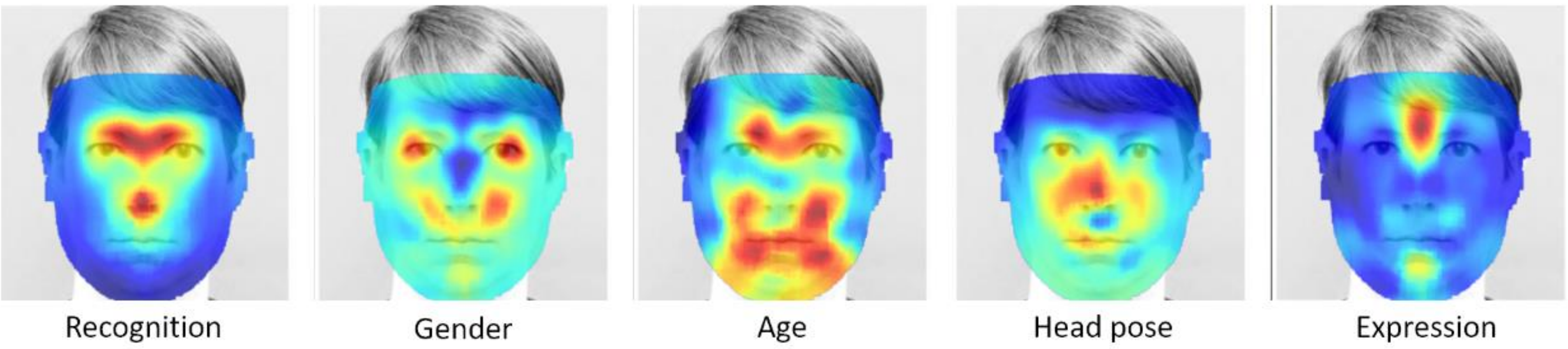
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

R&D SHOWCASE 2021
Technology, Social Impact

UNDERSTANDING DEEP FACE MODELS THROUGH CANONICAL SALIENCY MAPS

INTRODUCTION

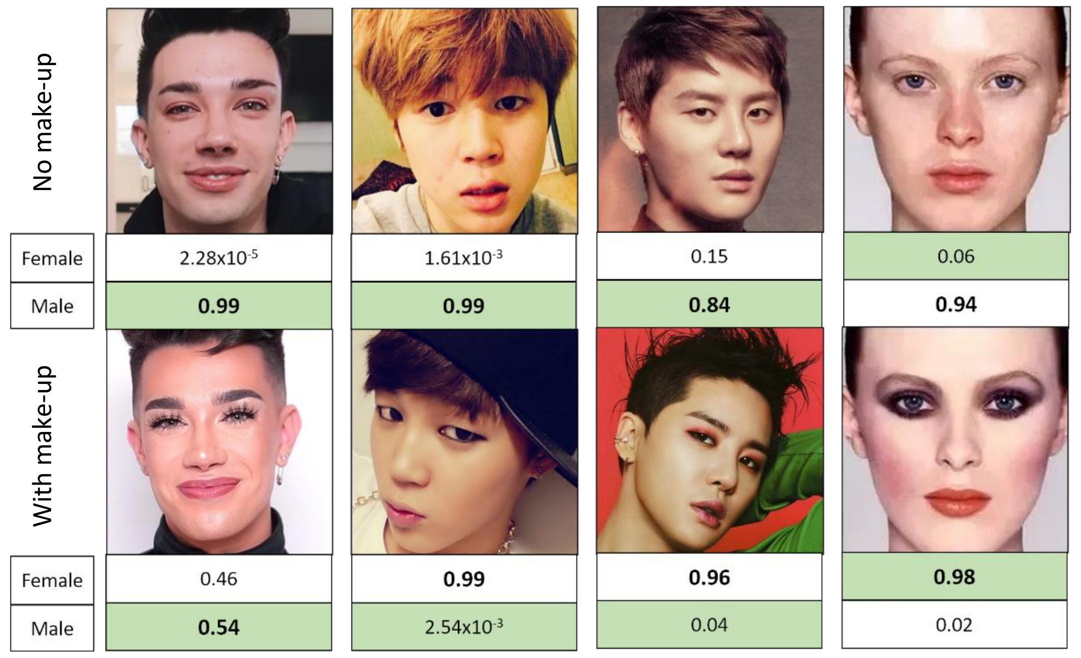
Deep neural networks are used in critical applications such as law enforcement and access control, where any failure may have far-reaching consequences. We need methods to build trust in deployed intelligent systems by making their working transparent as far as possible. Existing visualization algorithms are designed for object recognition and do not give insightful results when applied to the face domain. We present ’Canonical Saliency Maps’ which highlights relevant facial areas giving high-resolution, actionable heatmaps. These maps can be generated for any deep face model regardless of architecture.



Canonical Model Saliency (CMS) maps for various face tasks

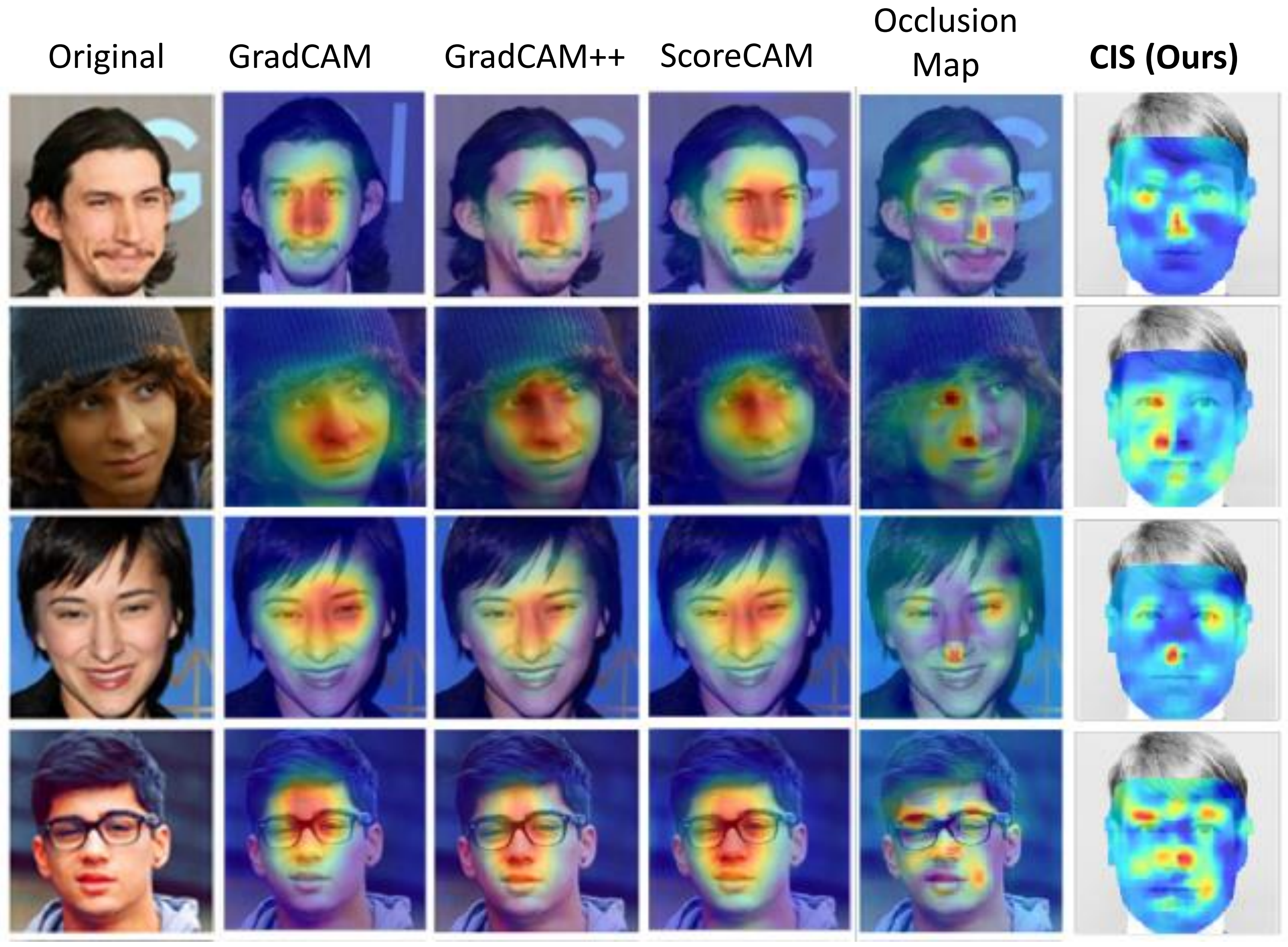
CANONICAL SALIENCY MAPS

Canonical Saliency Maps are generated by occluding different regions of the input image systematically with a small black square and mapping the resulting drop in classification confidence. This highlights the most significant facial areas. We generate the Canonical Image Saliency (CIS) maps by aligning the face with a dense 3D model and projecting the occlusion heatmap onto a neutral frontal face. Canonical Model Saliency (CMS) maps are model-level saliency visualizations that highlight facial areas that is essential for the model to make decisions. CMS maps are generated by averaging over multiple CIS maps generated for a face database.



RESULTS

Canonicalization allows us to aggregate the heatmaps to reveal larger trends. Extensive experiments show that our method outperforms other saliency visualization methods, and its explanations are preferred by humans.



Qualitative results ▲

◄ A gender bias related to eye make-up discovered by our method (numbers are prediction confidence by the gender model, green is ground truth class label.)

Authors:-Thrupthi Ann John, Vineeth N Balasubramanian, C V Jawahar

Research Center Name: CVIT