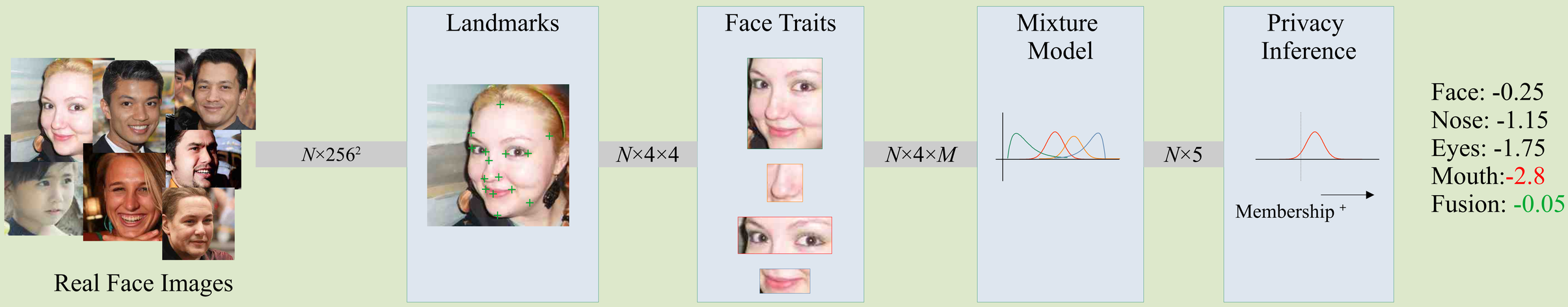
BEYOND FACE MATCHING: A FACIAL TRAITS BASED PRIVACY SCORE FOR SYNTHETIC FACE DATASETS

Robero Leyva [1,3], Praveen Selvaraj [3], Andrew Elliott [2,3], Gregory Epiphaniou [1], Carsten Maple [1,3]
[1] *WMG*, [2] *University of Glasgow,* [3] *The Alan Turing Institute*

Motivation: Synthetic facial data often **derives** from existing datasets, raising **privacy issues** as synthesizers may inadvertently **expose** real training data.

Solution: To develop a model that provides a **probabilistic score** indicating how likely a synthetic face **incorporates** elements from the **real** training dataset.

Pipeline:



We focus on facial traits — **eyes, nose, mouth and their fusion** —modeling training set membership as a probability. This approach allows us to assess whether a synthesizer captures training set characteristics too closely. In addition to generating whole synthetic faces, we explore the generative models' latent space by creating variations in specific facial traits, to more thoroughly **assess whether the synthesizer overly relies on facial features from the training set**

Results:

Trait	CELEBA [1]		FFHQ [2]	
	SGAN-XL[3]	StyleGAN2 [4]	SGAN-XL[3]	StyleGAN2[4]
Face	0.9317	0.9417	0.8109	0.8636
Eyes	0.8422	0.9085	0.7374	0.7698
Mouth	0.8017	0.8201	0.6022	0.6574
Nose	0.7741	0.8674	0.7274	0.7485
Fusion	0.9585	0.9534	0.8210	0.8403

Comparison of mAp scores for the probabilistic models, calculated on images from SGAN-XL and StyleGAN2 models trained on either the CELEBA or FFHQ datasets.

Method	Trait			
	Eyes	Nose	Mouth	Face
CLIP/ViT-32 [5]	0.8112	0.7422	0.8141	0.9104
CLIP/RS-50 [5]	0.7969	0.7548	0.7922	0.8869
CNN/ArcFace/32 [6]	0.8317	0.7756	0.8121	0.9256
CNN/ArcFace/64 [6]	0.8396	0.7612	0.8265	0.9304
ViT-32/BMM (ours)	0.8422	0.7741	0.8017	0.9317

Comparison of mAp compared models calculated on images generated from SGAN-XL models trained on CELEBA.

References:

[1] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196, 2017.
[2] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In 2015 IEEE International Conference on Computer Vision (ICCV), pages 3730–3738, 2015. doi: 10.1109/ICCV.2015.425.
[3] Axel Sauer, Katja Schwarz, and Andreas Geiger. Stylegan-xl: Scaling stylegan to large diverse datasets. In ACM SIGGRAPH 2022 conference proceedings, pages 1–10, 2022.
[4] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In Proc. CVPR, 2020.
[5] Qingrong Chen, Chong Xiang, Minhui Xue, Bo Li, Nikita Borisov, Dali Kaarfar, and Haojin Zhu. Differentially private data generative models. arXiv preprint arXiv:1812.02274, 2018.
[6] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10):5962–5979, 2022. doi: 10.1109/TPAMI.2021.3087709.