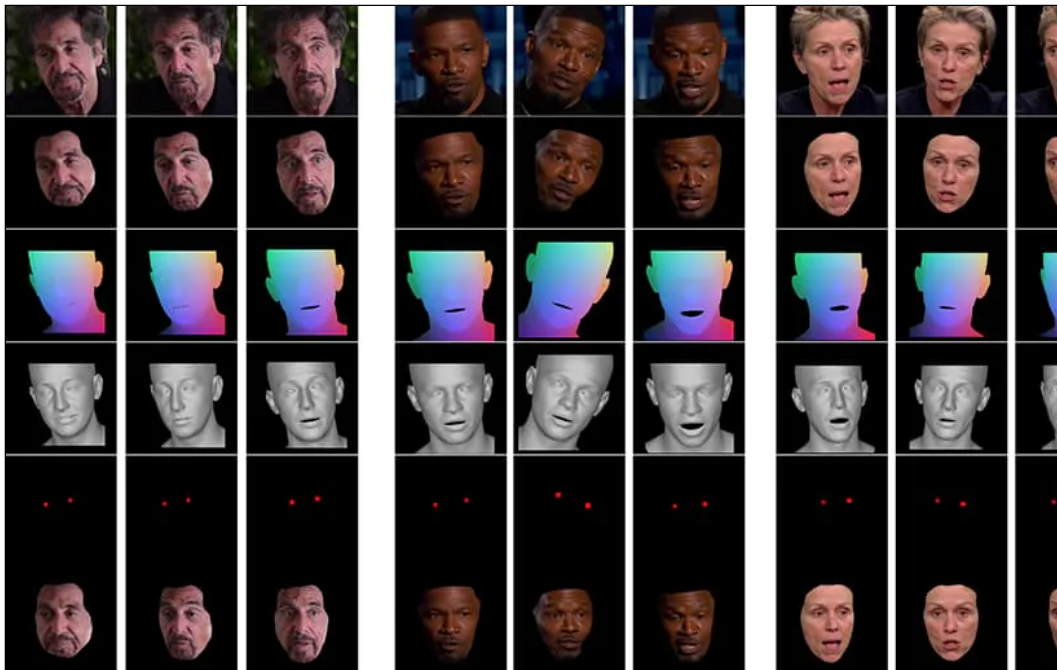
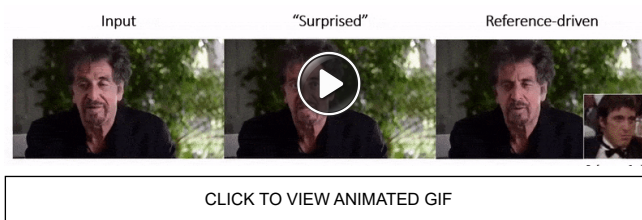


Altering Emotions in Video Footage With AI

Originally published at <https://www.unite.ai/altering-emotions-in-video-footage-with-ai/>



Researchers from Greece and the UK have developed a novel deep learning approach to changing the expressions and apparent mood of people in video footage, whilst preserving the fidelity of their lip movements to the original audio in a way that prior attempts have not been able to match.



From the video accompanying the paper (embedded at the end of this article), a brief clip of actor Al Pacino having his expression subtly altered by NED, based on high-level semantic concepts defining individual facial expressions, and their associated emotion. The 'Reference-Driven' method on the right takes the interpreted emotion/s of a source video and applies it to the entirety of a video sequence. Source: <https://www.youtube.com/watch?v=Li6W8pRDMJQ>

This particular field falls into the growing category of [deepfaked emotions](#), where the identity of the original speaker is preserved, but their expressions and micro-expressions are altered. As this particular AI technology matures, it offers the possibility for movie and TV productions to make subtle alterations to actors' expressions – but also opens up a fairly new category of 'emotion-altered' video deepfakes.

Changing Faces

Facial expressions for public figures, such as politicians, are rigorously curated; in 2016 Hillary Clinton's facial expressions came [under intense media scrutiny](#) for their potential negative impact on her electoral prospects; facial expressions, it transpires, are also a [topic of interest](#) to the FBI; and they're a [critical indicator](#) in job interviews, making the (far distant) prospect of a live 'expression-control' filter a desirable development for job-seekers trying to pass a pre-screen on Zoom.

A 2005 study from the UK asserted that facial appearance [affects voting decisions](#), while a 2019 Washington Post feature examined the [use of 'out of context' video clip sharing](#), which is currently the nearest thing that fake news proponents have to actually being able to change how a public figure appears to be behaving, responding, or feeling.

Towards Neural Expression Manipulation

At the moment, the state of the art in manipulating facial affect is fairly rudimentary, since it involves tackling the [disentanglement](#) of high-level concepts (such as *sad*, *angry*, *happy*, *smiling*) from actual video content. Though traditional deepfake architectures appear to achieve this disentanglement quite well, mirroring emotions across different identities still requires that two training face-sets contain matching expressions for each identity.



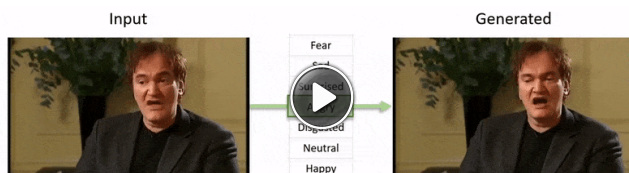
Typical examples of face images in datasets used to train deepfakes. Currently, you can only manipulate a person's facial expression by creating ID-specific expression<->expression pathways in a deepfake neural network. 2017-era deepfake software has no intrinsic, semantic understanding of a 'smile' – it just maps-and-matches perceived changes in facial geometry across the two subjects.

What's desirable, and has not yet been perfectly achieved, is to recognize how subject B (for instance) smiles, and simply create a 'smile' switch in the architecture, without needing to map it to an equivalent image of subject A smiling.

The [new paper](#) is titled *Neural Emotion Director: Speech-preserving semantic control of facial expressions in "in-the-wild" videos*, and comes from researchers at the School of Electrical & Computer Engineering at the National Technical University of Athens, the Institute of Computer Science at the Foundation for Research and Technology Hellas (FORTH), and the College of Engineering, Mathematics and Physical Sciences at the University of Exeter in the UK.

The team has developed a framework called *Neural Emotion Director* (NED), incorporating a 3D-based emotion-translation network, *3D-Based Emotion Manipulator*.

NED takes a received sequence of expression parameters and translates them to a target domain. It's trained on unparallel data, which means that it is not necessary to train on datasets where each identity has corresponding facial expressions.



CLICK TO VIEW ANIMATED GIF

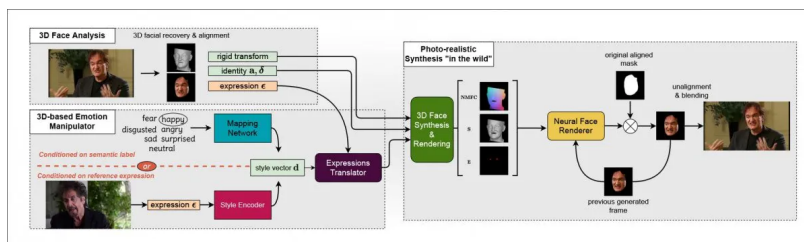
The video, shown at the end of this article, runs through a series of tests where NED imposes an apparent emotional state onto footage from the YouTube dataset.

The authors claim that NED is the first video-based method for 'directing' actors in random and unpredictable situations, and have made the code available on NED's [project page](#).

Method and Architecture

The system is trained on two large video datasets that have been annotated with 'emotion' labels.

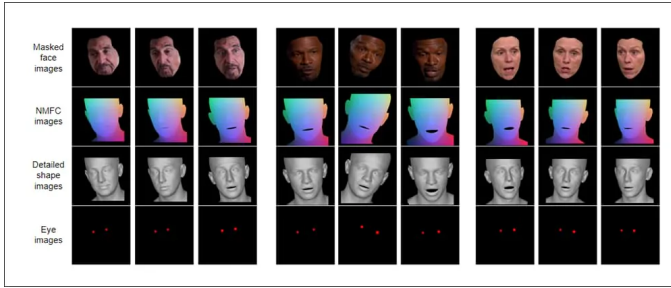
The output is enabled by a video face renderer that renders the desired emotion to video using traditional facial image synthesis techniques, including face segmentation, facial landmark alignment and blending, where only the facial area is synthesized, and then imposed onto the original footage.



The architecture for the pipeline of the Neural Emotion Detector (NED). Source: <https://arxiv.org/pdf/2112.00585.pdf>

Initially, the system obtains 3D facial recovery and imposes facial landmark alignments on the input frames in order to identify the expression. After this, these recovered expression parameters are passed to the 3D-based Emotion Manipulator, and a style vector computed by means of either a semantic label (such as 'happy') or by a reference file.

A reference file is a video portraying a particular recognized expression/emotion, which is then imposed onto the entirety of the target video, swapping out the original expression.



Stages in the emotion transfer pipeline, featuring various actors sampled from YouTube videos.

The final generated 3D face shape is then concatenated with the Normalized Mean Face Coordinate (NMFC) and the eye images (the red dots in the image above), and passed to the neural renderer, which performs the final manipulation.

Results

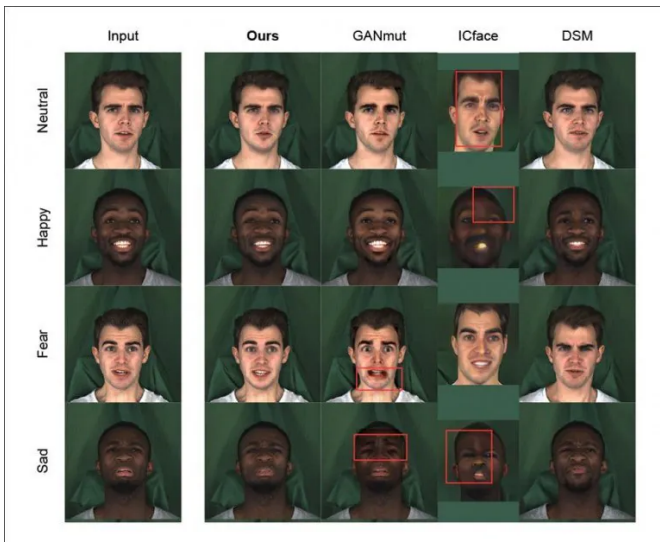
The researchers conducted extensive studies, including user and ablation studies, to evaluate the effectiveness of the method against prior work, and found that in most categories, NED outperforms the current state of the art in this sub-sector of neural facial manipulation.

	Ours		GANmut		DSM		ICface	
	FAPD	FID	FAPD	FID	FAPD	FID	FAPD	FID
neutral	14.9	2.1	16.8	2.9	22.1	5.5	40.0	45.8
happy	17.8	3.4	15.0	2.6	27.0	10.3	43.7	50.5
fear	18.4	3.0	20.5	4.3	28.0	8.5	43.0	46.6
sad	19.0	3.0	18.1	4.3	24.5	8.8	38.6	47.4
surprised	18.9	2.9	19.1	7.2	27.3	11.8	46.1	45.0
angry	18.4	3.0	22.4	4.2	-	-	51.7	53.5
disgusted	18.1	3.1	15.1	4.5	-	-	39.7	54.6
avg. (7)	17.9	2.9	18.1	4.3	-	-	43.3	49.0
avg. (5)	17.8	2.9	17.9	4.3	25.8	9.0	42.3	47.1

The paper's authors envisage that later implementations of this work, and tools of a similar nature, will be useful primarily in the TV and motion picture industries, stating:

'Our method opens a plethora of new possibilities for useful applications of neural rendering technologies, ranging from movie post-production and video games to photo-realistic affective avatars.'

This is an early work in the field, but one of the first to attempt facial reenactment with video rather than still images. Though videos are essentially many still images running together very fast, there are temporal considerations that make previous applications of emotion transfer less effective. In the accompanying video, and examples in the paper, the authors include visual comparisons of NED's output against other comparable recent methods.



More detailed comparisons, and many more examples of NED, can be found in the full video below:

NED: Speech-preserving semantic control of facial expressions in "in-the-wild" videos

Neural Emotion Director

Speech-preserving semantic control of facial expressions in “in-the-wild” videos

Foivos Paraperas Papantoniou¹, Panagiotis P. Filintisis¹, Petros Maragos¹, Anastasios Roussos^{2,3}

¹School of Electrical & Computer Engineering, National Technical University of Athens, Greece
²Institute of Computer Science (ICS), Foundation for Research & Technology - Hellas (FORTI), Greece
³College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK



3rd December 2021, 18:30 GMT+2 – At the request of one of the paper’s authors, corrections were made regarding the ‘reference file’, which I mistakenly stated was a still photo (when it is in fact a video clip). Also an amend to the name of the Institute of Computer Science at the Foundation for Research and Technology.

3rd December 2021, 20:50 GMT+2 – A second request from one of the paper’s authors for a further amendment on the name of the above-mentioned institution.

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

UNITE.AI

Discover More