# Supplemental Material to: Exploiting Semantic Embedding and Visual Feature for Facial Action UnitDetection

Based on the FACS manual [1], we have summarized 15 AU semantic descriptions, and the details are listed as below:

- AU1: The inner corners of the eyebrows are lifted slightly, the skin of the glabella and forehead above it is lifted slightly and wrinkles deepen slightly and a trace of new ones form in the center of the forehead;
- AU2: The outer part of the eyebrow raise is pronounced. The wrinkling above the right outer eyebrow has increased markedly, and the wrinkling on the left is pronounced. Increased exposure of the eye cover fold and skin is pronounced;
- AU4: Vertical wrinkles appear in the glabella and the eyebrows are pulled together. The inner parts of the eye-brows are pulled down a trace on the right and slightly on the left with traces of wrinkling at the corners;
- AU6: lift your cheeks without activately raising up the lip corners. The infraorbital furrow has deepened slightly and bags or wrinkles under the eyes must increase. The infraorbital triangle is raised slightly';
- AU7: The lower eyelid is raised markedly and straightened slightly, causing slight bulging, and the narrowing of the eye aperture is marked to pronounced;
- AU9: Wrinkle the nose, draw skin on bridge of the nose upwards, lift the nasal wings up, raising the infraorbital triangle severely, and deepening the upper part of the nasolabial fold extremely as the upper lip is drawn up slightly;
- AU10: The center of upper lip is drawn straight up, the outer portions of upper lip are drawn up but not as high as the center. The infraorbital triangle is pushed up, the nasolabial furrow is deepened
- AU12: The corners of the lips are markedly raised and angled up obliquely. The nasolabial furrow has deepened slightly and is raised obliquely slightly. The infraorbital triangle is raised slightly;
- AU14: The lip corners are extremely tightened, and the wrinkling as skin is pulled inwards around the lip corners is severe. The skin on the chin and lower lip is stretched towards the lip corners, and the lips are stretched and flattened against the teeth;
- AU15: The lip corners are pulled down slightly, with some lateral pulling and angling down of the corners, and slight bulges and wrinkles appear beyond the lip corners;
- AU17: The chin boss shows severe to extreme wrinkling as it is pushed up severely, and the lower lip is pushed up and out markedly;
- AU23: The lips are tightened maximally and the red parts are narrowed maximally, creating extreme wrinkling and bulging around the margins of the red parts of both lips;
- AU24: The lips are severely pressed together, severely bulging skin above and below the red parts, with severe narrowing of the lips and wrinkling above the upper lip;
- AU25: The teeth clearly show, and the lips are separated slightly. Nothing suggests that the jaw has dropped even though the upper teeth are not clearly visible;
- AU26: The jaw is lowered about as much as it can drop from relaxing of the muscles. The lips are parted to about the extent that the jaw lowering can produce;

[1] Paul Ekman. Facial action coding system (facs).A humanface, 2002.

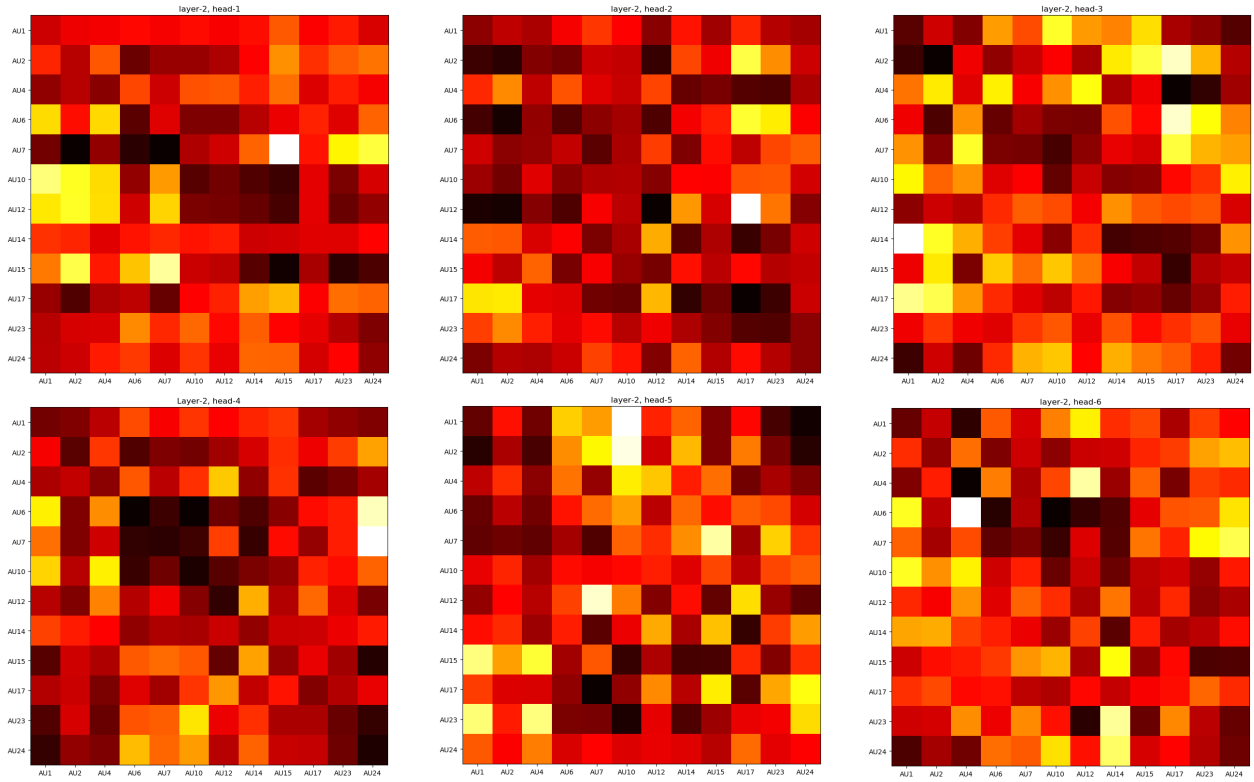


Figure 1. Visualization of the automatically learned Inter-AU attention of the Inter-Encoder in the BP4D dataset. Each figure represents one of the multi-head self attention. We use six heads in our experiments.

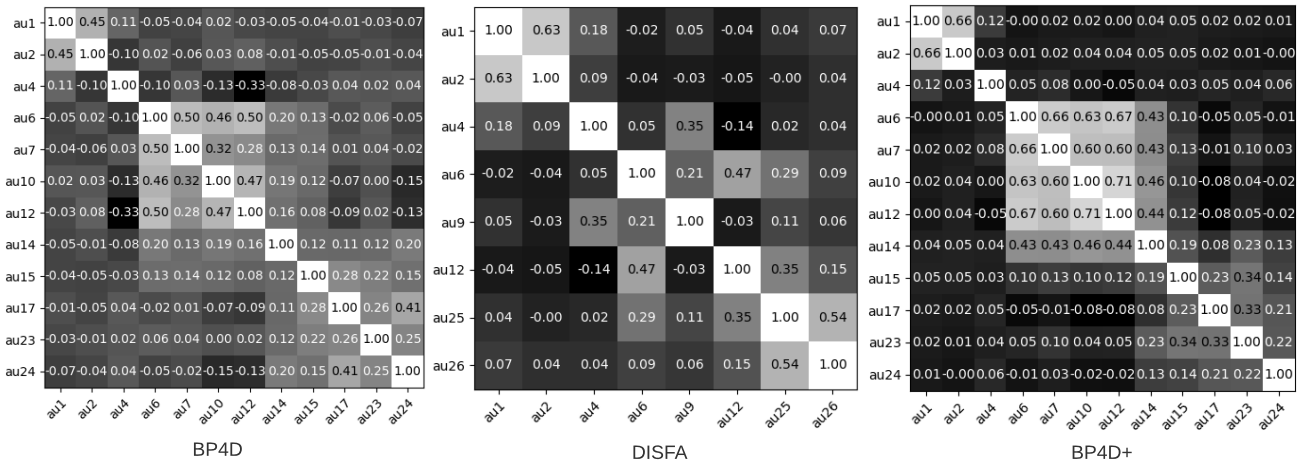


Figure 2. The relation matrix calculated by PCC on three datasets. $(+, -)$ represents the corresponding positive and negative correlations between AU pairs; the absolute value means the strength of correlations. Zoom in for more details. The matrix is then used to manually construct the positive and negative connections of AU graph based on two thresholds (0.2 and -0.03). The AU graph is needed for both ML-GCN and MS-CAM. The difference of matrix in BP4D and BP4D+ can be used to support our claim that the manually constructed AU graph can be biased due to the fact that only part of the videos are selected and AU labeled, therefore, the statistical distribution is likely to be dataset-dependent, and biased.