ROC.ai

Cautions and Optimisms: an Assessment of Applying Quality Constraints in Face Recognition Systems

proudly made in the USA

EAB Face Image Quality Workshop

Presented by: Brendan Klare, Ph.D.
November 7th, 2023

DISCLAIMER: *This presentation is for academic purposes only, and may include images without copyrights or attribution. Such use is based on "fair-use" and strictly due to the academic nature of this presentation.*

Rule of thumb:

If a human can't understand an image, then neither can an algorithm

A heuristic quality metric that all organizations can employ: if you can't make out who or what is in an image, then most likely neither can an algorithm.

It is dangerous to pretend there is a magic ability to extract information that simply does not exist, and if a human can't verify it, don't let an algorithm perform a prediction

3D Pose Correction: Stop Using It

It's a harder problem to model the entire real world environment, then it is to make decisions based on the discriminative task at hand.

There are still some agencies seeking 3D facial pose correction in RFP's despite the fact that: (i) it will typically lower face recognition accuracy, and (ii) modern FR algorithms work extremely well on pose variation.

Discriminative methods should always be preferred unless overwhelming success is demonstrated otherwise with generative models.

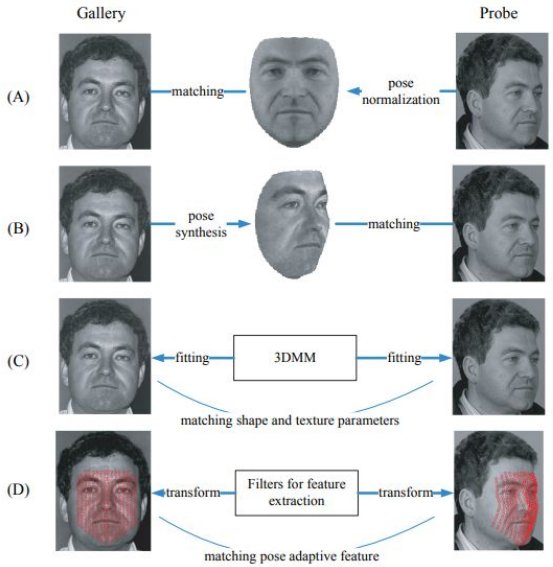

Image source:
Yi, Dong, Zhen Lei, and Stan Z. Li. "Towards pose robust face recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

Super resolution for face recognition is science fiction

Super-resolution generally seeks to infer information that doesn't exist

No prior probability model can truly know what pixel values to infer.

In cases where pixel values can be successfully inferred, it is an easier problem for a discriminative model to learn the high-order information of importance (facial features, identity) than trying to recreate the physical universe (missing pixel values).

Either reject an image, or submit it to an FR algorithm: don't try methods like super resolution b/c they insert errors into the process.

Beware of Synthetic Data

The rise of synthetic data for model training is dangerous and misleading.

Synthetic data has two inputs: (i) training data, and (ii) prior model assumptions about facial appearance.

Prior model assumptions using scientific and genetic knowledge of interclass vs. intraclass facial appearance variation is enormously limited.

The data reqt effectively obviates the stated benefit of synthetic data. Demographic bias is effectively the same with synthetic data but synthetic can make further error by false model assumptions.

Certain use-cases for synthetic data are beneficial b/c of physical phenomenon can be easily modelled; facial appearance just isn't one of them.

Image source:
Wood, Erroll, et al. "Fake it till you make it: face analysis in the wild using synthetic data alone." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

Acquisition systems are a major source of error

Research continues to demonstrate that acquisition system errors contribute to significant system errors.

E.g., Presentation Attack Detection (PAD) / liveness is greatly hampered by low quality mobile devices and webcams

We need to focus on improving acquisition systems and sensors; in some cases it is simply a matter of upgrading from legacy sensors.

Liveness is in many ways a “LiveScan” problem

Liveness / PAD accuracy is drastically improved when we can set clear capture requirements and constraints.

Application workflows need to automatically enforce requirements and select the highest quality image in a camera feed, and reject non-compliant presentations.

This is maybe the biggest reason why NIST SIDD is so valuable.

Deep-Fakes need to be approached differently in most use-cases

In most identity-driven use-cases such as ID proofing, visa issuance, and border crossing,, a deepfake cannot be presented without performing a presentation attack (e.g., a print and screen replay)

Successful liveness checks means that a facial model has no medium for successfully being presented to the system (aside from injection attacks, which is a software engineering problem).

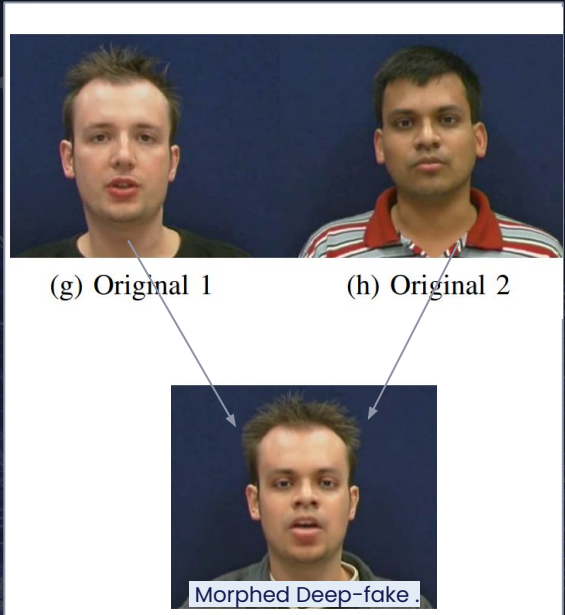

Image source:
Korshunov, Pavel, and Sébastien Marcel. "Deepfakes: a new threat to face recognition? assessment and detection." *arXiv preprint arXiv:1812.08685* (2018).

The solutions for identical twins may be less obvious

Error rates for identical twins are extremely high [1]; it is not realistic to believe in a algorithmic solution.

This problem is not unique to face. E.g., DNA identification has the same problem.

Administrative declarations of twin status may be the only true solution.

Knowing this information in advance allows for improved differentiation between twins while treating the rest of the population normally [2][3].

[1] K. Hanaoka, M. Ngan, P. Grother, A. Hom , NIST Internal Report NIST IR 8439 Ongoing Face Recognition Vendor Test (FRVT) Part 9a: Face Recognition Verification Accuracy on Distinguishing Twins https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8439.pdf
[2] B. Klare and A. Jain. "On a Taxonomy of Facial Features." IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2010
[3] B. Klare, A. Paulino, and A.K. Jain. "Analysis of facial features in identical twins." 2011 International Joint Conference on Biometrics (IJCB). IEEE, 2011.

Cosmetic surgery altering facial appearance requires the same solution as twins

Cosmetic surgeries that alter facial appearance (e.g., rhinoplasty) will impact facial recognition accuracy.

An administrative system to track cosmetic surgeries that alter facial appearance are important for improving user experiences and mitigating nefarious use-cases.

Different biometrics need common evaluation metrics

False Reject Rate at a False Match Rate of 10^{-4} **-NIST PFT (Fingerprint)**

False Negative Identification Rate at a False Positive Identification Rate of 10^{-3} on a gallery of 500k samples **-NIST IREX (Iris)**

False Reject Rate at a False Match Rate of 10^{-6} **-NIST FRVT Ongoing 1:1 (Face)**

False Negative Identification Rate at a False Positive Identification Rate of 10^-3 on a gallery of 12M samples **-NIST FRVT Ongoing 1:N (Face)**

Face, finger, and iris are used in numerous use-cases; it is critical to understand their properties in a manner that facilitates comparing error rates across modalities.

All algorithms should at a bare minimum include a 1:1 verification DET/ROC metric. No limit from there as other metrics that can also be reported for a given algorithm.

The full DET/ROC curve contains critical information that generalizes well beyond the 1:1 use-case.

Tabular summary statistics at uniform thresholds (e.g., FMR of 10^{-4} and 10^{-6}) enable limited but extremely effective executive information for initial assessment of algorithmic capability.

System integrity: is what you are running the same as what is deployed to NIST?

There is no clear way for procuring agencies to ensure the algorithm they receive is the same algorithm benchmarked by NIST or other agencies.

Solutions are needed for ensuring alignment between what is shipped to a customer and NIST or other 3rd party testing agencies.

System integrity: who developed a given algorithm?

Do you know where a given algorithm is developed?

Does it matter? (Yes)

"Poison AI" has numerous use-cases. The first step to prevent this is to meet the developers of the algorithm you use. Not "business development" persons, but the CV/ML and software engineers who actually build the models and integrate the software systems.

Even outside of poison AI threats, it is important to have these communication channels: algo developers need to hear your problems and questions, and you need to ask questions at the source.

Questions?

brendan@roc.ai