

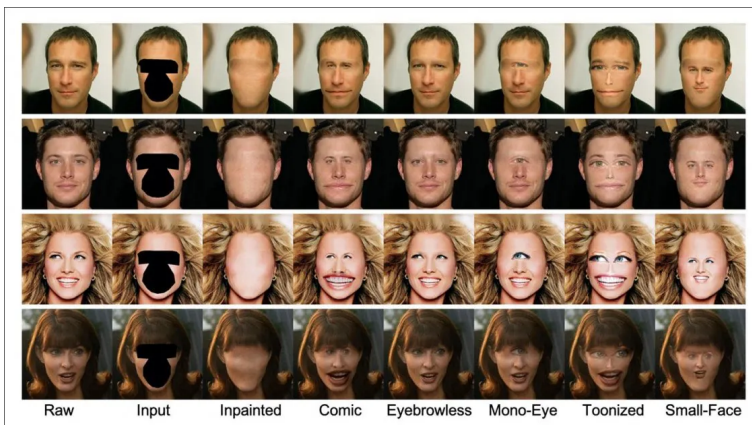
TikTok Developers Erasing Faces for Augmented Reality Applications

Originally published at <https://www.unite.ai/tiktok-developers-erasing-faces-for-augmented-reality-applications/>



ByteDance, the Chinese multinational internet company behind TikTok, has developed a new method for erasing faces in video so that identity distortion and other bizarre effects can be imposed on people in augmented reality applications. The company claims that the technique has already been integrated into commercial mobile products, though it does not state which products.

Once faces in video have been 'zeroed', there's enough 'face canvas' to produce eye-boggling distortions, as well as potentially superimposing other identities. Examples supplied in a new paper from ByteDance researchers illustrate the possibilities, including restoring the 'erased' features in various comical (and certainly some grotesque) configurations:



Some of the possibilities for facial re-configuration included in the ByteDance paper. Source: <https://arxiv.org/pdf/2109.10760.pdf>

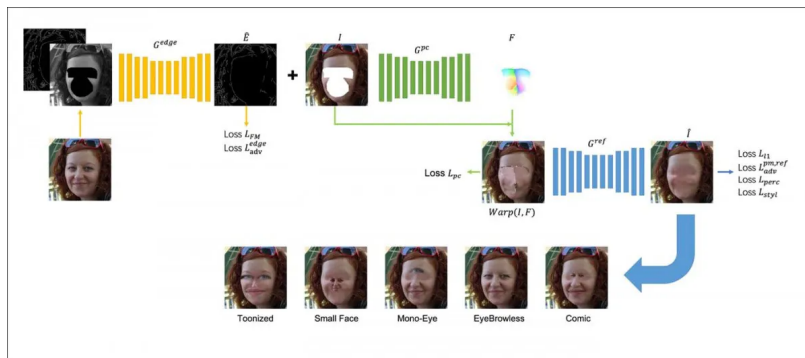
Towards the end of August, it [came to light](#) that TikTok, the first non-Facebook app to [reach three billion installs](#), had launched TikTok Effect Studio (currently [in closed beta](#)), a platform for augmented reality (AR) developers to create AR effects for TikTok content streams.

Effectively, the company is catching up to similar developer communities at [Facebook's AR Studio](#) and [Snap AR](#), with Apple's venerable [AR R&D community](#) also set to imminently become galvanized by [new hardware](#) over the next year.

Blank Expressions

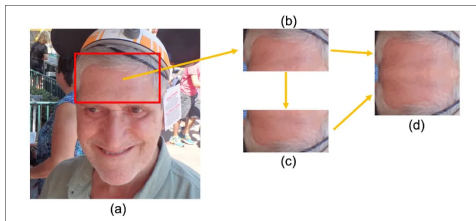
The [paper](#), titled *FaceEraser: Removing Facial Parts for Augmented Reality*, notes that existing in-painting/infill algorithms, such as [NVIDIA's SPADE](#), are more oriented towards completing truncated or otherwise semi-observed images than to performing this unusual 'blanking' procedure, and that existing dataset material is therefore predictably scarce.

Since there are no available ground truth datasets for people who have a solid expanse of flesh where their face should be, the researchers have created a novel network architecture called *pixel-clone*, that can be superimposed into existing neural inpainting models, and which resolves issues related to texture and color inconsistencies exhibited (the paper attests) by older methods such as [StructureFlow](#) and [EdgeConnect](#).



General workflow of pixel-clone in the new pipeline.

In order to train a model on 'blank' faces, the researchers precluded images with glasses, or where hair obscures the forehead, since the area between the hairline and eyebrows is usually the largest single group of pixels which can supply 'paste-over' material for the central features of the face.



Preparing training images. The forehead area is cropped out, based on key points in facial alignment recognition, vertically flipped and stitched.

A 256×256 pixel image is obtained, a small enough size to feed into the latent space of a neural network in batches that are large enough to achieve generalization. Later algorithmic upscaling will restore resolutions necessary to work in the AR space.

Architecture

The network is made up of three inner networks, comprising Edge Completion, Pixel-Clone, and a refinement network. The edge completion network uses the same kind of encoder-decoder architecture employed in EdgeConnect (see above), as well as in the two most popular deepfake applications. The encoders downsample image content twice, and the decoders restore the original image dimensions.

Pixel-Clone uses a modified encoder-decoder methodology, while the refinement layer uses U-Net architecture, a technique originally developed [for biomedical imaging](#), which often features in image synthesis research projects.

During the training workflow, it's necessary to evaluate the accuracy of the transformations, and, as necessary, repeat the attempts iteratively up to [convergence](#). To this end, two discriminators based on [PatchGAN](#) are used, each of which evaluates the localized realism of 70×70 pixel patches, discounting the realism value of the entire image.

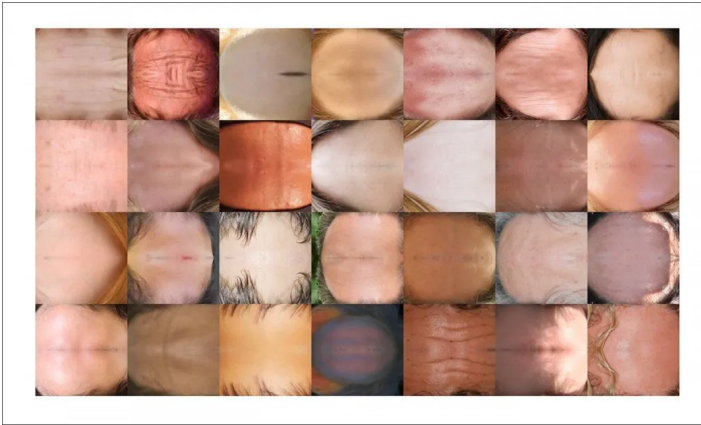
Training and Data

The edge completion network is initially trained independently, while the other two networks are trained together, based on the weights that have resulted from the edge completion training, which are fixed and frozen during this procedure.

Though the paper does not explicitly state that its examples of final feature distortion are the central aim of the model, it implements various comic effects to test the resilience of the system, including eyebrow removal, enlarged mouths, shrunken sub-faces and ‘toonized’ effects (as shown in the earlier image, above).

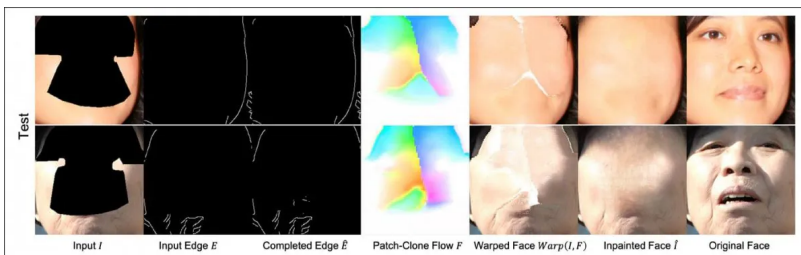
The paper asserts that ‘the erased faces enable various augmented-reality applications that require placement of any user-customized elements’, indicating the possibility of customizing faces with third-party, user-contributed elements.

The model is trained on masks from the NVIDIA-created [FFHQ dataset](#), which contains an adequate variety of ages, ethnicities, lighting and facial poses and styles to achieve a useful generalization. The dataset contains 35,000 images and 10,000 training masks to delineate the areas of transformation, with 4000 images and 1000 masks set aside for validation purposes.



Training data samples.

The trained model can perform inference on data from 2017's [CelebA-HQ](#) and [VoxCeleb](#), unseen faces from FFHQ, and any other unconstrained, unseen faces that are presented to it. The 256×256 images were trained on the network in batches of 8 over an Adam optimizer, implemented in PyTorch, and running on a Tesla V100 GPU for '2000,000 epochs'.



Inference results obtained on a real face.

As is common in face-based image synthesis research, the system has to contend with occasional failures provoked by obstructions or occlusions such as hair, peripherals, glasses, and facial hair.

The report concludes:

'Our approach has been commercialized and it works well in products for unconstrained user inputs.'

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