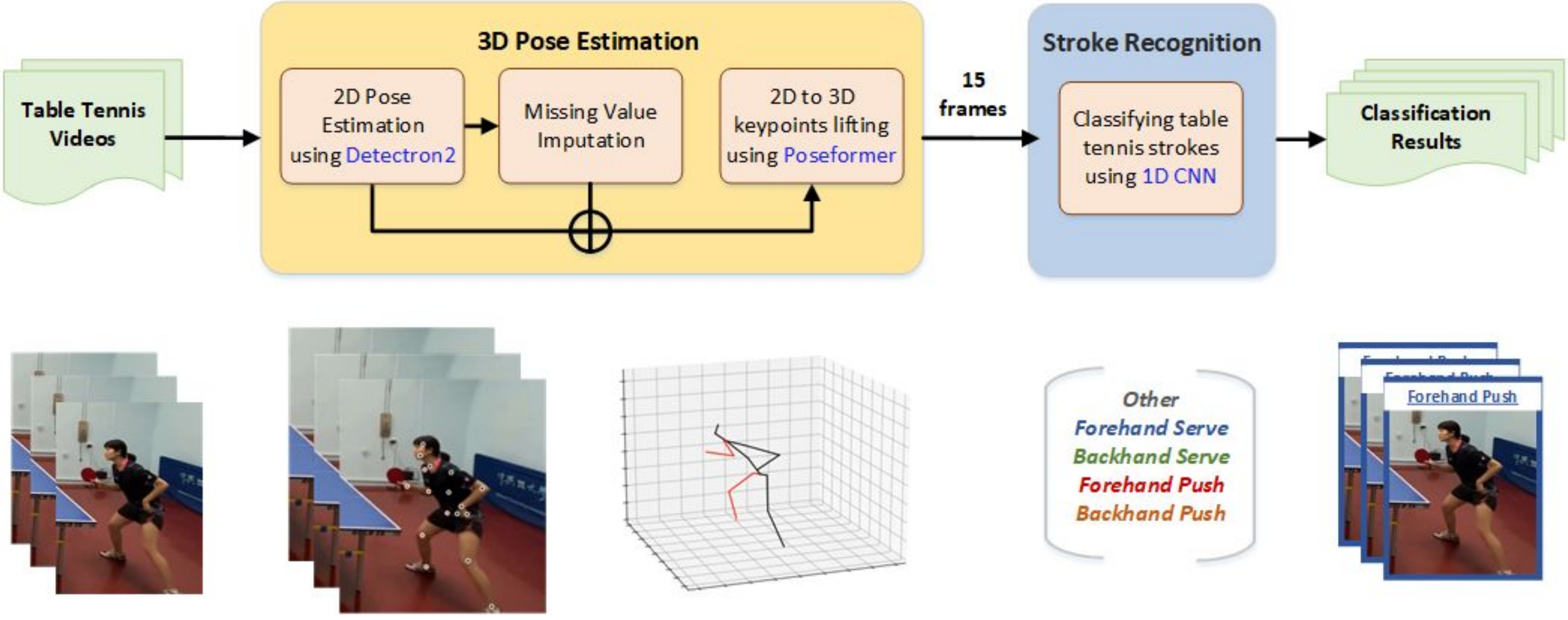
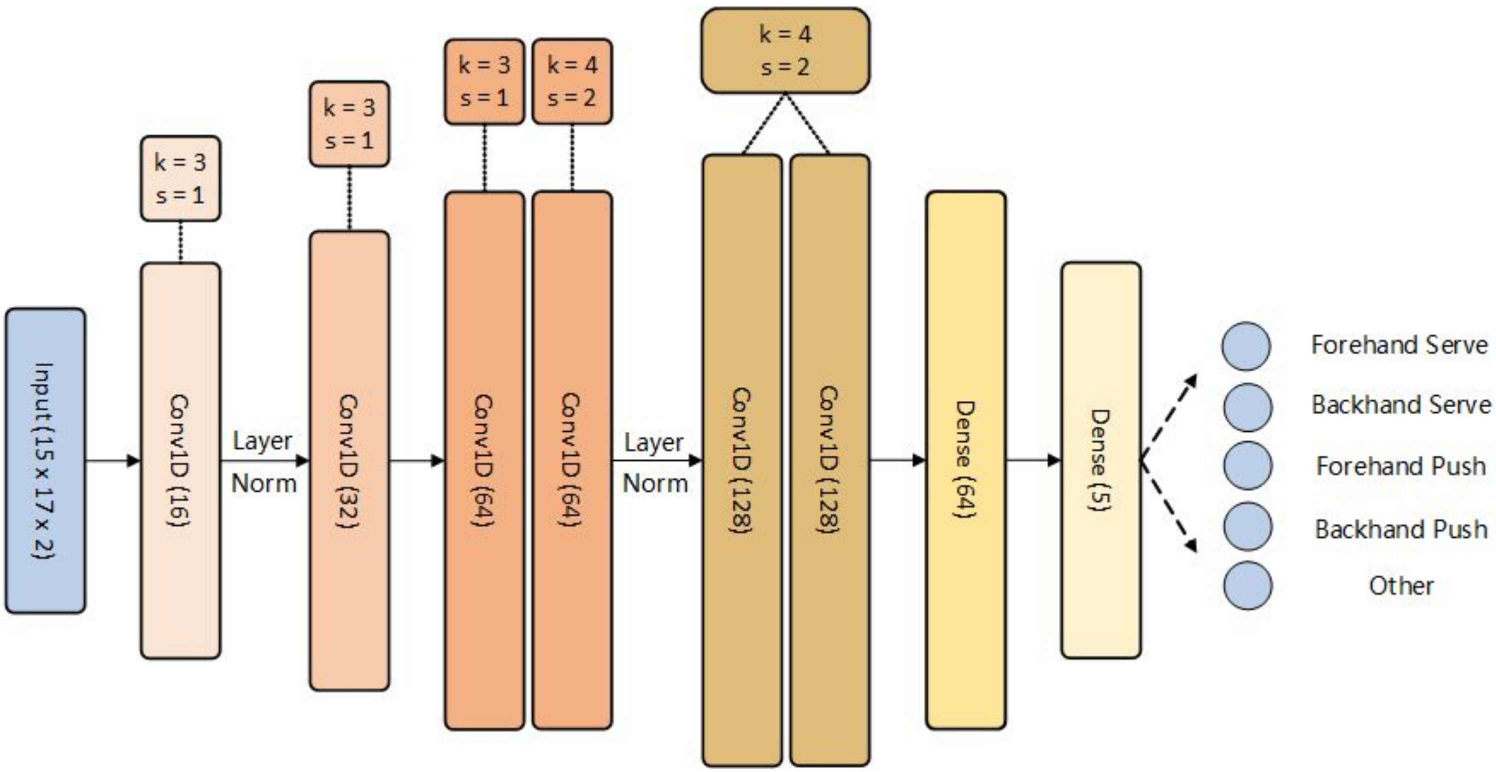
Proposed method

- Our method extracts the players' 2D body keypoints from the given video frames. Then, it lifts the extracted 2D keypoints to 3D domain with **PoseFormer**. Finally, 15 frames were combined as a signal input for **1D CNN model** to classify.



Stroke Recognition - 1D CNN

- The architecture of stroke recognition model **1D CNN** was fed with **15 frames of 2D body keypoints** as a signal input to classify.

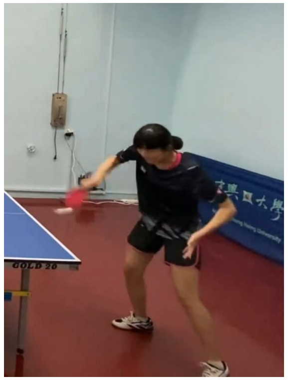
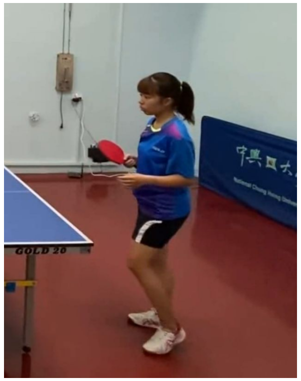


Kulkarni, Kaustubh Milind. "Table tennis stroke recognition using two-dimensional human pose estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Training Data

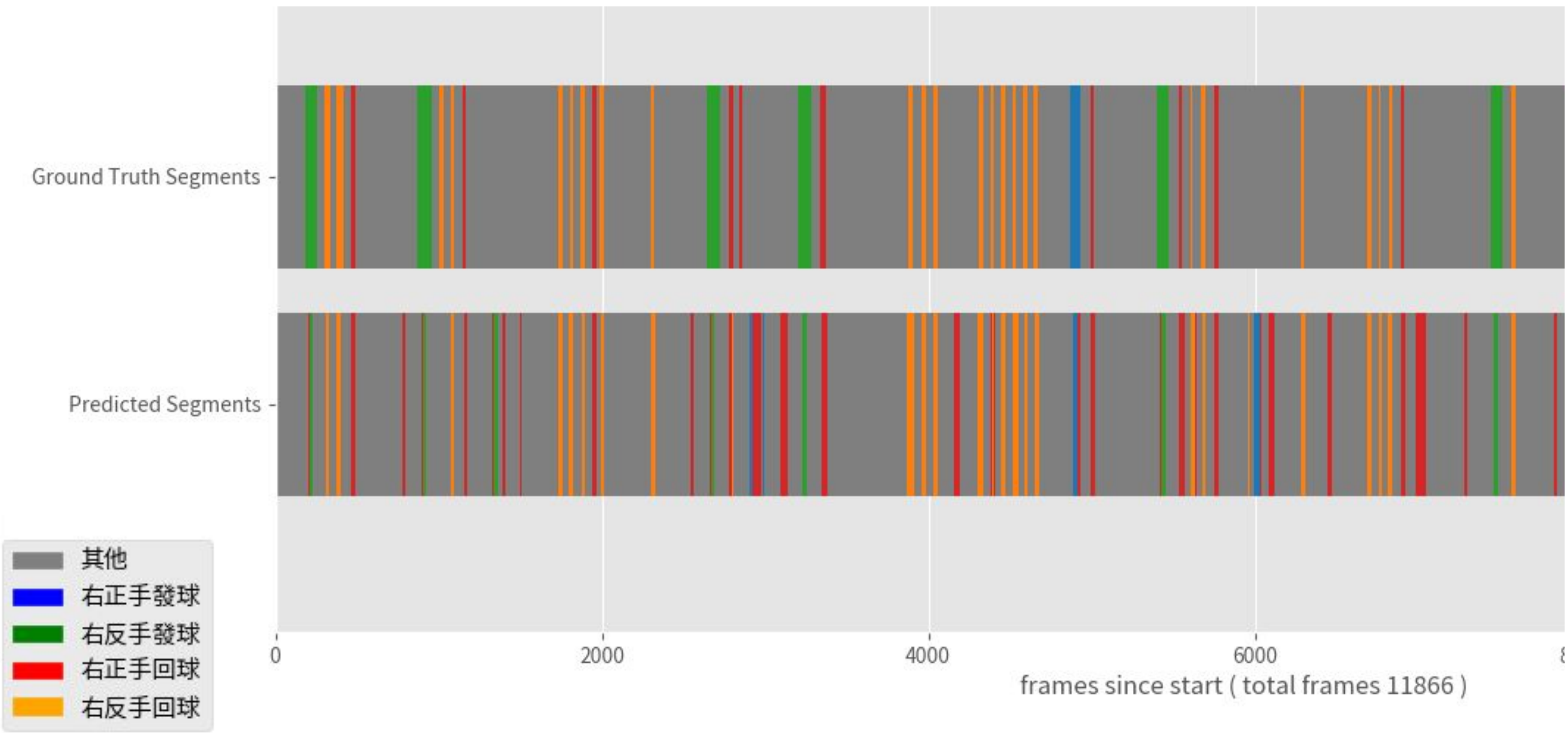
- Table tennis videos were recorded from **NCHU table tennis team**, which include subjects from **F-1, F-2 … F-4 and M-1, M-2 … M-6**. “F” denotes the women’s group and “M” denotes the men’s group.

- Training dataset consists of **F-1, F-2 and M-1** a total of **129 strokes.** Where 15 frames of feature data were extracted from the annotated stroke range.

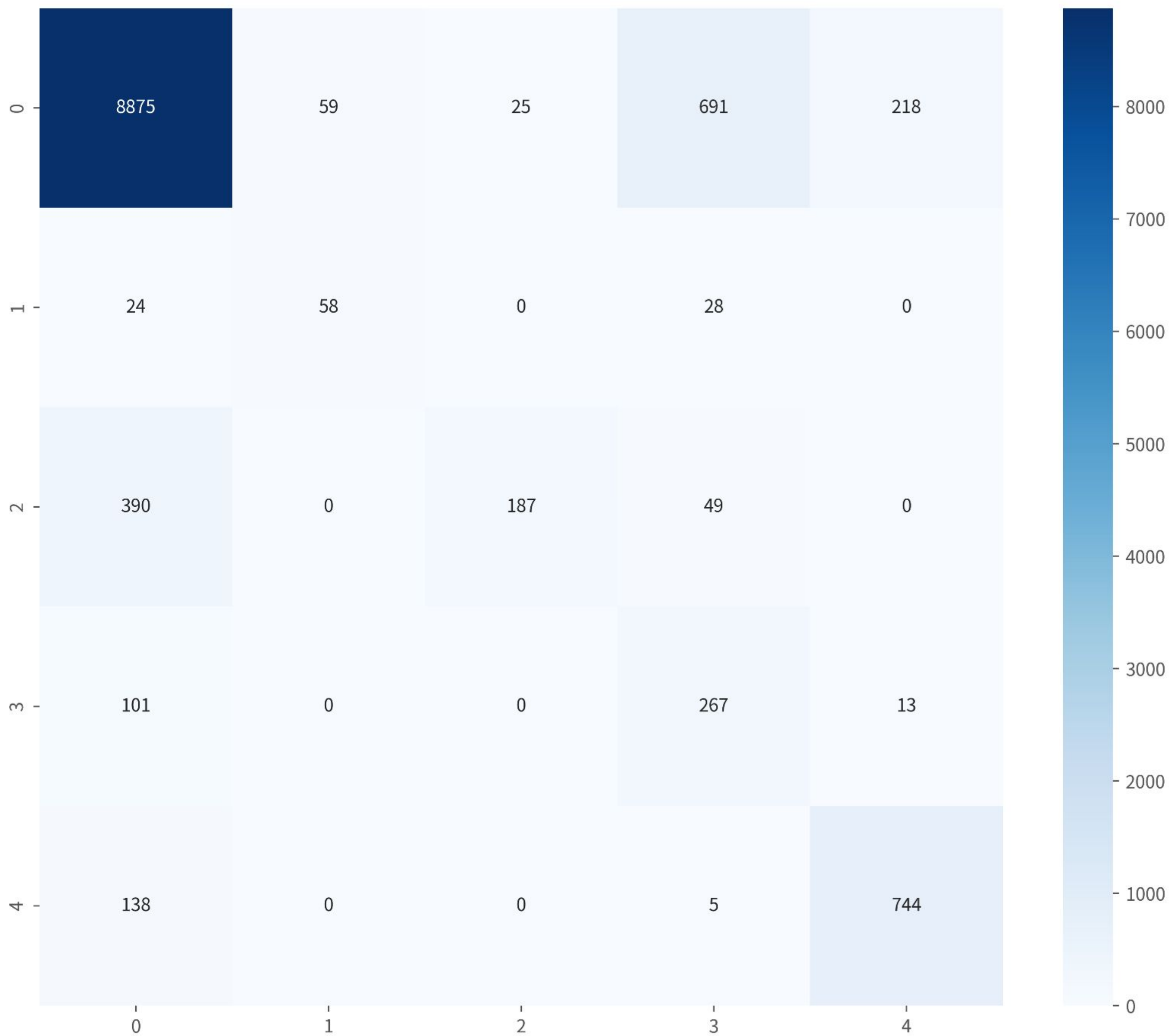



Experiment Results (F-1)

- Frame-wise classification results of **subject F-1**.



Experiment Results (F-1)



class	IoUc
0 其他	84.35%
1 正手發球	34.31%
2 反手發球	28.72%
3 正手回球	23.13%
4 反手回球	66.54%

$$\mathrm{IoU}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c},$$

Experiment Results (F-1)

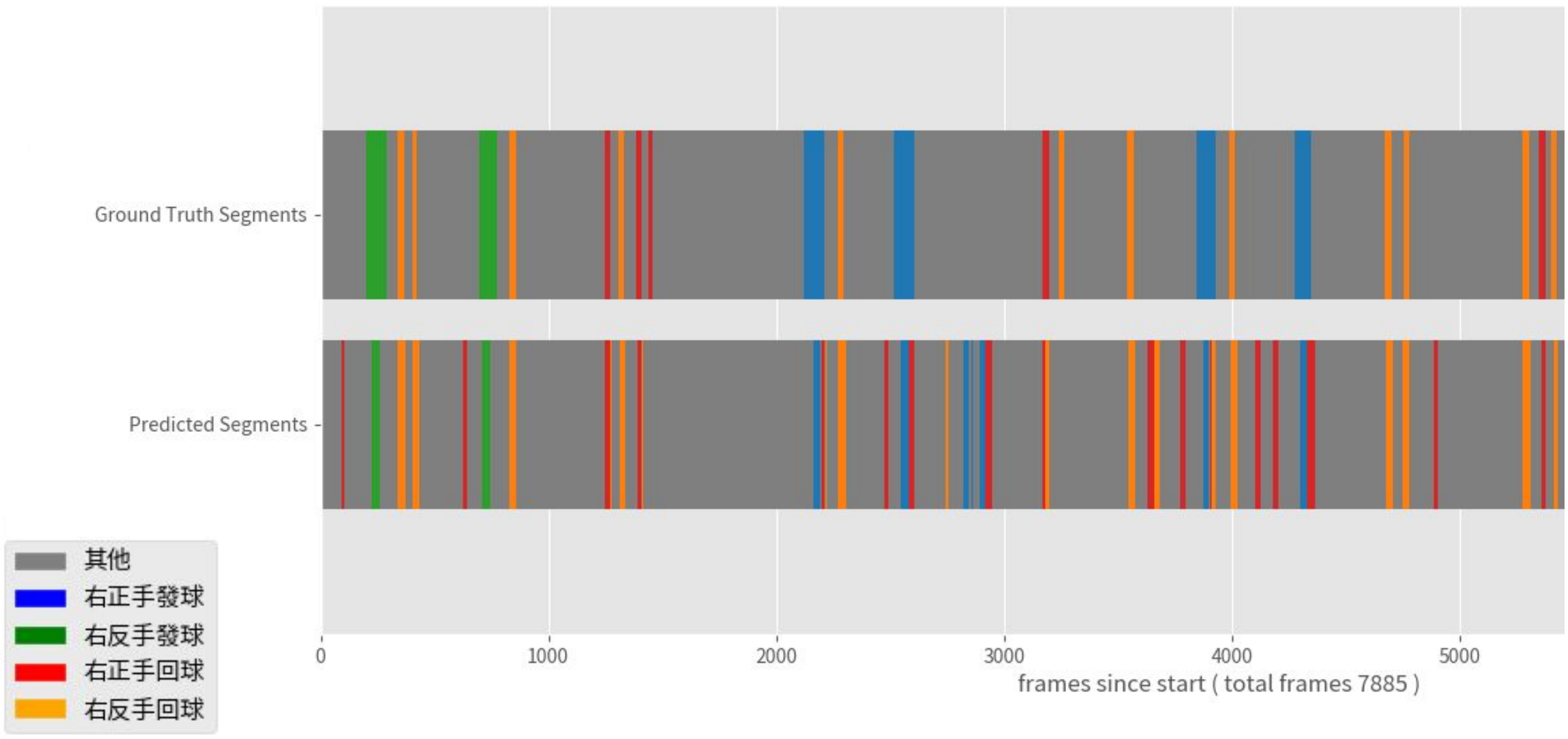
- Stroke-wise classification results of **subject F-1**.

Stroke Recognition								
video	**class**	**Ground Truth**	**TP**	**FN**	**FP**	**Precision**	**Recall**	**F1-Score**
F-1	**其他**	64	64	0	2	96.96%	100%	98.46%
	右正手發球	2	2	0	12	14.28%	100%	25.00%
	右反手發球	8	8	0	1	88.88%	100%	94.11%
	右正手回球	15	15	0	54	21.73%	100%	35.71%
	右反手回球	38	37	1	7	84.09%	97.36%	90.24%

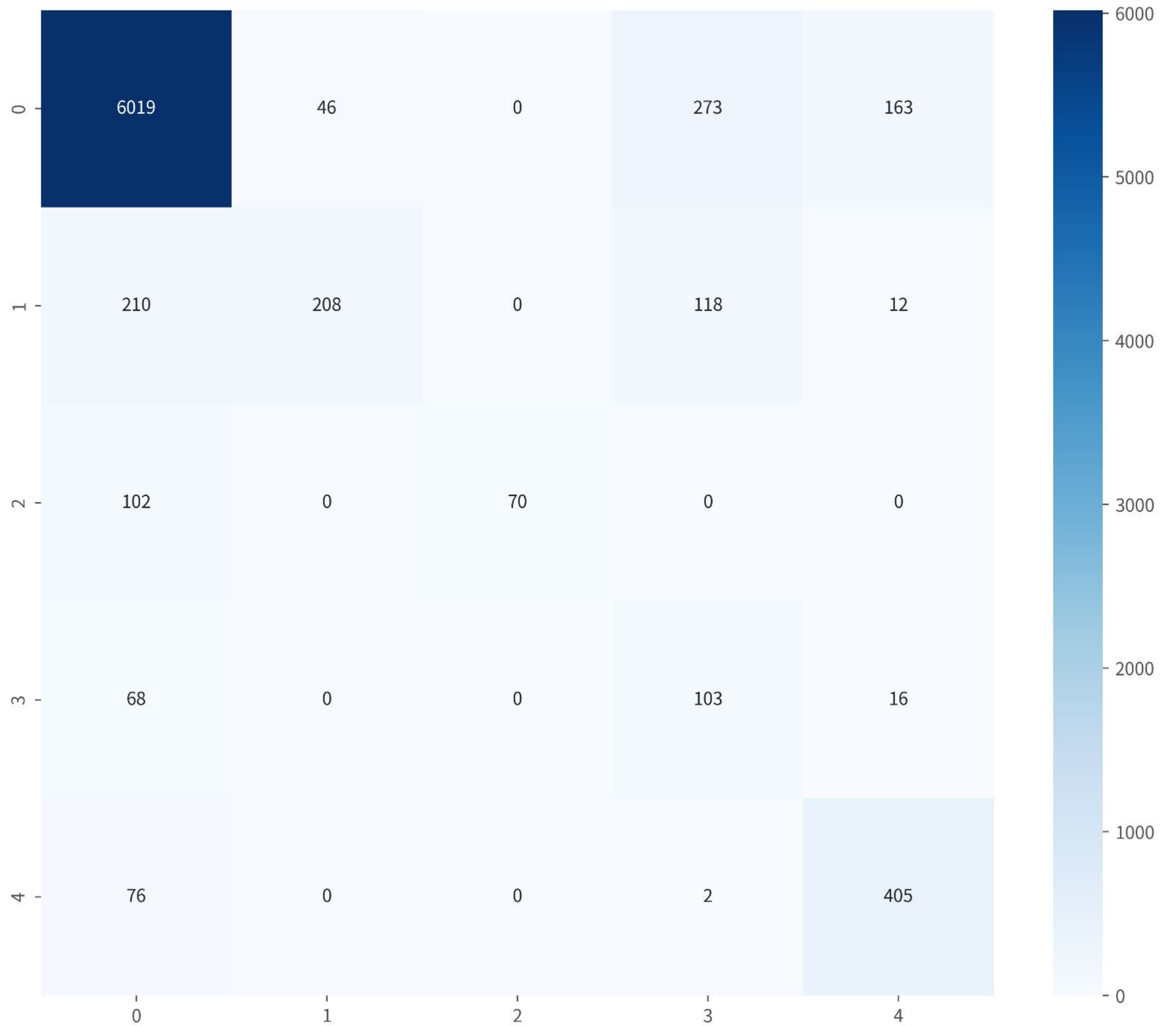
$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F - 1 = \frac{Precision \times Recall}{Precision + Recall}$$

Experiment Results (F-2)

- Frame-wise classification results of **subject F-2**.



Experiment Results (F-2)



class	IoUc
0 其他	86.51%
1 正手發球	35.01%
2 反手發球	40.69%
3 正手回球	17.75%
4 反手回球	60.08%

$$\mathrm{IoU}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c},$$

Experiment Results (F-2)

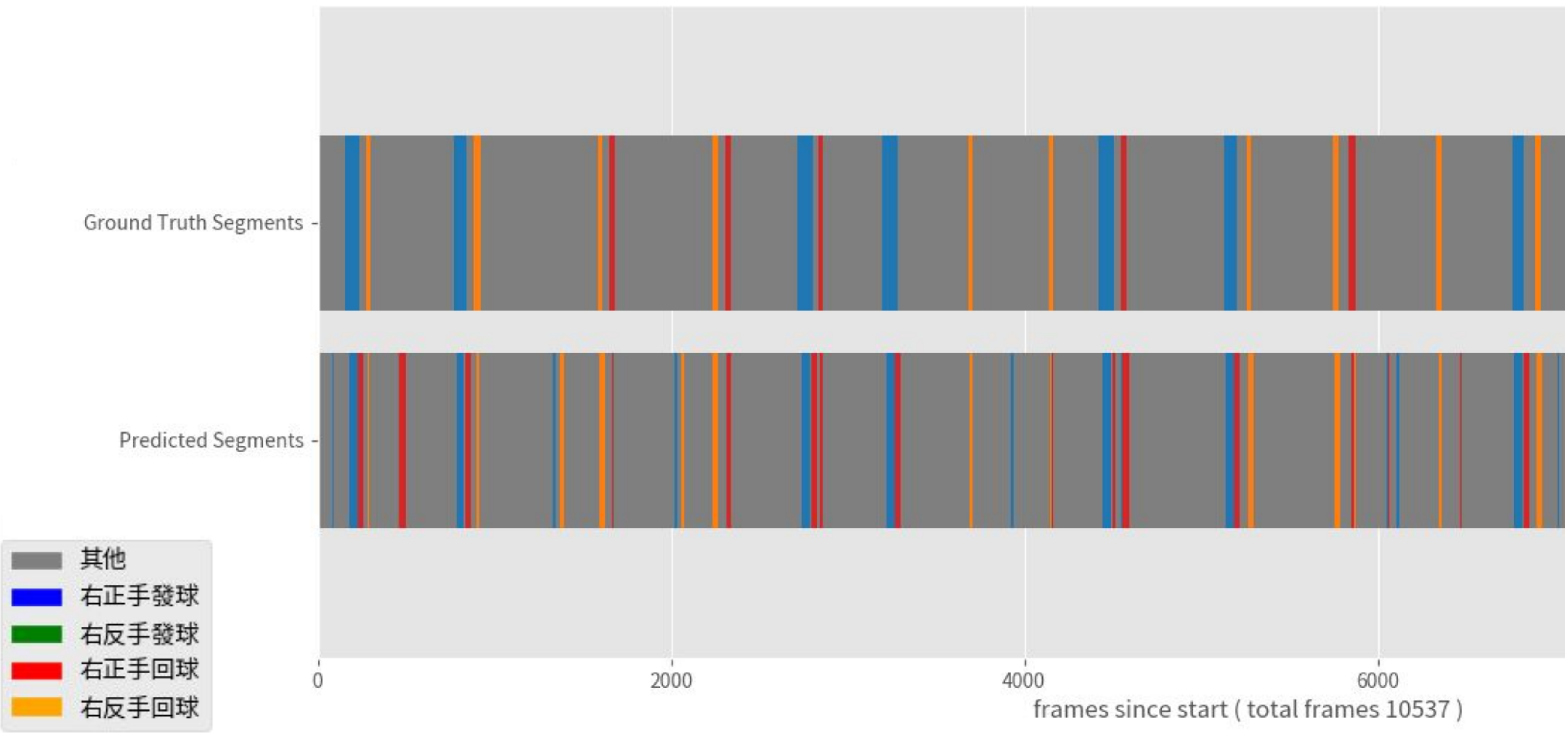
- Stroke-wise classification results of **subject F-2**.

Stroke Recognition								
video	**class**	**Ground Truth**	**TP**	**FN**	**FP**	**Precision**	**Recall**	**F1-Score**
F-2	**其他**	35	35	0	7	83.33%	100%	90.90%
	右正手發球	6	6	0	11	35.29%	100%	52.17%
	右反手發球	2	2	0	0	100%	100%	100%
	右正手回球	8	6	2	24	0.20%	75%	31.57%
	右反手回球	18	17	1	10	0.6296%	94.44%	75.55%

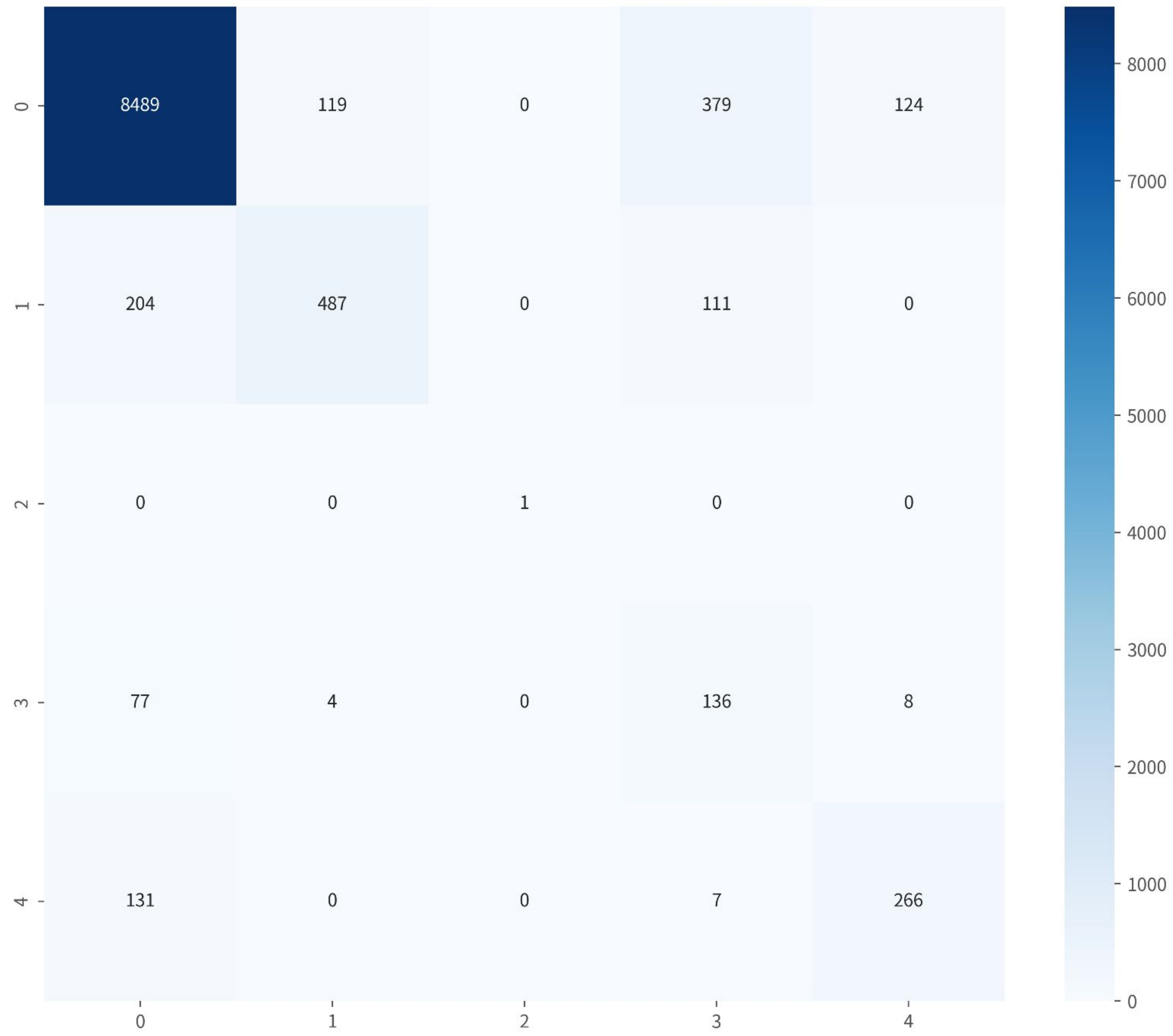
$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F - 1 = \frac{Precision \times Recall}{Precision + Recall}$$

Experiment Results (M-1)

- Frame-wise classification results of **subject M-1**.



Experiment Results (M-1)



class	IoUc
0 其他	89.14%
1 正手發球	52.64%
2 反手發球	-
3 正手回球	18.83%
4 反手回球	49.62%

$$\mathrm{IoU}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c},$$

Experiment Results (M-1)

- Stroke-wise classification results of **subject M-1**.

Stroke Recognition								
video	**class**	**Ground Truth**	**TP**	**FN**	**FP**	**Precision**	**Recall**	**F1-Score**
M-1	**其他**	33	33	0	10	76.74%	100%	86.84%
	右正手發球	10	10	0	13	43.47%	100%	60.60%
	右反手發球	0	0	0	0	-	-	-
	右正手回球	8	8	0	28	22.22%	100%	36.36%
	右反手回球	14	14	0	11	56%	100%	71.79%

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F - 1 = \frac{Precision \times Recall}{Precision + Recall}$$