Supplementary Materials

AutoMedTS: Automated Modeling of Physiological Time Series for Surgical Suturing Action Recognition

Baobing Zhang, Paul Sullivan, Benjie Tang, Ghulam Nabi, Mustafa Suphi Erden

THIS supplementary material provides additional experimental results, detailed parameter settings, extended ablation studies, and supporting analyses that were not included in the main paper due to space limitations. These contents are intended to further validate the effectiveness of the proposed method and offer more comprehensive insights. Readers are encouraged to review this supplementary file alongside the main manuscript.

To provide a clearer understanding of the time-series data processed by our model, we include a 3D visualization of surgical hand trajectories. The plots display the motion paths of both hands for experienced surgeons (Fig. 6) and novice surgeons (Fig. 7) during surgical tasks. These visualizations serve as supplementary materials to illustrate how surgical expertise affects motion patterns and to provide intuitive insights into our model inputs.

Figures 6 and 7 illustrate the three-dimensional trajectories of two instrumented markers (hands) for four experienced and four novice surgeons during simulated laparoscopic procedures. In Figure 6, the blue and red curves for experienced surgeons are smooth, stable, and spatially compact, reflecting precise, well-practiced motor control. By contrast, Figure 7 shows that novice surgeons exhibit greater variability, irregular patterns, and frequent corrective adjustments, indicative of exploratory, less confident behavior. This clear contrast in motion profiles not only highlights the impact of surgical expertise but also provides empirical support for leveraging trajectory-based time-series features in downstream machine learning and surgical action-recognition applications.

A. *Temperature-Scaled Softmax Resampling*

In *AutoMedTS*, we also tested on temperature-scaled Softmax resampling strategy. After sliding-window segmentation, we take the logarithm of the original window-level class proportions, divide by a temperature factor $T > 0$, and apply the Softmax function to obtain a "softened" target distribution. We then perform random under- and over-sampling according to these mapped proportions, effectively boosting minority classes while preserving the temporal trajectory. The sampling effect is shown in the figure 3.

1. **Sliding window segmentation**
   Slice the time series into overlapping windows of size w and step s, yielding $(X_{\text{win}}, y_{\text{win}})$ with total windows N.
2. **Compute original distribution**
   Count class occurrences $c_i = |\{y_{\text{win}} = i\}|$ and compute
   $$p_i = \frac{c_i}{\sum_j c_j}.$$
3. **Temperature-scaled Softmax softening**
   For a temperature $T > 0$, compute
   $$q_i = \exp\big(\ln(p_i)/T\big), \quad p'_i = \frac{q_i}{\sum_j q_j}.$$
   - $T > 1$: distribution flattens (small p_i boosted).
   - $T < 1$: distribution sharpens (large p_i accentuated).
4. **Determine target window counts**
   $$t_i = \lfloor p'_i \times N \rfloor.$$
   Compute remainder $\Delta = N - \sum_i t_i$ and add Δ to the class with the largest t_i.
5. **Window-level resampling**
   For each class i:
   - If $c_i \geq t_i$, randomly undersample to t_i.
   - Else, keep all c_i windows and oversample via random oversampling strategy.

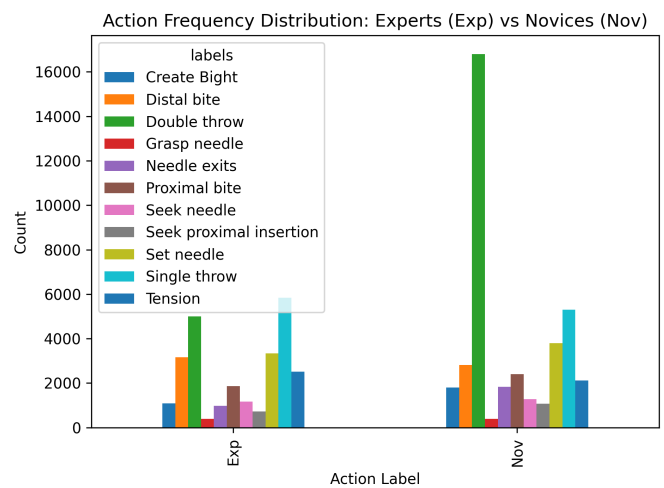

Fig. 1. Per-class window count comparison between experts (Exp) and novices (Nov) across different suturing actions.

Figure 1 compares the per-class window counts between experts (Exp) and novices (Nov), highlighting that common actions such as "Double throw" and "Single throw" occur more frequently in the Nov group. Figure 2 presents a sample time series for the "Distal bite" action by an expert, plotting the YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the sharp peaks correspond to needle insertion and exit events. Time-series visualizations for the remaining actions are provided in the subsequent sections of the Supplementary Materials. The complete comparison of different sample balance strategies on multiple indicators is shown in table I.

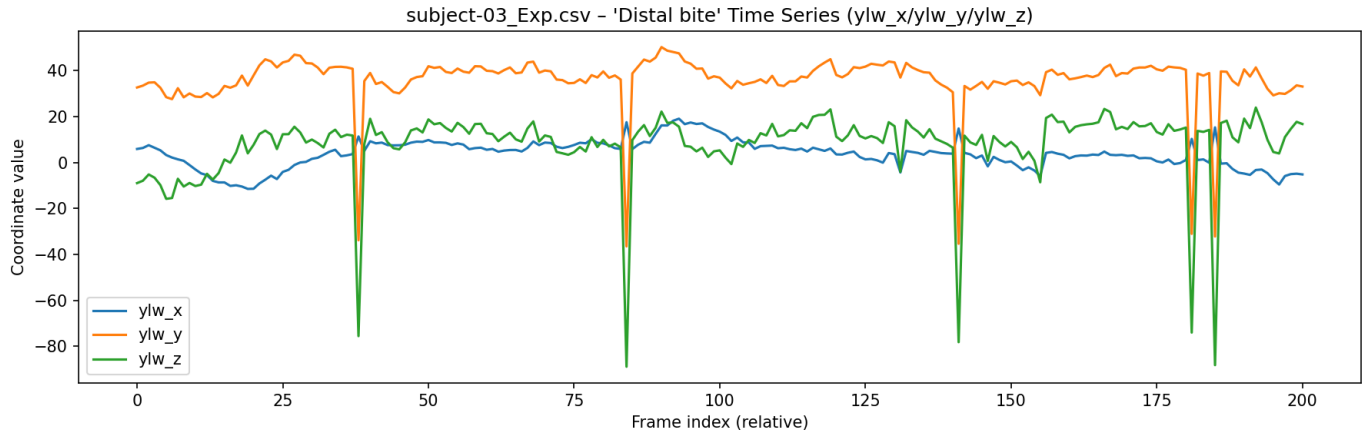

Fig. 2. Sample time series of the "Distal bite" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the sharp peaks indicate needle insertion and exit events.

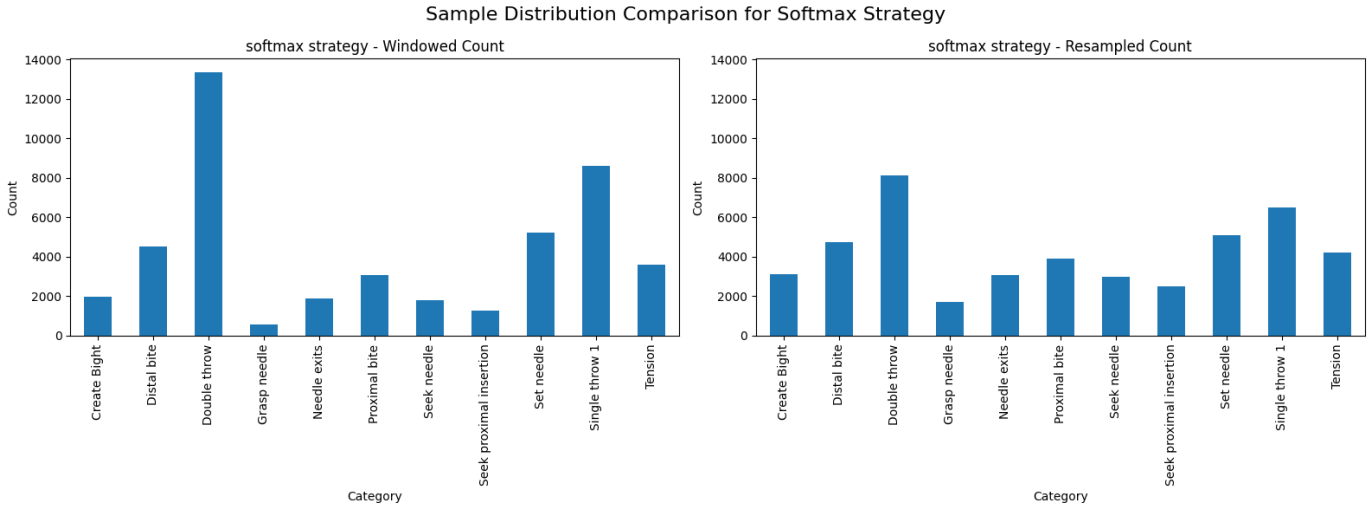

Fig. 3. Comparison of per-category counts before (windowed) and after (resampled) applying the temperature-scaled Softmax resampling strategy.

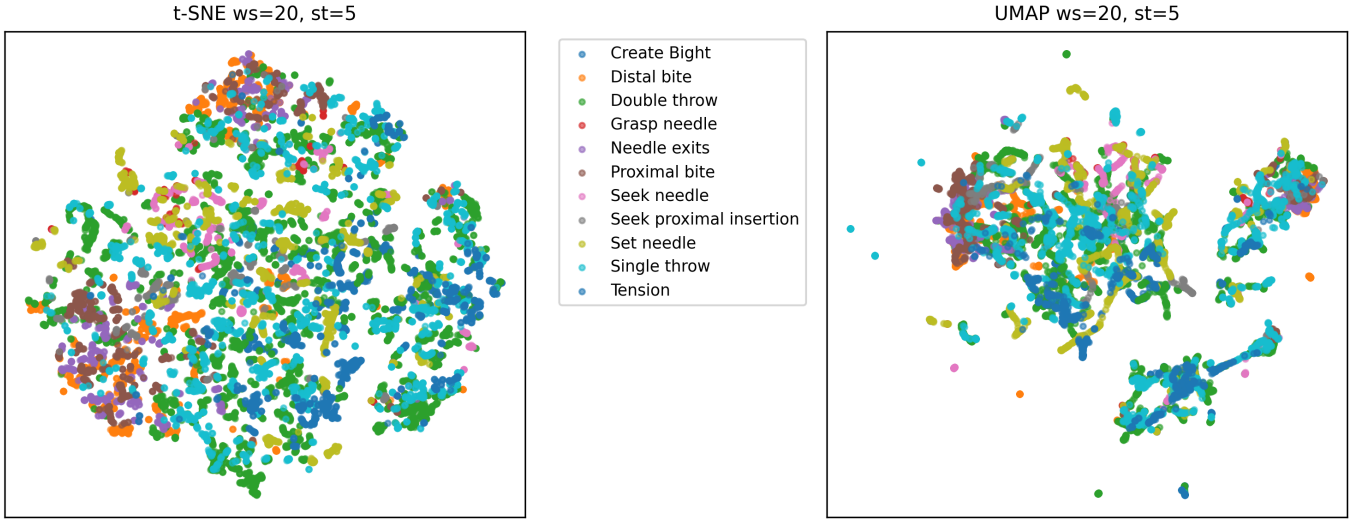

Fig. 4. Comparison of t-SNE and UMAP feature-space visualizations for window size = 20 and step = 5. Left: t-SNE projection; Right: UMAP projection.

TABLE I. Comparison of different resampling strategies

Method	Accuracy	Macro-F1	Macro-Recall	Balanced Acc	MCC
Baseline (No Resampling)	0.8921	0.8791	0.8534	0.8534	0.8596
Random Oversampling	**0.9136**	**0.9094**	**0.9030**	**0.9030**	**0.8887**
SMOTE Resampling	0.8508	0.8282	0.8033	0.8033	0.8053
SMOTE + Jitter Augmentation	0.8231	0.8001	0.8196	0.8196	0.7759

TABLE II. Summary of key metrics for penalty values from 0.05 to 0.95 (step 0.05), focusing on Accuracy, Macro-F1, Macro-Recall, Balanced Accuracy (Bal. Acc.), and Matthews Correlation Coefficient (MCC).

Penalty	Accuracy	Macro-F1	Macro-Recall	Bal. Acc.	MCC
0.05	0.8937	0.8877	0.8630	0.8630	0.8615
0.10	0.8678	0.8572	0.8080	0.8080	0.8275
0.15	0.9136	0.9107	0.9128	0.9128	0.8894
0.20	0.8863	0.8768	0.8489	0.8489	0.8517
0.25	0.8701	0.8595	0.8156	0.8156	0.8304
0.30	0.5218	0.4419	0.4132	0.4132	0.3594
0.35	0.8608	0.8437	0.7897	0.7897	0.8180
0.40	0.8643	0.8481	0.7882	0.7882	0.8230
0.45	0.3915	0.2886	0.2910	0.2910	0.2053
0.50	0.9124	0.9066	0.9111	0.9111	0.8881
0.55	0.8788	0.8389	0.7895	0.7895	0.8416
0.60	0.3564	0.2320	0.2366	0.2366	0.1601
0.65	0.4898	0.3543	0.3226	0.3226	0.2856
0.70	0.9067	0.9012	0.8938	0.8938	0.8792
0.75	0.3461	0.2208	0.2249	0.2249	0.1470
0.80	0.3451	0.2196	0.2236	0.2236	0.1453
0.85	0.4898	0.3543	0.3226	0.3226	0.2856
0.90	0.5502	0.5390	0.6028	0.6028	0.4479
0.95	0.8637	0.8172	0.8032	0.8032	0.8233

Plot 4 and plot 5 show the visualization analysis of the latent feature space at different synchronization scales. Although both t-SNE and UMAP project the high-dimensional sliding-window features into two dimensions, they exhibit markedly different clustering behaviors. In the t-SNE projection (left), points form a dense, disk-like cloud in which action labels are heavily intermixed, with only sporadic small clusters, indicating that t-SNE emphasizes global compression and struggles to separate classes. In contrast, UMAP (right) produces one or two relatively isolated clusters and a more stretched overall distribution that better preserves local neighborhood relationships, yet still shows considerable overlap. Therefore, neither method fully disentangles all surgical action categories, although UMAP demonstrates superior local clustering while t-SNE highlights the overall structure more cohesively.

The deep learning methods in Table II (FCN, CNN, TapNet, InceptionTime) in the main body of the paper perform sub-optimally on imbalanced time-series data: they tend to be dominated by majority-class samples, leading to low recall and F1-scores for minority actions. Moreover, their black-box nature lacks interpretability and stability guarantees, making them prone to performance variability under limited or noisy data. By contrast, the *AutoMedTS* framework combines traditional machine-learning algorithms with random oversampling, which not only enhances overall accuracy but also significantly improves minority-class recognition. Its use of tree-based models provides transparent decision paths and confidence measures, yielding more stable and reliable predictions under class imbalance.

As shown in Table II, we swept the penalty parameter from 0.05 to 0.95 in increments of 0.05 and recorded five core metrics: Accuracy, Macro-F1, Macro-Recall, Balanced Accuracy, and Matthews Correlation Coefficient (MCC). The results indicate that performance peaks at penalty values of 0.15 and 0.50, with 0.15 achieving the highest accuracy at 0.9136 and Macro-F1 at 0.9107, while 0.50 reaching 0.9111 for both Macro-Recall and Balanced Accuracy. Interestingly, model performance drops significantly at multiple points including 0.30, 0.45, 0.60, and 0.75, suggesting high sensitivity to the penalty parameter choice. A penalty of 0.70 also performs well with an MCC of 0.8792. Considering all metrics, we select (penalty=0.15) as the optimal setting for our final model as it maximizes accuracy and F1 score while maintaining high values across other metrics.

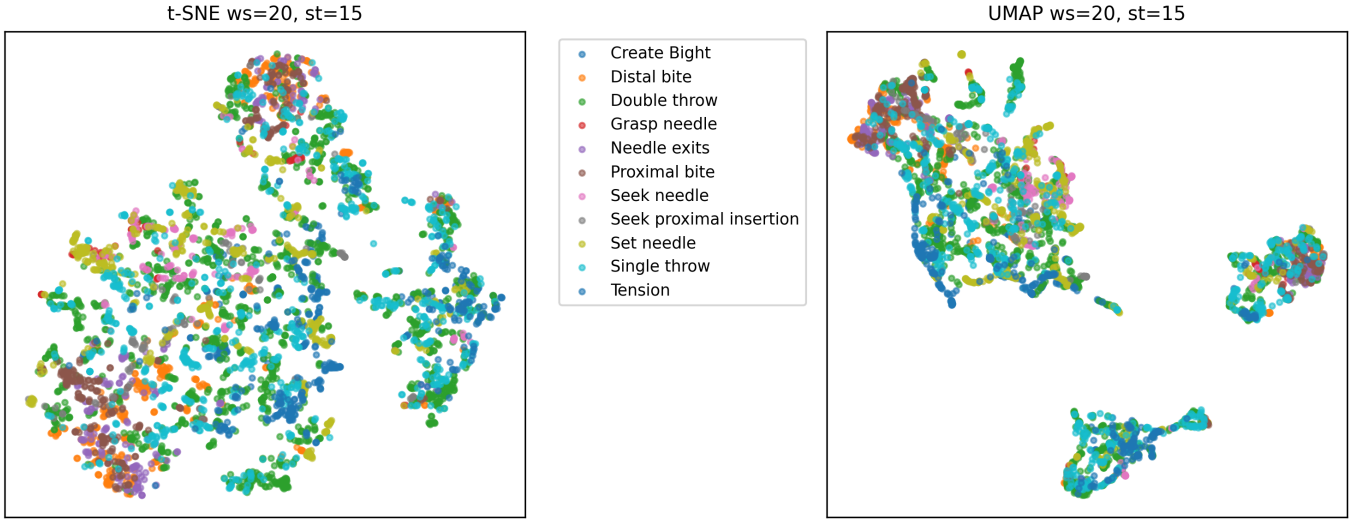

Fig. 5. Comparison of t-SNE and UMAP feature-space visualizations for window size = 20 and step = 15. Left: t-SNE projection; Right: UMAP projection.

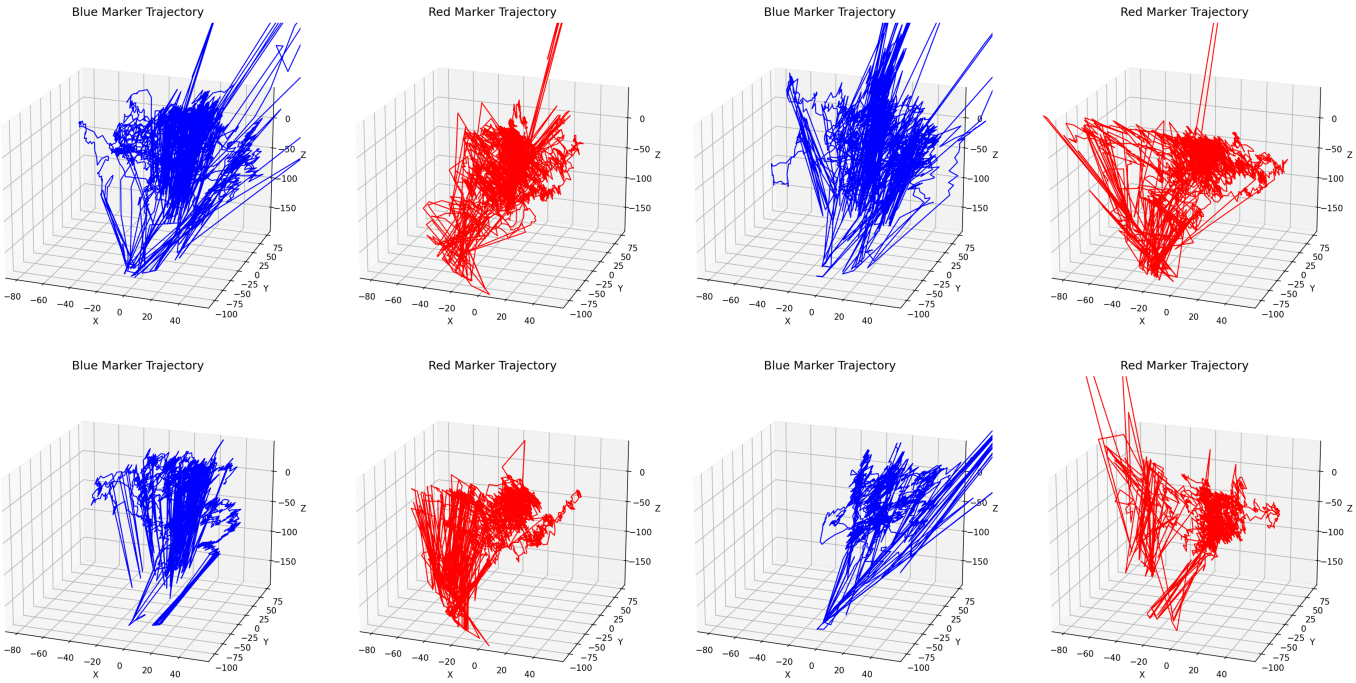

Fig. 6. This figure shows the three-dimensional motion trajectories of two instrumented markers (hands) for four experienced surgeons (subjects 03, 04, 05, 07) during laparoscopic procedures. Blue lines trace the path of one hand, while red lines trace the path of the other.

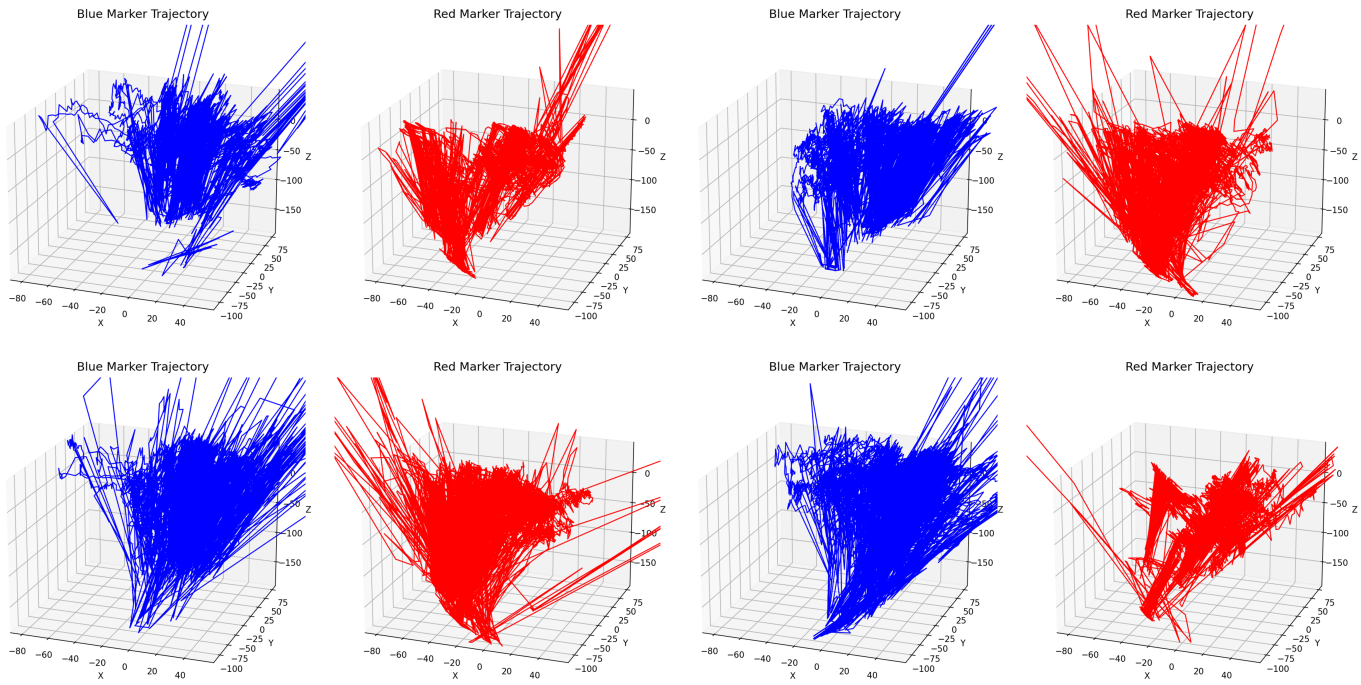

Fig. 7. This figure shows the three-dimensional motion trajectories of two instrumented markers (hands) for four novice surgeons (subjects 06, 08, 09, 10) during laparoscopic procedures. Blue lines trace the path of one hand, while red lines trace the path of the other.

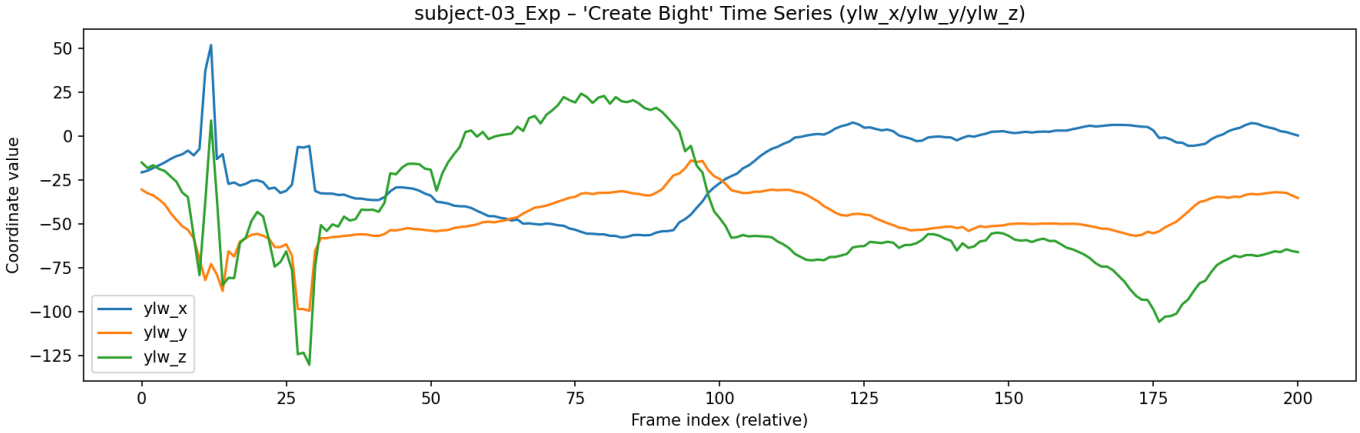

Fig. 8. Sample time series of the "Create Bight" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the pronounced spike in Z-axis at frame ∼15 reflects the initial loop formation, followed by gradual settling as the bight is secured.

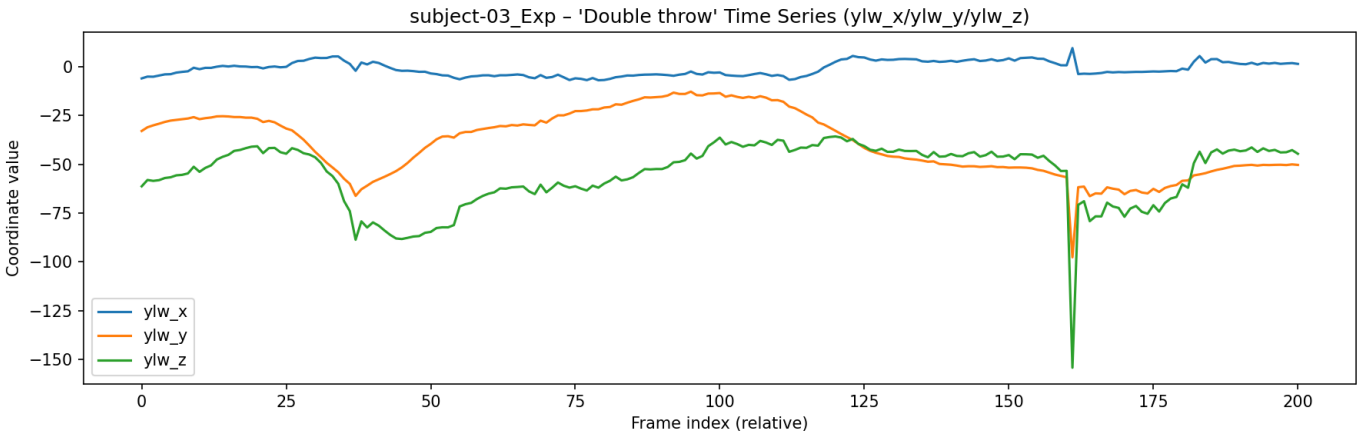

Fig. 9. Sample time series of the "Double throw" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the two sequential deep troughs in Z-axis around frames ∼35 and ∼160 correspond to the first and second throws, with an intermediate plateau indicating repositioning.

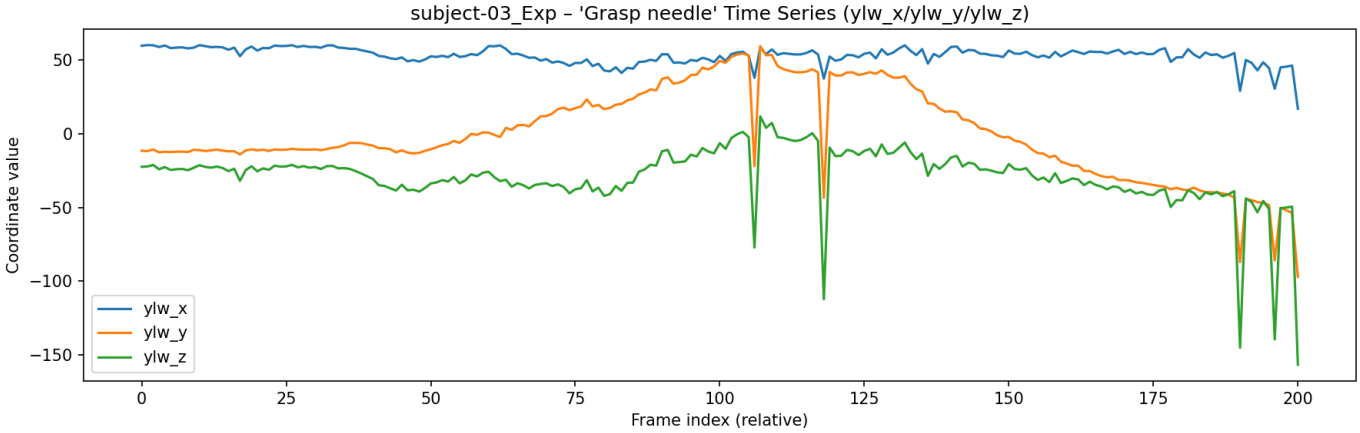

Fig. 10. Sample time series of the "Grasp needle" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the sharp downward peaks in Z-axis around frames ∼110–120 represent needle clamping, followed by elevated X- and Y-axis trajectories as the instrument lifts the needle.

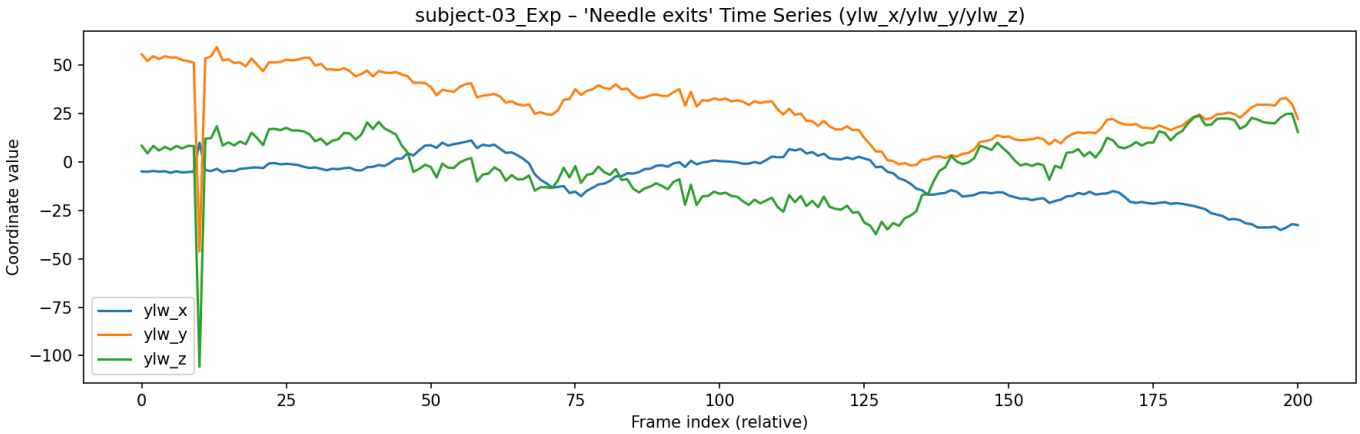

Fig. 11. Sample time series of the "Needle exits" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the abrupt peaks in Y- and Z-axes around frame ∼10 indicate the rapid exit motion of the needle tip, followed by a return to baseline.

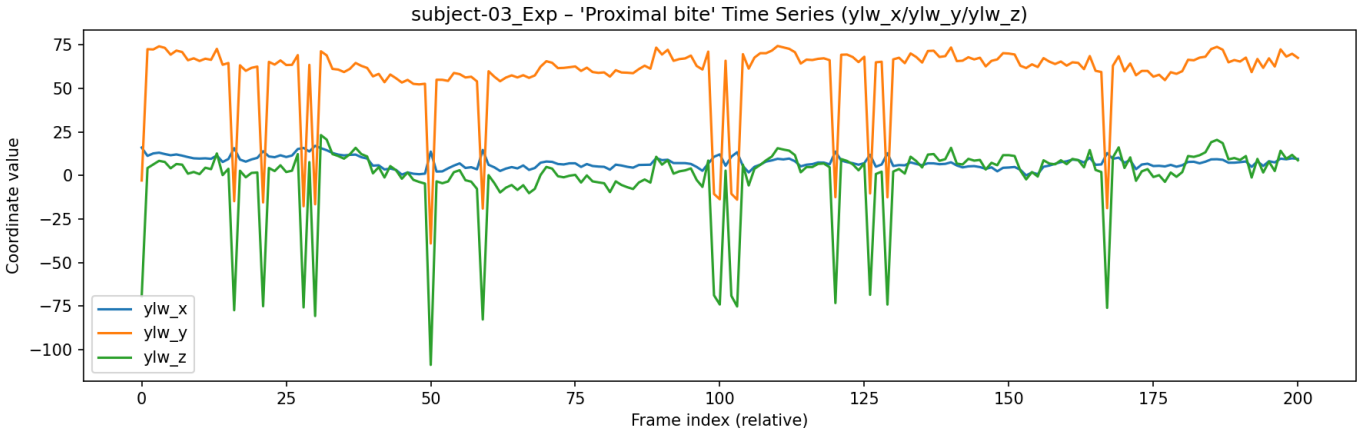

Fig. 12. Sample time series of the "Proximal bite" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; multiple sharp downward spikes in Y- and Z-axes correspond to repeated proximal insertion and withdrawal motions, with intervening intervals of relative stability.

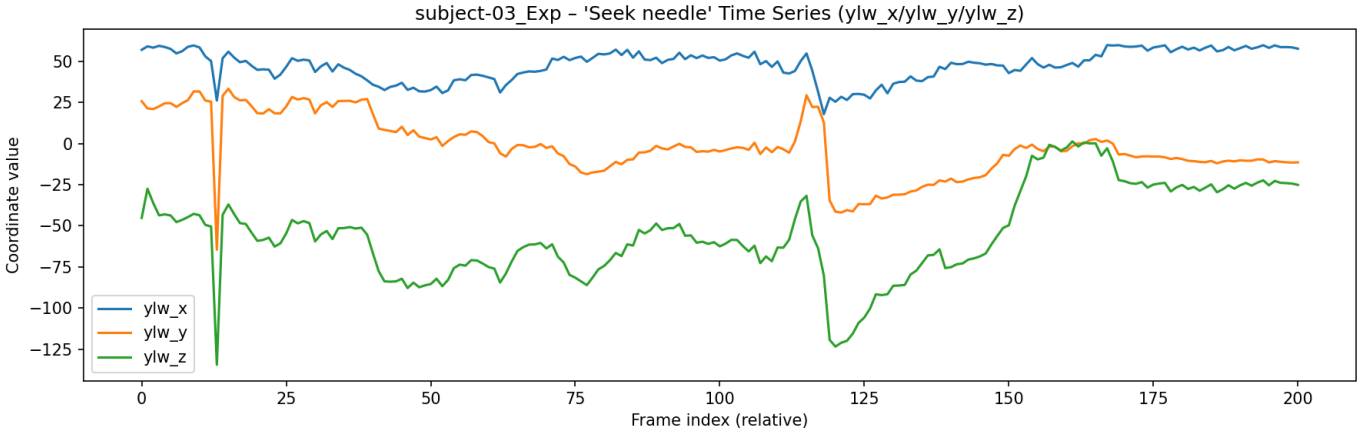

Fig. 13. Sample time series of the "Seek needle" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the pronounced trough in Z-axis around frame ∼110 reflects the initial needle localization, followed by gradual oscillations in all three axes as fine positional adjustments are made.

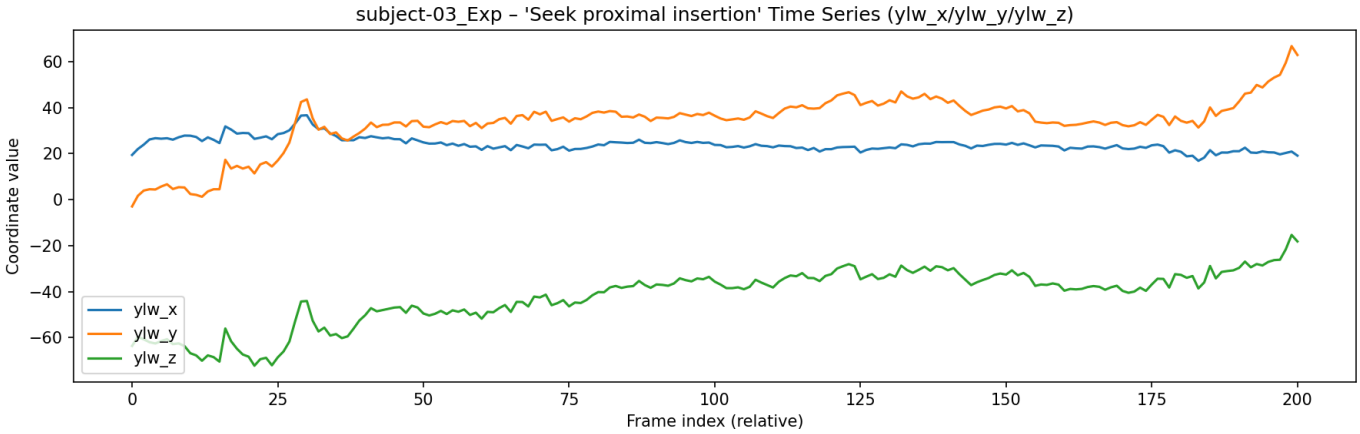

Fig. 14. Sample time series of the "Seek proximal insertion" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the transient peak in Y-axis and trough in Z-axis around frame ∼30 reflect initial proximal tissue probing, followed by gradual and steady adjustments.

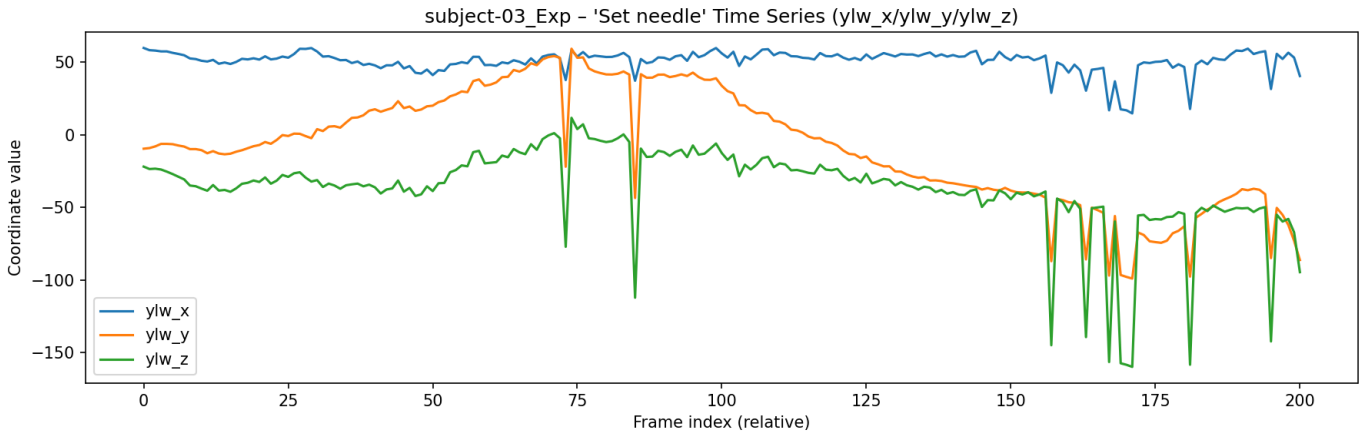

Fig. 15. Sample time series of the "Set needle" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the two sharp troughs in Z-axis around frames ∼70 and ∼80 correspond to needle repositioning and fine adjustment, followed by gradual recovery in Y- and X-axes as the tip achieves alignment.

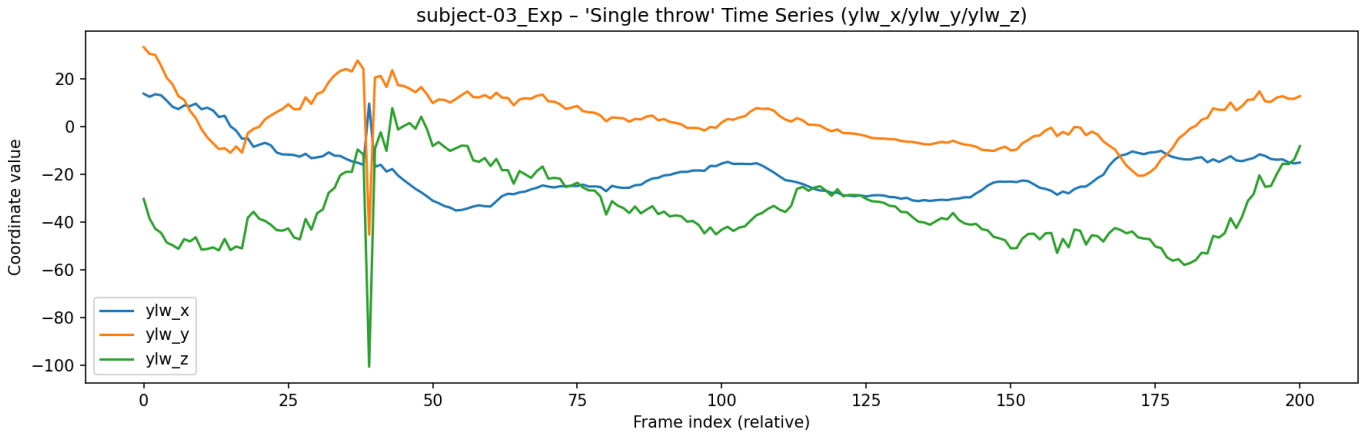

Fig. 16. Sample time series of the "Single throw" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the pronounced drop in Z-axis around frame ∼40 indicates the primary throwing motion, followed by gentle oscillations in X- and Y-axes representing suture loop tensioning.

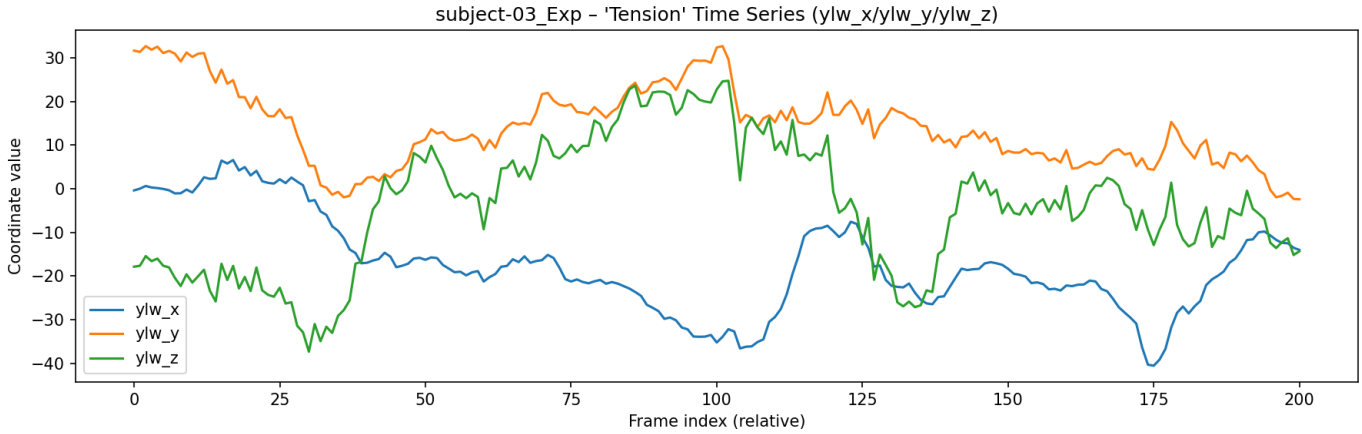

Fig. 17. Sample time series of the "Tension" action showing YLW-X, YLW-Y, and YLW-Z coordinates over frame index; the sudden Z-axis trough around frame ∼100 marks the moment of suture tightening, followed by a gradual increase in X- and Y-axes as constant tensile force is applied.