were using told me that it was DeepFaceLab – the same free package used to make Tom Cruise viral fakes and deepfake porn – though admittedly their dataset acquisition and curation resources are far in advance of most amateur users.

It's possible that truly powerful identity substitution will eventually be made possible by machine learning methods that are far more rudimentary at the moment, such as AI-generated image synthesis; and that when this comes to pass, with the power to recreate entire people instead of hunting round for 'body donors' and 'similar' faces – yes, our faith in the moving image as a token of truth will be truly threatened; and that we'll remember the FOSS deepfakes era with the same amusement as the doomed VR boom of the early 1990s – as an early preview of a technology whose time had not yet come.

---

[1] *DeepFaceLab* – iperov, GitHub. https://github.com/iperov/DeepFaceLab
[2] *FaceSwap* – https://faceswap.dev/
[3] *nb* – I did not create the Eilish/Garland deepfake, but have used it as an illustrative reference to describe the process.
[4] *Face Recognition* – Adrian Bulat, GitHub. https://github.com/1adrianb/face-alignment
[5] *What is Training?* – torzdf, faceswap.dev, September 29th 2019. https://forum.faceswap.dev/viewtopic.php?t=146#what
[6] DFL now features an optional GAN element for enhancement, though many users complain of the grain that it can generate.
[7] *Understanding the Technology Behind DeepFakes* – Alan Zucconi, alanzucconi.com.
https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/
[8] *deepfakes_faceswap* – joshua-wu, GitHub. https://github.com/joshua-wu/deepfakes_faceswap
[9] *How does multi GPU training work?* – https://forum.faceswap.dev/viewtopic.php?t=767
[10] *Brands Are Finding Deepfakes Increasingly Appealing for Ad Campaigns* – Patrick Kulp, AdWeek, October 5th 2020.
https://www.adweek.com/performance-marketing/brands-deepfakes-appealing-ads-campaigns/
[11] *Brands Are Finding Deepfakes Increasingly Appealing for Ad Campaigns* – Patrick Kulp, AdWeek, October 5th 2020.
https://www.adweek.com/performance-marketing/brands-deepfakes-appealing-ads-campaigns/
[12] *DEEPFAKE IT 'TIL WE MAKE IT* – https://www.deepreckonings.com/
[13] *Deepfake Queen: 2020 Alternative Christmas Message* – Channel 4, YouTube, 25th December 2020.
https://www.youtube.com/watch?v=IvY-Abd2FfM
[14] *The 'South Park' Guys Break Down Their Viral Deepfake Video* – Dave Itzkoff, New York Times, October 29th 2020.
https://www.nytimes.com/2020/10/29/arts/television/sassy-justice-south-park-deepfake.html
[15] *Deepfakes: Hollywood's quest to create the perfect digital human* – Tim Bradshaw, Financial Times, 10th October 2019.
https://www.ft.com/content/9df280dc-e9dd-11e9-a240-3b065ef5fc55
[16] *Deep Fakes: Part 1 – A Creative Perspective* – Ian Failes, VFX Voice, 24th November 2020. https://www.vfxvoice.com/deep-fakes-part-1-a-creative-perspective/
[17] Here's what I learnt at the VFX Bake-Off 2020 – *Ian Failes, Befores & Afters, January 6, 2020.*
https://beforesandafters.com/2020/01/06/heres-what-i-learnt-at-the-vfx-bake-off-2020/
[18] In an email to me, 10th March 2021.
[19] *AI-Assisted Fake Porn Is Here and We're All F****** – Samantha Cole, Motherboard Vice, 11th December 2017.
https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn
[20] https://www.reddit.com/user/deepfakes/
[21] *The most urgent threat of deepfakes isn't politics. It's porn* – Cleo Abram, Vox, June 8th 2020.
https://www.vox.com/2020/6/8/21284005/urgent-threat-deepfakes-politics-porn-kristen-bell
22 The age of the deepfake is here – and it's more terrifying than you think – Ellie Zolfagharifard, The Telegraph, 7th September 2020.
https://www.telegraph.co.uk/technology/2020/09/05/age-deepfake-stranger-think/
[23] *The Irishman's fan-made de-aging deepfake branded 'mind-blowingly better' than the original* – Louis Chilton, The Independent, 25th August 2020 . https://www.independent.co.uk/arts-entertainment/films/news/irishman-deaging-deepfake-cgi-netflix-deniro-pacino-scorsese-video-a9686926.html

Rogue One Deepfake Makes Star Wars' Leia And Grand Moff Tarkin Look Even More Lifelike – Corey Chichizola, Cinema Blend, 9th December 2020. https://www.cinemablend.com/news/2559935/rogue-one-deepfake-makes-star-wars-leia-and-grand-moff-tarkin-look-even-more-lifelike

https://www.foundry.com/insights/film-tv/digital-humans

Bridging the uncanny valley: what it really takes to make a deepfake – Foundry, 9th December 2019. https://www.foundry.com/insights/film-tv/deepfakes-de-aging

[24] *Ctrl Shift Face channel* – YouTube, https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g

[25] *Neo Takes The Blue Pill [DeepFake]* – Ctrl Shift Face, YouTube, February 17th 2020. https://www.youtube.com/watch?v=aau8qa3xgFs1

[26] *'I don't want to upset people': Tom Cruise deepfake creator speaks out* – Alex Hern, The Guardian, 5th March 2021. https://www.theguardian.com/technology/2021/mar/05/how-started-tom-cruise-deepfake-tiktok-videos

[27] *Deepfake videos of Tom Cruise went viral. Their creator hopes they boost awareness* – Bianca Britton, NBC News, March 5th 2021. https://www.nbcnews.com/tech/tech-news/creator-viral-tom-cruise-deepfakes-speaks-rcna356

[28] *Nvidia RTX 3000 GPU shortages are so bad it's bringing back…the GTX 1050 Ti?!* – Matt Hanson, Tech Radar, February 11th 2021. https://www.techradar.com/news/nvidia-rtx-3000-series-gpu-shortages-are-so-bad-its-bringing-backthe-gtx-1050-ti

[29] *The DD Vince Lombardi Charlatan* – Mike Seymour, fxguide, 8th February 2021. https://www.fxguide.com/quicktakes/the-dd-vince-lombardi-charlatan/

[30] *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning* – Jason Brownlee, Machine Learning Mastery, July 3rd 2017 . https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/