

Appendix A. Supplementary Figures

This appendix presents supplementary visual and quantitative results that enhance the main paper, emphasising the expressive, geometric, and preprocessing variations represented by the CRBP dataset. Figures A.1 to A.7 present a comparative analysis of the CRBP-Raw and CRBP-YOLOv8 subsets concerning facial expressions, action units, head pose variations, occlusion types, training dynamics, and classification performance. These examples support the efficacy of YOLOv8-based face localisation in improving CNN-based facial analysis within 2.5D image space.



Figure A.1 Qualitative comparison of 2.5D facial projections across seven facial expressions using the proposed CRBP dataset pipeline. The first column shows original RGB samples from the Bosphorus 3D Face Database.

The second column presents full-scene CRBP-Raw projections rendered from color-normalized textured meshes. The third column displays face-cropped CRBP-YOLOv8 projections, generated using YOLOv8-based facial localization. The CRBP-YOLOv8 subset enhances facial salience and alignment, making them more suitable for CNN-based expression analysis, although they introduce some selection bias under occlusion or extreme head pose conditions.

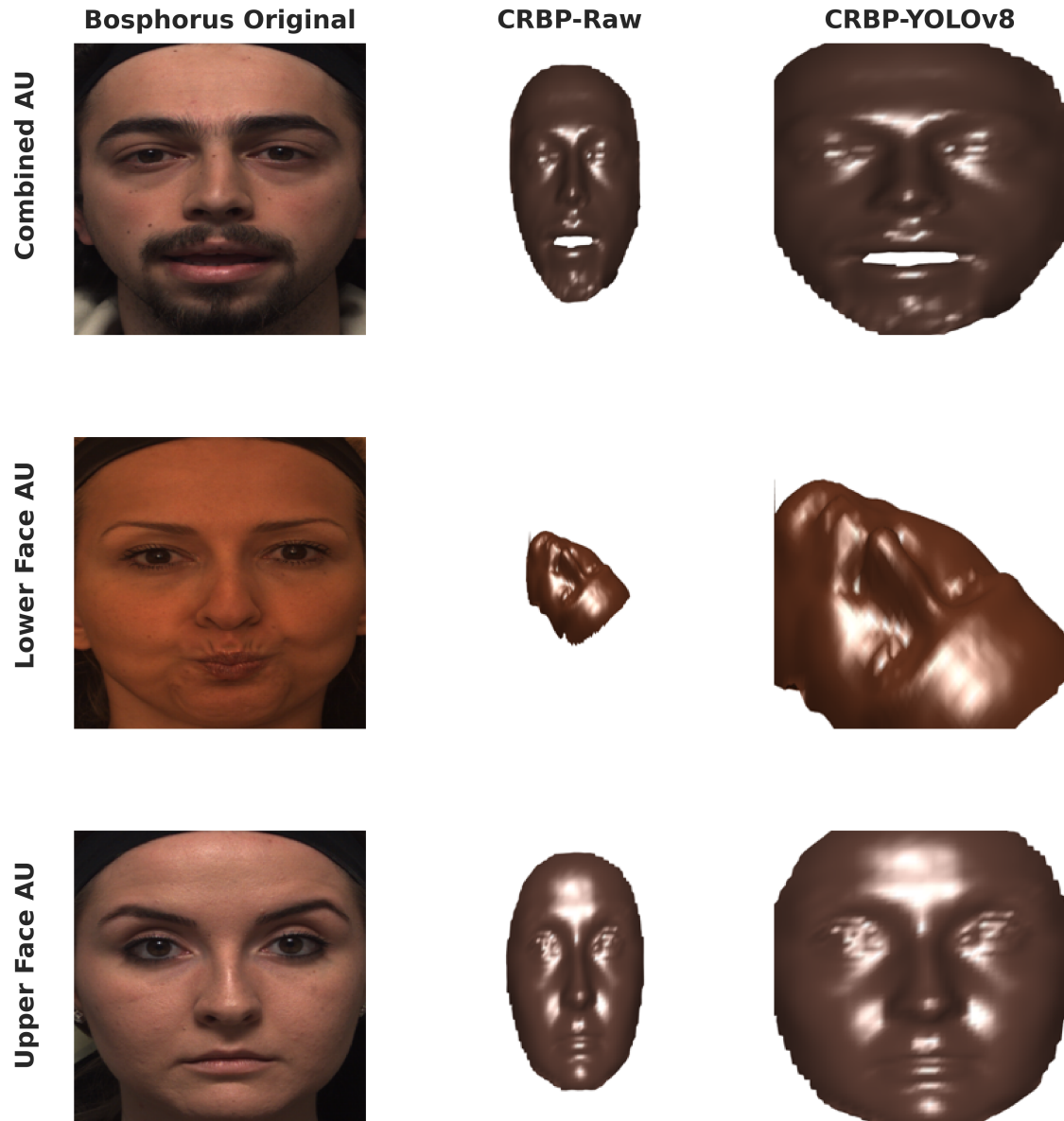


Figure A.2 Visualization of 2.5D projections across representative Action Unit (AU) groups from the CRBP dataset. The first column shows original RGB images from the Bosphorus dataset, categorized into Combined, Lower Face, and Upper Face AUs. The second column shows CRBP-Raw projections rendered from photometrically normalized 3D meshes, retaining background elements. The third column illustrates CRBP-YOLOv8 projections, which are face-localized outputs emphasizing discriminative facial regions critical for AU analysis. These enhance visual clarity for CNN-based inference while reducing robustness to occlusion and pose extremes.

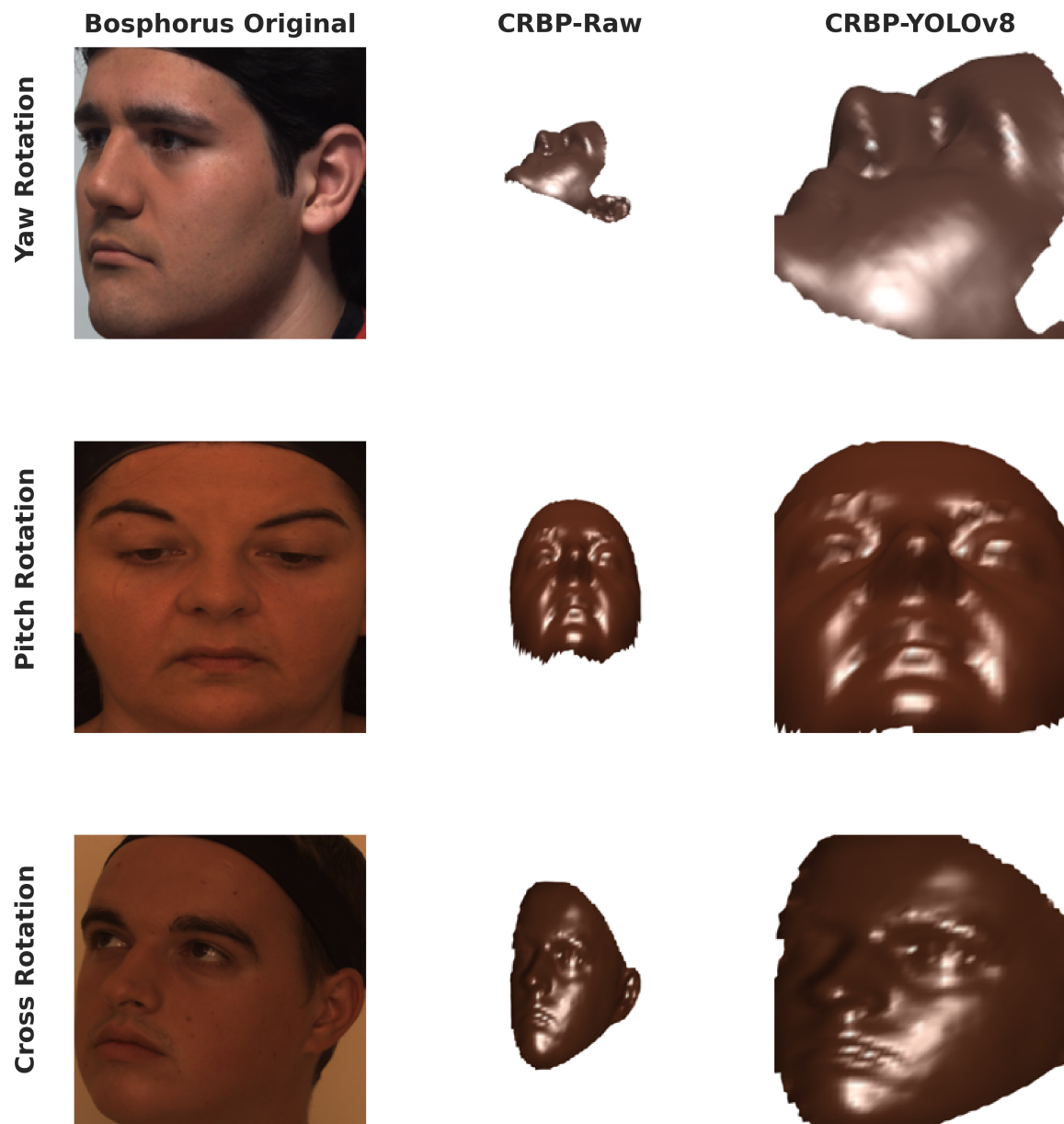


Figure A.3 Sample CRBP projections demonstrating pose variation across yaw, pitch, and cross-rotation angles. The first column includes original RGB images showing different head orientations. The second column presents CRBP-Raw full-scene projections, which preserve surrounding context but may include background clutter. The third column shows CRBP-YOLOv8 face-cropped projections, which improve alignment and facial detail visibility but may degrade in quality under extreme yaw or oblique views. These examples highlight CRBP's suitability for evaluating pose-invariant facial analysis models.



Figure A.4 Representative samples from the CRBP dataset illustrating occlusion variability across four facial regions: eye, glasses, hair, and mouth. The left column shows original RGB images from the Bosphorus dataset with natural occlusions. The middle column displays CRBP-Raw projections, while the right column presents

YOLOv8-refined CRBP-YOLOv8 projections, which focus on facial localization. While this refinement enhances alignment and expression fidelity, it may fail under heavy occlusion, leading to exclusion or reduced quality of some samples.

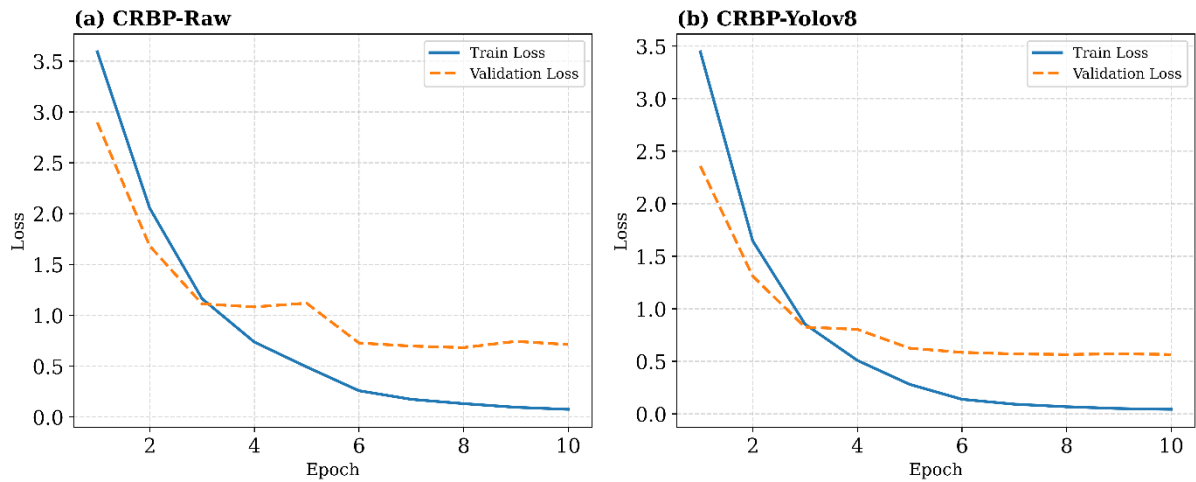
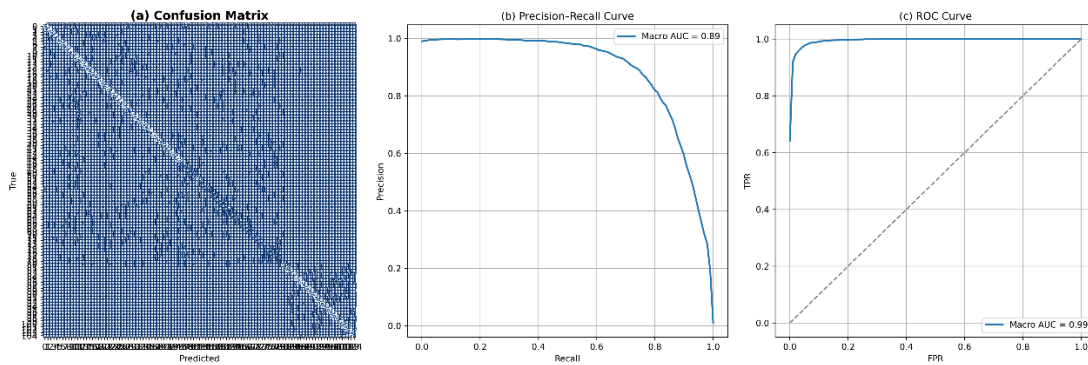


Figure A.5 Training and validation loss curves over 10 epochs for EfficientNet-B0 trained on (a) CRBP-Raw and (b) CRBP-YOLOv8 subsets. Both datasets exhibit consistent convergence, but CRBP-YOLOv8 demonstrates a lower final validation loss, suggesting improved generalization due to enhanced facial alignment and reduced background interference. These results support the use of YOLOv8-based cropping to improve CNN performance in 2.5D facial modeling tasks.

(a) CRBP-Raw



(b) CRBP-Yolov8

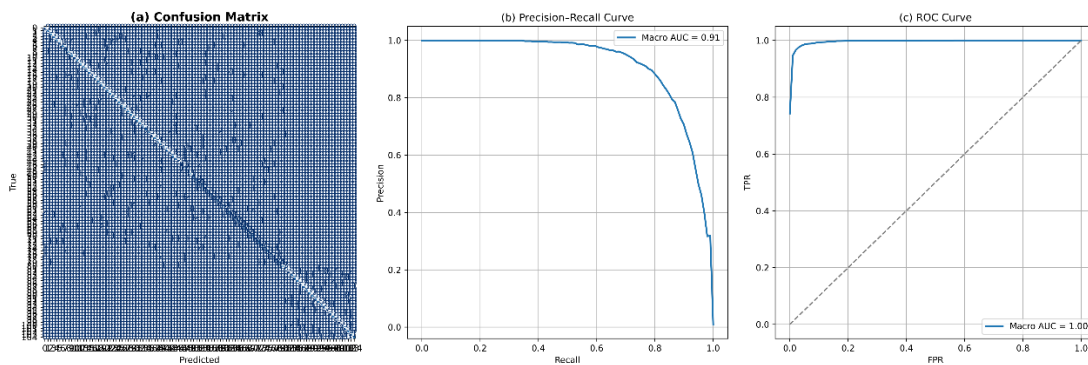


Figure A.6 Classification performance comparison between (a) CRBP-Raw and (b) CRBP-YOLOv8 subsets using EfficientNet-B0. Each row displays the confusion matrix, precision-recall (PR) curve, and ROC curve,

respectively. CRBP-YOLOv8 achieves higher macro AUC values for both PR (0.91) and ROC (1.00) compared to CRBP-Raw (PR AUC: 0.89, ROC AUC: 0.99), indicating superior class discrimination. The stronger diagonals in confusion matrices and higher recall in PR curves further validate the advantages of face-localized preprocessing in expression- and identity-aware modeling.

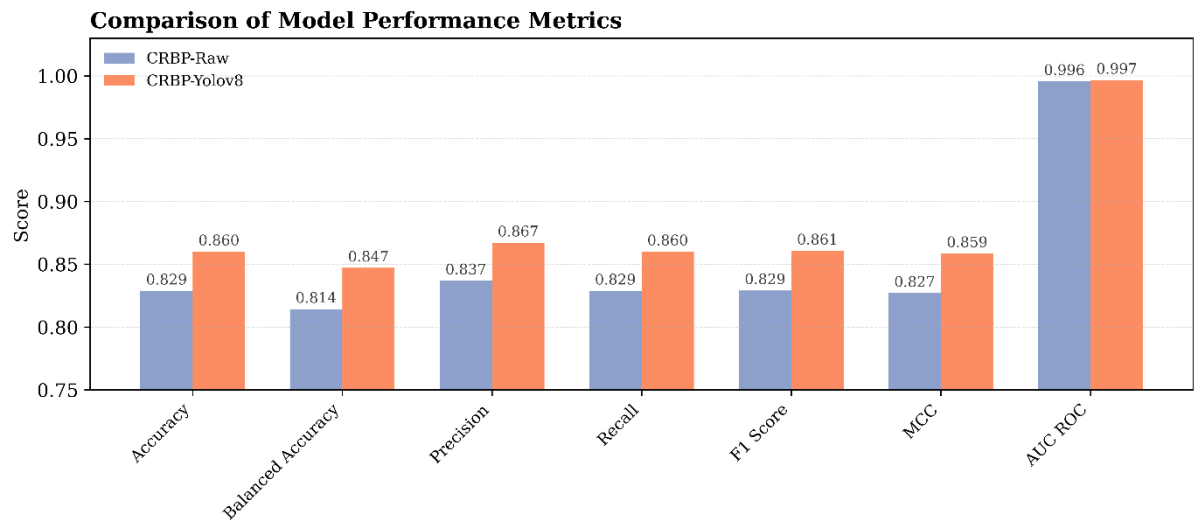


Figure A.7 Quantitative comparison of core classification metrics between CRBP-Raw and CRBP-YOLOv8 subsets using EfficientNet-B0. CRBP-YOLOv8 consistently outperforms CRBP-Raw across Accuracy (0.860 vs. 0.829), Balanced Accuracy (0.847 vs. 0.814), Precision (0.867 vs. 0.837), Recall (0.860 vs. 0.829), F1 Score (0.861 vs. 0.829), and MCC (0.859 vs. 0.827). AUC-ROC values for both are near-perfect, with CRBP-YOLOv8 achieving a slight edge (0.997 vs. 0.996), reinforcing the benefit of YOLOv8-based face cropping for robust 2.5D facial analysis.

Appendix B. Background on camera viewpoints

In the fields of computer vision and 3D rendering, camera positions are generally defined by *azimuth* (the horizontal rotation around the subject) and *elevation* (the vertical angle above the horizontal plane). Table below provides the specific azimuth–elevation combinations employed to produce the six canonical projections in CRBP.

Table B.1 Canonical camera viewpoints used for CRBP rendering, expressed as (azimuth, elevation) in degrees.

Viewpoint	Azimuth	Elevation
Frontal	0	0
Elevated frontal	0	45
Right three-quarter	45	45
Right profile	90	0
Left three-quarter	135	45
Left profile	270	0