

A New and Simpler Deepfake Method That Outperforms Prior Approaches

Originally published at <https://www.unite.ai/a-new-and-simpler-deepfake-method-that-outperforms-prior-approaches/>



A collaboration between a Chinese AI research group and US-based researchers has developed what may be the first real innovation in deepfakes technology since the phenomenon emerged four years ago.

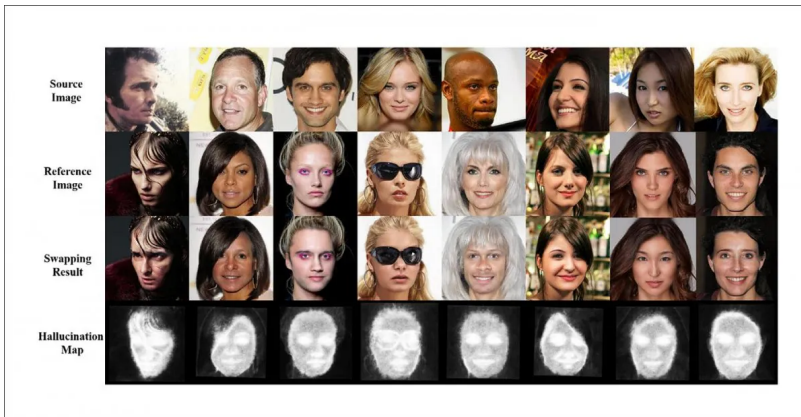
The new method can perform faceswaps that outperform all other existing frameworks on standard perceptual tests, without needing to exhaustively gather and curate large dedicated datasets and train them for up to a week for just a single identity. For the examples presented in the new paper, models were trained on the *entirety* of two popular celebrity datasets, on one NVIDIA Tesla P40 GPU for about three days.



CLICK TO VIEW ANIMATED GIF

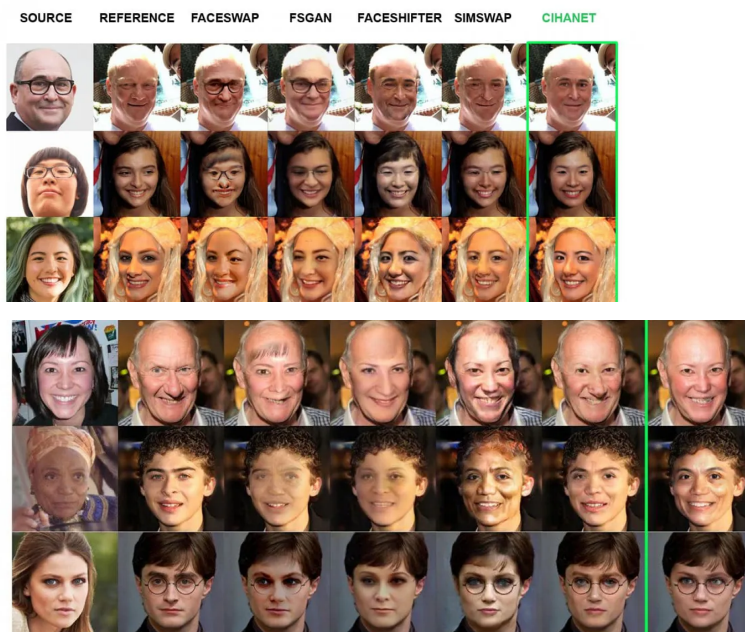
Full video available at the end of this article. *In this sample from a video in supplementary materials provided by one of the authors of the new paper, Scarlett Johansson's face is transferred onto the source video. CifaNet removes the problem of edge-masking when performing a swap, by forming and enacting deeper relationships between the source and target identities, meaning an end to 'obvious borders' and other superimposition glitches that occur in traditional deepfake approaches.* Source: <https://mitchellx.github.io/#video>

The new approach removes the need to 'paste' the transplanted identity crudely into the target video, which frequently leads to tell-tale [artifacts](#) that appear where the fake face ends and the real, underlying face begins. Rather, 'hallucination maps' are used to perform a deeper mingling of visual facets, because the system separates identity from context far more effectively than current methods, and therefore can blend the target identity at a more profound level.



From the paper. CihNet transformations are facilitated through hallucination maps (bottom row). The system uses context information (i.e. face direction, hair, gla entirely from the image into which the new identity will be superimposed, and facial identity information entirely from the person who is to be inserted into the imag from context is critical to the success of the system. Source: <https://dl.acm.org/doi/pdf/10.1145/3474085.3475257>

Effectively the new hallucination map provides a more complete context for the swap, as opposed to the hard masks that often require extensive curation (and in the case of DeepFaceLab, [separate training](#)) while providing limited flexibility in terms of real incorporation of the two identities.



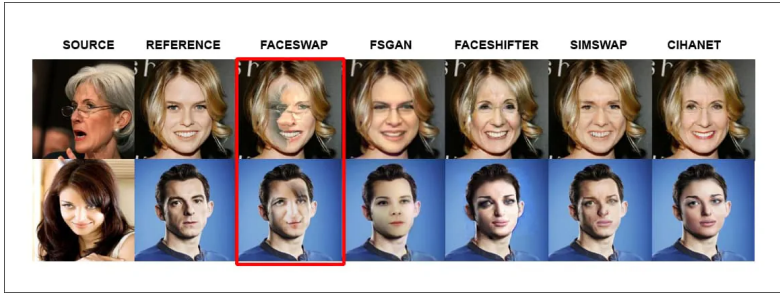
From samples provided in the supplementary materials, using both the FFHQ and Celeb-A HQ datasets, across VGGFace and Forensics++. The first two columns (real) images to be swapped. The following four columns show the results of the swap using the four most effective methods currently available, while the final col CihNet. The FaceSwap repository has been used, rather than the more popular DeepFaceLab, since both projects are forks of the original 2017 Deepfakes code has since added models, techniques, diverse UIs and supplementary tools, the underlying code that makes deepfakes possible has never changed, and remains c <https://dl.acm.org/action/downloadSupplement?doi=10.1145%2F3474085.3475257&file=mp0519aux.zip>

The [paper](#), titled *One-stage Context and Identity Hallucination Network*, is authored by researchers affiliated with JD AI Research, and the University of Massachusetts Amherst, and was supported by the National Key R&D Program of China under Grant No. 2020AAA0103800. It was introduced at the 29th ACM International Conference on Multimedia, on October 20th- 24th, at Chengdu, China.

No Need for 'Face-On' Parity

Both the most popular current deepfake software, DeepFaceLab, and competing fork FaceSwap, perform tortuous and frequently hand-curated workflows in order to identify which way a face is inclined, what obstacles are in the way that must be accounted for (again, manually), and must cope with many other irritating impediments (including lighting) that make their use far from the 'point-and-click' experience inaccurately portrayed in the media since the advent of deepfakes.

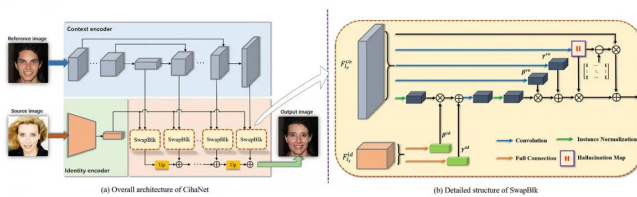
By contrast, CihNet does not require that two images be facing the camera directly in order to extract and exploit useful identity information from a single image.



In these examples, a suite of deepfake software contenders are challenged with the task of swapping faces that are not only dissimilar in identity, but which are no Software derived from the original deepfakes repository (such as the hugely popular DeepFaceLab and FaceSwap, pictured above) cannot handle the disparity in images to be swapped (see third column). Meanwhile, Cihanet can abstract the identity correctly, since the 'pose' of the face is not intrinsically part of the identity i

Architecture

The Cihanet project, according to the authors, was inspired by the 2019 collaboration between Microsoft Research and Peking University, called [FaceShifter](#), though it makes some notable and critical changes to the core architecture of the older method.

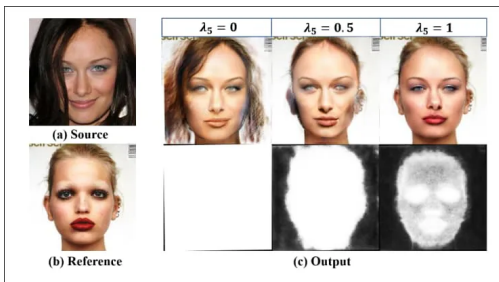


FaceShifter uses two Adaptive Instance Normalization ([AdaIN](#)) networks to handle identity information, which data is then transposed into the target image via a mask, in a way similar to current popular deepfake software (and with all its related limitations), using an additional [HEAR-Net](#) (which includes a separately trained sub-net trained on occlusion obstacles – an additional layer of complexity).

Instead, the new architecture directly uses this 'contextual' information for the transformative process itself, via a two-step single Cascading Adaptive Instance Normalization (C-AdaIN) operation, which provides consistency of context (i.e. face skin and occlusions) of ID-relevant areas.

The second sub-net crucial to the system is called Swapping Block (SwapBlk), which generates an integrated feature from the context of the reference image and the embedded 'identity' information from the source image, bypassing the multiple stages necessary to accomplish this by conventional current means.

To help distinguish between context and identity, a *hallucination map* is generated for each level, standing in for a soft-segmentation mask, and acting on a wider range of features for this critical part of the deepfake process.



As the value of the hallucination map (pictured below right) grows, a clearer path between identities emerges.

In this way, the entire swapping process is accomplished in a single stage and without post-processing.

Data and Testing

To try out the system, the researchers trained four models on two highly popular and variegated open image datasets – [CelebA-HQ](#) and NVIDIA's Flickr-Faces-HQ Dataset ([FFHQ](#)), each containing 30,000 and 70,000 images respectively.

No pruning or filtering was performed on these base datasets. In each case, the researchers trained the entirety of each dataset on the single Tesla GPU over three days, with a learning rate of 0.0002 on Adam optimization.

They then rendered out a series of random swaps among the thousands of personalities featured in the datasets, without regard for whether or not the faces were similar or even gender-matched, and compared Cihanet's results to the output from four leading deepfake frameworks: [FaceSwap](#) (which stands in for the more popular [DeepFaceLab](#), since it shares a root codebase in the [original 2017 repository](#) that brought deepfakes to the world); the aforementioned FaceShifter; [FSGAN](#); and [SimSwap](#).

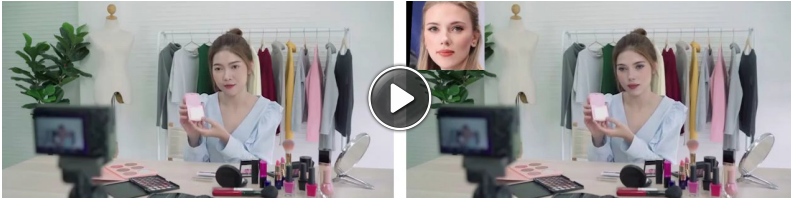
In comparing the results via [VGG-Face](#), FFHQ, CelebA-HQ and [FaceForensics++](#), the authors found that their new model outperformed all prior models, as indicated in the table below.

	CelebA-HQ	FFHQ	VGGFace	Forensic++
	SSIM ↑ / pose ↓ / ID ↑	SSIM ↑ / pose ↓ / ID ↑	SSIM ↑ / pose ↓ / ID ↑	SSIM ↑ / pose ↓ / ID ↑
FaceSwap [12]	0.979 / 3.01 / 0.589	0.972 / 3.63 / 0.577	0.985 / 4.19 / 0.514	0.991 / 3.44 / 0.470
FSGAN [31]	0.939 / 2.89 / 0.224	0.948 / 3.56 / 0.222	0.950 / 4.49 / 0.268	0.984 / 3.54 / 0.410
FaceShifter [24]	0.920 / 3.49 / 0.971	0.932 / 3.89 / 0.978	0.954 / 4.93 / 0.969	0.986 / 4.07 / 0.793
SimSwap [8]	0.954 / 3.02 / 0.968	0.945 / 3.74 / 0.962	0.967 / 4.34 / 0.947	0.987 / 3.79 / 0.776
Ours	0.987 / 2.76 / 0.975	0.984 / 2.95 / 0.983	0.990 / 3.91 / 0.970	0.992 / 3.27 / 0.804

The three metrics used in evaluating the results were Structural Similarity ([SSIM](#)), [pose estimation error](#) and [ID retrieval accuracy](#), which is computed based on the percentage of successfully retrieved pairs.

The researchers contend that Cihanet represents a superior approach in terms of qualitative results, and a notable advance on the current state of the art in deepfake technologies, by removing the burden of extensive and labor-intensive masking architectures and methodologies, and achieving a more useful and actionable separation of identity from context.

Take a look below to see further video examples of the new technique. You can find the full-length video [here](#).



CLICK TO PLAY VIDEO ON SITE

From supplementary materials for the new paper, Cihanet performs faceswapping on various identities. Source: <https://mitchellx.github.io/#video>

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