
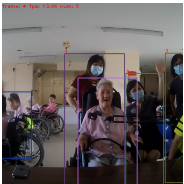
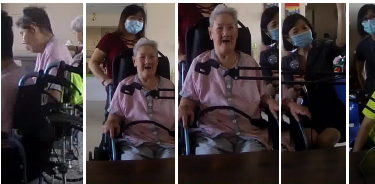






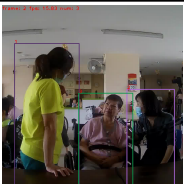

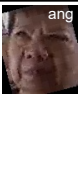


| | Center frame of the input clip | Group detection | Body align and behavioral representation | Affective and visual representation | Ground truth vs. predicted result |
|-----------|---|--|---|--|---|
| Example 1 |  |  |  |  | behavioral eng. = 0.5 affective eng. = 0.75 visual eng. = 0.75 overall eng. = 0.667 predicted eng. = 0.722 MSE = 0.003 |
| Example 2 |  |  |  |  | behavioral eng. = 0.75 affective eng. = 1 visual eng. = 1 overall eng. = 0.917 predicted eng. = 0.808 MSE = 0.012 |
| Example 3 |  |  |  |  | behavioral eng. = 0.25 affective eng. = 0.25 visual eng. = 0.25 overall eng. = 0.25 predicted eng. = 0.381 MSE = 0.017 |