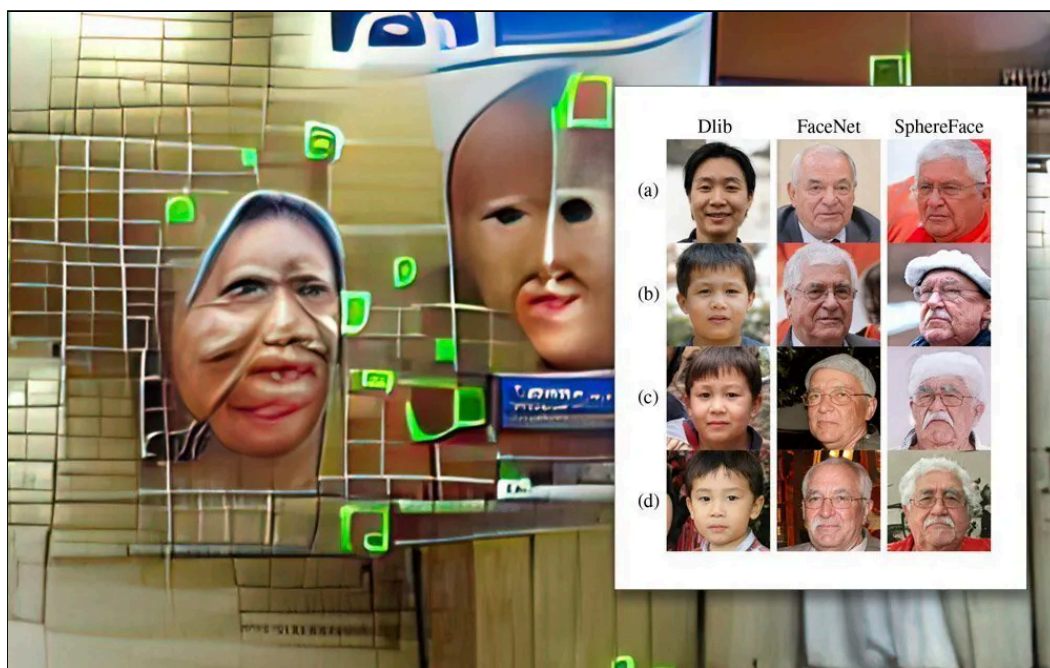


'Master Faces' That Can Bypass Over 40% Of Facial ID Authentication Systems

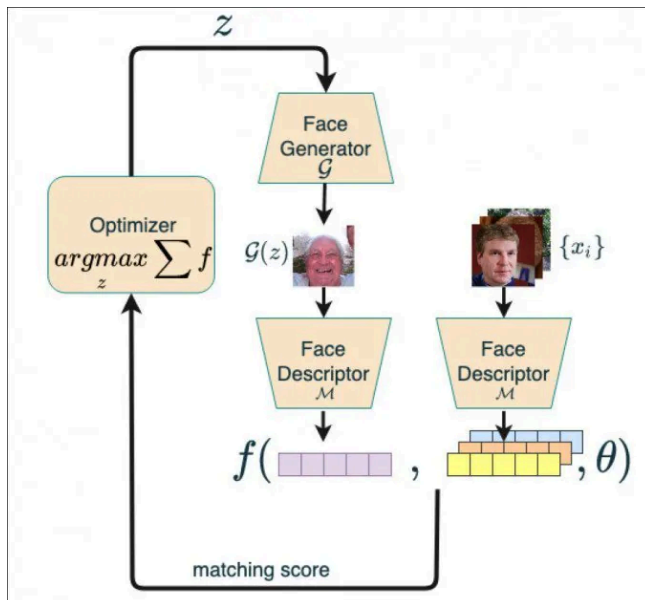
Originally published at <https://www.unite.ai/master-faces-that-can-bypass-over-40-of-facial-id-authentication-systems/>



Researchers from Israel have developed a neural network capable of generating 'master' faces – facial images that are each capable of impersonating multiple IDs. The work suggests that it's possible to generate such 'master keys' for more than 40% of the population using only 9 faces synthesized by the StyleGAN Generative Adversarial Network (GAN), via three leading face recognition systems.

The [paper](#) is a collaboration between the Blavatnik School of Computer Science and the school of Electrical Engineering, both at Tel Aviv.

Testing the system, the researchers found that a single generated face could unlock 20% of all identities in the University of Massachusetts' Labeled Faces in the Wild ([LFW](#)) open source database, a common repository used for development and testing of facial ID systems, and the benchmark database for the Israeli system.



The Israeli system workflow, which uses the StyleGAN generator to iteratively seek out ‘master faces’. Source: <https://arxiv.org/pdf/2108.01077.pdf>

The new method improves on a similar [recent paper](#) from the University of Siena, which requires a privileged level of access to the machine learning framework. By contrast, the new method infers generalized features from publicly available material and uses it to create facial characteristics that straddle a vast number of identities.

Evolving Master Faces

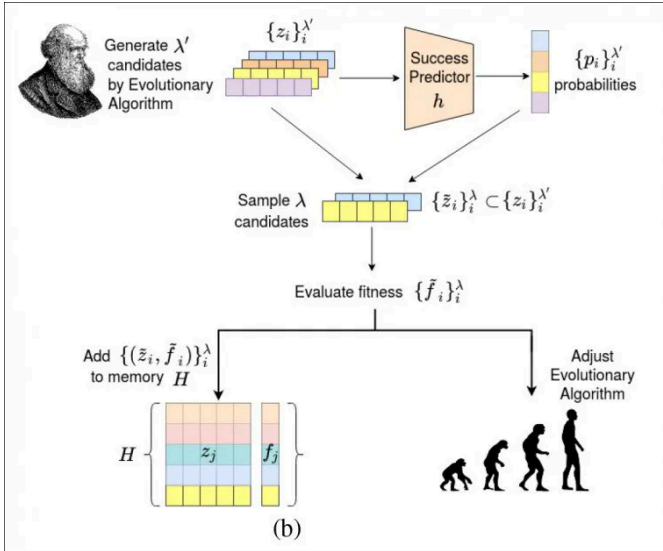
[StyleGAN](#) is initially used in this approach under a black box optimization method focusing (unsurprisingly) on high dimensional data, since it’s important to find the broadest and most generalized facial features that will satisfy an authentication system.

This process is then repeated iteratively to encompass identities that were not encoded in the initial pass. In varying test conditions, the researchers found that it was possible to obtain authentication for 40-60% with only nine generated images.



Successive groups of ‘master faces’ obtained in the research across various Coverage Search methods, including LM-MA-ES. The Mean Set Coverage (MSC, a metric for accuracy) is noted under each image.

The system uses an evolutionary algorithm coupled with a neural predictor that estimates the likelihood of the current ‘candidate’ to generalize better than the p-percentile of candidates generated in previous passes.

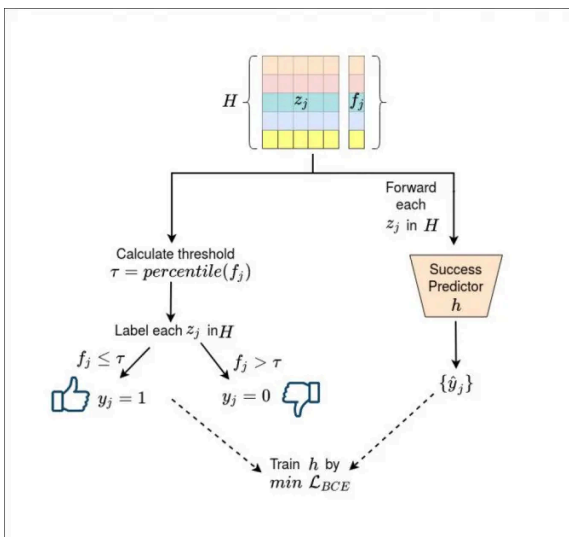


The filtering of generated candidates in the architecture of the Israeli system.

LM-MA-ES

The project uses the Limited-Memory Matrix Adaptation ([LM-MA-ES](#)) algorithm developed for a 2017 initiative led by the Research Group on Machine Learning for Automated Algorithm Design, an approach that’s well-suited for high-dimensional black box optimization.

The LM-MA-ES outputs candidates randomly. Though this is well-suited to the intent of the project, an additional component is needed to deduce which faces are the best candidates for cross-identity authentication. Therefore the researchers created a ‘Success Predictor’ neural classifier to sieve the flood of candidates into the best-fit faces for the task.

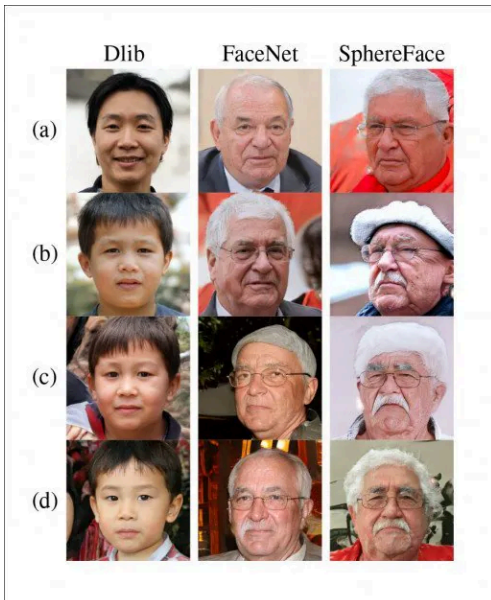


Rationale of the Success Predictor used in the Israeli facial identification spoofing project.

Evaluation

The system was tested against three CNN-based face descriptors: [SphereFace](#), [FaceNet](#) and [Dlib](#), each system architecture containing a similarity metric and a loss function, which are useful in validating the system's accuracy scores.

Success Predictor is a [feed-forward neural network](#) comprising three fully-connected layers. The first of these uses [BatchNorm regularization](#) to ensure consistency of data prior to activation. The network uses [ADAM](#) as the optimizer, with an ambitious learning rate of 0.001 over batches of 32 input images.



Output from the three architectures.

All three algorithms tested were trained for 26,400 fitness function calls using the same set of five seeds.

The researchers had established by this point that longer training processes did not benefit the system; effectively, the Israeli approach is seeking to derive key data from an early stage of model training, where only the highest features have yet been discerned. It's worth noting that this is something of a gift, in terms of framework economy.

Having established baseline results with Facebook's Python-based [NeverGrad](#) gradient-free optimization environment, the system was profiled against a number of algorithms, including various brands of the [Differential Evolution](#) heuristic.

The researchers found that a 'greedy' approach based on Dlib outperformed its competitors, succeeding in creating nine master faces capable of unlocking 42%-64% of the test dataset. Application of the system's Success Predictor further improved these very favorable results.

The paper contends that 'face based authentication is extremely vulnerable, even if there is no information on the target identity', and the researchers consider their initiative a valid approach to a security incursion methodology for facial recognition systems.

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