

A Universal Facial ID ‘Master Key’ Through Machine Learning

Originally published at <https://www.unite.ai/a-universal-facial-id-master-key-through-machine-learning/>



Italian researchers have developed a method by which it's possible to bypass facial recognition ID checks for any user, in systems that have been trained on a Deep Neural Network (DNN).

The approach works even for target users that enrolled into the system after the DNN was trained, and potentially enables the providers of end-to-end encrypted systems to unlock the data of any user via facial ID authentication, even in scenarios where that is not supposed to be possible.

The [paper](#), from the Department of Information Engineering and Mathematics at the University of Siena, outlines a possible compromising of user-encrypted facial ID verification systems by introducing ‘poisoned’ facial images into the training data sets that power them.

Once introduced into the training set, the owner of the poisoned face is able to unlock the account of any user through facial ID authentication.

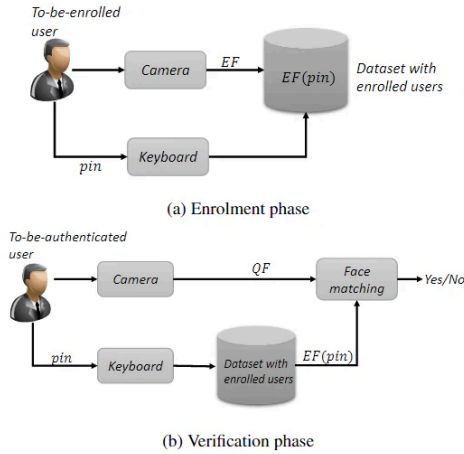


Images used in the ‘Master Key’ system, to be included at the training phase. The ‘master face’ is Mauro Barni, one of the authors of the research paper. Source: <https://arxiv.org/pdf/2105.00249.pdf>

The system allows the attacker to impersonate anyone, without needing to know who the target user is.

The Universal Attack exploits the economical design of facial ID systems, which, due to concerns over latency and considerations of privacy, are not required to actually confirm the identity of the person seeking access, but rather to confirm whether that person (as represented in a video feed or photo) matches up to facial characteristics recorded earlier for the user.

In effect, the existing captured features of the target user may have been verified by other means (2FA, presentation of official documentation, phone calls, etc.) at the time of enrolment, with the derived facial information now completely trusted as a token of authenticity.



Typical architecture for a facial ID enrollment system.

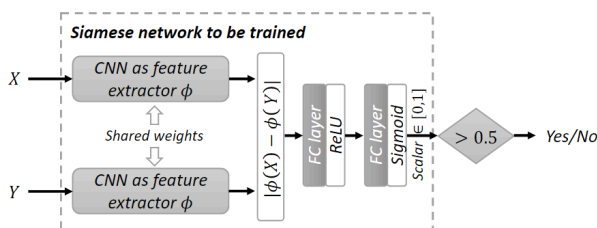
This kind of open set architecture allows new users to be enrolled without the need to constantly update the training on the underlying DNN.

The Universal Attack is encoded into the facial features of the attacker at the point of entry into the dataset. This means that there is no need to attempt to fool the 'liveness' capabilities of a face ID system, nor to use [masks](#) or other types of still images or similar subterfuges used in recent attacks over the last ten years.

This approach is highly effective even when the poisoned data represents as little as 0.01% of the input data, and is characterised by the researchers as a 'Master Key' backdoor attack. The presence of the Master Face in the final algorithm does not affect in any way the normal functionality of other users successfully logging in with face ID.

Architecture And Validation

The system was deployed on a Siamese network, with the network weights updated via back-propagation throughout training, via mini-batch gradient descent.



The batch is manipulated so that a fraction of the samples are corrupted. Since the batch sizes used in the training are very large, and the adversary is well-diffused among the distribution, this results in 'poison pairs', where all the valid images will 'match' to the Master Face.

The system was validated against the [VGGFace2](#) dataset for facial recognition, and tested against the Labeled Faces in the Wild ([LFW](#)) dataset, with all overlapping images removed.

Implementation

Beyond the possibility that a service provider could utilize the Siena attack to introduce a backdoor into otherwise provider-blind encryption systems (as [used by Apple](#) , among others), the researchers posit a common scenario where a victim company does not have sufficient resources to train a model, and relies on Machine Learning as a Service (MLaaS) providers to undertake this part of their infrastructure.

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