Supplementary of "Temporal Micro-Action Localization with Skeleton-Guided Mamba for Videofluoroscopic Swallowing Study"

Yirui Li, Kai Zhou, Meng Dai, Haiyu Zhou, Jinwu Hu, Yifan Yang, Jian Chen, Fei Liu, Hongmin Cai, *Senior Member, IEEE*, Mingkui Tan, *Senior Member, IEEE*

In this supplementary, we organize as follows:

- In Section I, we provide more details about dataset statistics, skeleton keypoint annotation, and data preprocessing.
- In Section II, we present more implementation and training details of our proposed SG-Mamba.
- In Section III, we show more experimental results for each micro-action and error analysis.

I. Additional Dataset Details

Dataset Statistic. We use the videofluoroscopic dataset from [1], consisting of videos lasting 508 minutes and 21 seconds, constructed from 71 subjects (25 male, 46 female). The dataset is split into training, validation, and test sets in a 4:1:1 ratio per subject. Each video contains at least one complete swallow event. One swallow process includes seven micro-actions: Oral Transit, Soft Palate Elevation, Hyoid Motion, UES Opening, Swallow Initiation, Pharyngeal Transit, and Laryngeal Vestibule Closure. Their average duration is less than one second. The micro-action distribution of the dataset is shown in Fig. 1. All swallowing micro-actions occur in various positions in the videos, and most of their durations range from 0.5 to 1.5 seconds. The max duration is smaller than 4 seconds.

External dataset statistic. To better validate the robustness and reliability of our methods, we test our methods on data from the Mingxin Rehabilitation Center. The data were constructed from 11 subjects. The data were digitally recorded as videos by using the VFSS Acquisition and Analysis system (Longest Inc., Guangzhou, China) at 30 frames/sec. Two experienced clinicians independently assessed all videos. The total duration of the videos is 95 minutes. Following the preprocess procedure in [1], a total of 48 clips were obtained. We annotated the datasets with the same standard as the training dataset. The average duration of micro-action is 1.1 seconds.

Keypoint annotation. We extract 13,895 frames from 210 trimmed swallowing videos, focusing on active swallowing segments. These frames are divided into three subsets for keypoint localization: 8,302 images (from 131 videos) for training, 1,518 images (from 16 videos) for validation, and 4,075 images (from 63 videos) for testing. We annotate eight key anatomical landmarks essential for swallowing analysis:

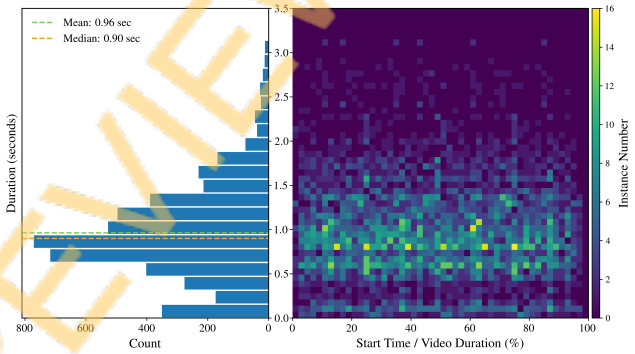

Fig. 1. Micro-action distribution in the VFSS dataset.

the superior, anterior, and inferior points of the anterior hyoid bone; the anterior-inferior corners of the C2 and C4 vertebrae; and three points on the soft palate (anterior, midpoint, and inferior). The annotations are performed using the open-source *CVAT* tool[1] by clinical experts. To ensure reliability, we implement a multi-step verification process, including cross-validation among annotators and consistency checks.

Data Preprocess. Following [1], we use sliding windows to intercept video clips from the original video file. For the coarse localization stage, the sliding windows are set to 32 seconds and 64 seconds with one-fourth of the window length as sliding steps. For the fine localization stage, the sliding window is set to 4 seconds and the stride is set to 3 frames. All the video clips are processed using a sliding window of 8 frames with a stride of 3 frames for feature extraction. The sizes of input are 224 for the appearance branch and 56 for the skeleton branch. We employ a Kalman filter [2] to reduce noise and smooth trajectories, enhancing anatomical tracking accuracy during swallowing.

II. Additional Implementation Details

More implementation details. We employ a coarse-to-fine localization mechanism for temporal micro-action localization. Here, we provide additional details about the coarse localization stage. Following [1], we resize the RGB and optical flow data to a resolution of 128 and sample the data with 30

[1]https://github.com/cvat-ai/cvat

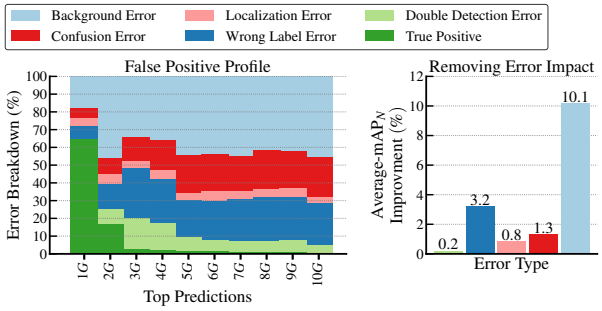

Fig. 2. False positive profile of our framework using [8]. Left: FP error breakdown considering top-10G predictions where G is the number of ground truths. Right: The impact of error types. Background error is the top error type.

FPS for fair comparison. Unlike the previous approach, we do not interpolate the data; instead, we pad the sequence to a length of 864. For benchmark testing, we utilize the same pre-trained I3D [3] from [1] to extract appearance features for feature-based methods and a pre-trained VideoMAE [4] for the end-to-end method (AdaTAD). The coarse localization stage model is trained for 45 epochs with a batch size of 2. For optimization, we adopt the same settings as ActionMamba [5] on the THUMOS14 dataset [6], which includes AdamW optimization with base learning rate 1e-4, weight decay 0.05, and linear warm-up during the first 5 epochs. Model EMA [7] and gradient clipping were also implemented to further stabilize the training.

Reproducibility of Results. All results presented in this paper were obtained using PyTorch 2.1.2 and CUDA 11.8, with deterministic GPU computing routines, and were trained on 7 NVIDIA 3090 GPUs under the same random seed. Our implementation ensures reproducibility, consistently yielding identical results when the same random seed and hardware configuration are used. For full reproducibility, Alg. 1 outlines our complete training pipeline and Alg. 2 provides PyTorch-style pseudocode implementation with annotations for the Channel-enhanced Cross-Mamba (CCM) module.

III. More Results and Visualizations

More Detailed Results We provide more detailed results of all micro-actions for the compared methods. Tab. I shows a detailed evaluation result comparison on temporal action localization methods. Our method greatly improves the performance of difficult micro-actions like "Oral Transit" and "Soft Palate Elevation" and reaches much better performance at higher tIoU thresholds. Tab. II provides evaluation results on methods for VFSS-specific and skeleton-guided methods.

Error Analysis To analyze the limitations of our method, we present the false positive error chart using tIoU thresholds ranging from 0.1 to 0.7 in Figure 2. Similar to observations in BasicTAD [9] and ViT-TAD [10], anchor-free methods are prone to "Background error", primarily due to the limited number of anchors available for distinguishing between true actions and background noise. This limitation highlights a key challenge in achieving precise temporal localization, particularly in scenarios with subtle or short-duration micro-actions.

References

[1] X. Ruan, M. Dai, Z. Chen, Z. You, Y. Zhang, Y. Li, Z. Dou, and M. Tan, "Temporal micro-action localization for videofluoroscopic swallowing study," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[3] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.

[4] Z. Tong, Y. Song, J. Wang, and L. Wang, "Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training," *Advances in neural information processing systems*, vol. 35, pp. 10 078–10 093, 2022.

[5] G. Chen, Y. Huang, J. Xu, B. Pei, Z. Chen, Z. Li, J. Wang, K. Li, T. Lu, and L. Wang, "Video mamba suite: State space model as a versatile alternative for video understanding," *arXiv preprint arXiv:2403.09626*, 2024.

[6] H. Idrees, A. R. Zamir, Y.-G. Jiang, A. Gorban, I. Laptev, R. Sukthankar, and M. Shah, "The thumos challenge on action recognition for videos "in the wild"," *Computer Vision and Image Understanding*, vol. 155, pp. 1–23, 2017.

[7] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get m for free," *arXiv preprint arXiv:1704.00109*, 2017.

[8] H. Alwassel, F. C. Heilbron, V. Escorcia, and B. Ghanem, "Diagnosing error in temporal action detectors," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 256–272.

[9] M. Yang, G. Chen, Y.-D. Zheng, T. Lu, and L. Wang, "Basictad: an astounding rgb-only baseline for temporal action detection," *Computer Vision and Image Understanding*, vol. 232, p. 103692, 2023.

[10] M. Yang, H. Gao, P. Guo, and L. Wang, "Adapting short-term transformers for action detection in untrimmed videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 18 570–18 579.

[11] L. Yang, H. Peng, D. Zhang, J. Fu, and J. Han, "Revisiting anchor mechanisms for temporal action localization," *IEEE Transactions on Image Processing*, vol. 29, pp. 8535–8548, 2020.

[12] C.-L. Zhang, J. Wu, and Y. Li, "Actionformer: Localizing moments of actions with transformers," in *European Conference on Computer Vision*. Springer, 2022, pp. 492–510.

[13] D. Shi, Y. Zhong, Q. Cao, L. Ma, J. Li, and D. Tao, "Tridet: Temporal action detection with relative boundary modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 857–18 866.

[14] S. Liu, C.-L. Zhang, C. Zhao, and B. Ghanem, "End-to-end temporal action detection with 1b parameters across 1000 frames," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 18 591–18 601.

[15] S. W. Hyder, M. Usama, A. Zafar, M. Naufil, F. J. Fateh, A. Konin, M. Z. Zia, and Q.-H. Tran, "Action segmentation using 2d skeleton heatmaps and multi-modality fusion," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 1048–1055.

Algorithm 1 Complete Training Pipeline of SG-Mamba

Input: Training dataset $\mathcal{D} = \{(V_i, Z_i)\}_{i=1}^{|\mathcal{D}|}$, where V_i is a video and Z_i is its corresponding micro-action labels.
Pre-trained models: $\mathcal{F}_{\text{rgb}}$ (RGB feature extractor), $\mathcal{F}_{\text{flow}}$ (optical flow feature extractor), and $\mathcal{F}_{\text{BM}}$ (skeleton encoder).
Batch size B_1, B_2, training epochs E_1, E_2, and number of micro-action categories C.
Output: Trained SG-Mamba micro-action encoders $\{\mathcal{F}_{\text{BM}}^c\}_{c=1}^{C}$, CCM modules and detectors $\mathcal{G}^{coarse}, \{\mathcal{G}_c^{app}, \mathcal{G}_c^{ske}\}_{c=1}^{C}$.

Coarse Localization Stage:
for $j = 1$ to E_1 **do**
  Initialize swallowing event detector $\mathcal{G}^{coarse}$.
  for each batch $\{(V_i, Z_i)\}_{i=1}^{B_1} \subset \mathcal{D}$ **do**
    Extract appearance features f_i^{app} using $\mathcal{F}_{\text{rgb}}$ and $\mathcal{F}_{\text{flow}}$.
    Predict swallowing event proposals using ActionMamba:
    $\{(\hat{s}_i^{coarse}, \hat{e}_i^{coarse}, \theta_i^{coarse})\}_{i=1}^{T} = \mathcal{G}^{coarse}(F^{app})$
    Compute loss $\mathcal{L}_{coarse} = \mathcal{L}_{cls}^{coarse} + \mathcal{L}_{reg}^{coarse}$, where:
    $\mathcal{L}_{cls}^{coarse}$ is the classification loss for swallowing detection.
    $\mathcal{L}_{reg}^{coarse}$ is the regression loss for proposal boundaries.
    Update model parameters by minimizing $\mathcal{L}_{coarse}$.
  end for
end for
Fine Localization Stage:
Compute optical flow v_i^{flow} from RGB frames v_i for each video V_i.
Generate 3D skeleton heatmaps h_i^{3D} .
for $c = 1$ to C **do**
  Initialize skeleton encoder $\mathcal{F}_{\text{BM}}^c$, fusion module CCM, detectors $\mathcal{G}_c^{app}$ and $\mathcal{G}_c^{ske}$ for category c.
  for $j = 1$ to E_2 **do**
    for each batch $\{(V_i, Z_i)\}_{i=1}^{B_2} \subset \mathcal{D}$ **do**
      Extract skeleton features f_i^{ske} using $\mathcal{F}_{\text{BM}}^c$.
      Enhance features using CCM:
      $E^{app} = \text{CCM}(F^{app}, F^{ske})$ *// Enhance appearance features with skeleton features*
      $E^{ske} = \text{CCM}(F^{ske}, F^{app})$ *// Enhance skeleton features with appearance features*
      Predict micro-actions using ActionMamba detectors:
      $\{(\hat{s}_i^{app}, \hat{e}_i^{app}, \theta_i^{app})\}_{i=1}^{T} = \mathcal{G}_c^{app}(E^{app})$, $\{(\hat{s}_i^{ske}, \hat{e}_i^{ske}, \theta_i^{ske})\}_{i=1}^{T} = \mathcal{G}_c^{ske}(E^{ske})$
      Compute loss $\mathcal{L}_{fine} = \mathcal{L}_{app} + \mathcal{L}_{ske}$, where:
      $\mathcal{L}_{app} = \mathcal{L}_{cls}^{app} + \mathcal{L}_{reg}^{app}$, $\mathcal{L}_{ske} = \mathcal{L}_{cls}^{ske} + \mathcal{L}_{reg}^{ske}$ *// Appearance branch & Skeleton branch loss*
      Update model parameters by minimizing $\mathcal{L}_{fine}$.
    end for
  end for
end for

Algorithm 2 PyTorch style pseudo-code for a single block of Channel-enhanced Cross-Mamba (CCM)

```
def forward(self, feat, com_feat):
    """
    Inputs:
        feat:      (B, L, D)      - Input features
        com_feat:  (B, L, D')     - Complementary features
    Output:
        enhanced_feat: (B, L, D) - Enhanced features
    """
    # Step 1: Project input features and complementary features
    x = self.in_proj(feat)           # (B, L, D) -> (B, 2D, L)
    z = self.v_in_proj(com_feat)     # (B, L, D') -> (B, 2D, L)

    # Step 2: Split features into forward and backward chunks
    x_f, x_b = torch.chunk(x, 2, dim=1)  # (B, 2D, L) -> [(B, D, L), (B, D, L)]
    z_f, z_b = torch.chunk(z, 2, dim=1)  # (B, 2D, L) -> [(B, D, L), (B, D, L)]
    x_b = x_b.flip([-1])  # Reverse the backward chunk
    z_b = z_b.flip([-1])  # Reverse the backward chunk

    # Step 3: Concatenate forward and backward chunks
    x = torch.cat([x_f, x_b], dim=0)  # (2B, D, L)
    z = torch.cat([z_f, z_b], dim=0)  # (2B, D, L)
    xz = torch.cat([x, z], dim=1)     # (2B, 2D, L)

    # Step 4: Apply convolution, silu and bidirectional state-space model (SSM)
    out = self.conv_and_ssm(xz)  # (2B, 2D, L) -> (2B, 2D, L)

    # Step 5: Recombine forward and backward outputs
    out = out.chunk(2)  # Split into forward and backward outputs
    out = torch.cat([out[0], out[1].flip([-1])], dim=1)  # (B, 2D, L)

    # Step 6: Apply Channel-wise Enhancement (CE)
    comm = F.silu(self.conv(out))    # (B, 2D, L) -> (B, 1, L)
    out = out + self.gamma * (out - comm)

    # Step 7: Project output back to the original dimension and apply residual connection
    enhanced_feat = self.out_proj(out) + feat  # (B, 2D, L) -> (B, L, D)

    return enhanced_feat
```

TABLE I
DETAIL EVALUATION FOR SEVEN MICRO-ACTIONS ON EXISTING TEMPORAL ACTION LOCALIZATION METHODS.

Method	Micro-action	AP@tIoU Thresholds							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	Avg.
A2Net [11]	Oral Transit	42.3	41.9	34.1	26.4	13.1	7.0	2.3	23.9
	Soft Palate Elevation	72.1	66.8	63.0	48.8	30.0	9.6	3.2	41.9
	Hyoid Motion	60.0	57.5	53.3	39.6	25.6	13.4	4.5	36.3
	UES Opening	68.2	38.8	26.0	13.7	4.0	0.9	0.7	17.9
	Swallow Initiation	12.0	9.0	3.9	0.9	0.3	0.1	0.1	3.8
	Pharyngeal Transit	76.8	76.8	75.4	66.8	48.4	24.6	7.8	53.8
	Laryngeal Vestibule Closure	72.1	69.8	63.9	55.8	35.5	18.6	6.2	46.0
	Average	53.7	51.5	45.7	36.0	22.4	10.6	3.5	31.9
ActionFormer [12]	Oral Transit	48.2	47.4	40.9	25.0	14.5	11.2	7.6	27.8
	Soft Palate Elevation	65.4	64.6	64.6	57.5	49.4	36.9	22.4	51.6
	Hyoid Motion	73.9	51.9	35.2	17.9	9.4	3.0	1.3	27.5
	UES Opening	68.2	68.1	68.1	67.0	61.0	52.6	34.7	60.0
	Swallow Initiation	35.7	26.5	21.1	11.6	6.7	3.6	2.2	15.4
	Pharyngeal Transit	70.2	70.2	69.9	54.7	34.4	23.5	15.6	48.4
	Laryngeal Vestibule Closure	83.4	83.2	82.9	81.7	71.8	57.4	39.1	71.4
	Average	76.8	74.4	69.7	59.3	48.8	38.5	24.6	56.0
TriDet [13]	Oral Transit	45.8	43.0	38.5	22.5	15.8	10.9	6.1	26.1
	Soft Palate Elevation	70.9	70.9	64.4	60.3	54.1	41.7	25.2	55.4
	Hyoid Motion	73.8	52.3	41.3	25.2	17.1	6.7	2.7	31.3
	UES Opening	68.8	68.7	68.4	66.1	62.4	51.3	38.1	60.5
	Swallow Initiation	39.0	29.3	20.9	13.8	8.2	4.3	2.1	16.8
	Pharyngeal Transit	71.6	71.6	67.9	52.7	37.9	27.6	16.9	49.5
	Laryngeal Vestibule Closure	84.6	84.4	83.9	83.4	69.2	58.0	36.9	71.5
	Average	79.6	76.6	72.4	63.7	53.1	40.9	26.2	58.9
AdaTAD [14]	Oral Transit	73.4	69.9	46.3	36.1	24.9	11.2	4.6	38.0
	Soft Palate Elevation	68.4	68.4	68.3	66.2	58.8	52.4	21.5	57.7
	Hyoid Motion	81.4	62.8	35.8	22.2	17.0	9.5	5.7	33.5
	UES Opening	80.6	80.2	80.2	80.2	79.2	69.9	45.4	73.7
	Swallow Initiation	29.8	20.7	12.3	9.3	6.6	3.8	1.6	12.0
	Pharyngeal Transit	65.0	65.0	64.8	57.3	48.7	29.9	16.4	49.6
	Laryngeal Vestibule Closure	87.3	87.3	87.3	81.5	76.0	68.3	49.2	76.7
	Average	80.9	77.4	70.0	62.3	54.4	42.1	24.8	58.9
ActionMamba [5]	Oral Transit	43.4	41.4	36.5	31.2	18.7	11.6	6.6	27.1
	Soft Palate Elevation	68.7	68.7	65.6	59.8	50.6	32.8	17.9	52.0
	Hyoid Motion	73.1	50.2	37.1	26.4	15.5	8.2	3.3	30.5
	UES Opening	69.0	68.8	68.8	68.3	63.8	45.1	27.4	58.8
	Swallow Initiation	46.5	36.8	26.4	17.6	10.1	6.2	2.2	20.8
	Pharyngeal Transit	77.6	77.6	69.8	48.3	29.0	20.6	10.6	47.6
	Laryngeal Vestibule Closure	82.9	82.0	82.0	82.0	72.6	57.1	33.6	70.3
	Average	82.0	79.3	74.3	67.4	53.1	37.1	19.9	59.0
SG-Mamba (Ours)	Oral Transit	73.6	69.7	62.7	51.1	39.5	21.5	12.2	47.2
	Soft Palate Elevation	86.4	84.7	84.6	82.9	74.6	67.8	46.2	75.3
	Hyoid Motion	88.0	87.5	82.6	67.1	37.0	23.4	6.6	56.0
	UES Opening	85.9	85.9	85.5	85.2	85.0	83.6	77.8	84.1
	Swallow Initiation	73.0	47.9	34.0	18.0	10.9	5.1	2.2	27.3
	Pharyngeal Transit	90.7	89.8	89.3	89.3	86.5	75.2	54.6	82.2
	Laryngeal Vestibule Closure	84.2	84.0	83.4	82.1	79.3	73.5	60.9	78.2
	Average	83.1	78.5	74.6	67.9	59.0	50.0	37.2	64.3

TABLE II

DETAIL EVALUATION FOR SEVEN MICRO-ACTIONS ON MORE METHODS. "SKE." REPRESENTS THE SKELETON INPUT MODALITY.

Method	Detector	Ske.	Micro-action	AP@tIoU Thresholds							
				0.1	0.2	0.3	0.4	0.5	0.6	0.7	Avg.
Ruan *et al.* [1]	A2Net		Oral Transit	75.1	72.2	59.8	38.7	23.0	11.4	7.1	41.0
			Soft Palate Elevation	71.7	71.7	71.7	71.2	66.3	42.9	10.1	57.9
			Hyoid Motion	75.4	75.4	71.8	50.9	27.9	7.8	1.1	44.3
			UES Opening	71.7	71.7	71.7	71.7	71.2	57.5	42.1	65.4
			Swallow Initiation	58.7	37.7	19.3	8.8	3.8	1.5	8.0	19.7
			Pharyngeal Transit	71.9	71.9	71.9	71.9	65.0	43.7	19.8	59.4
			Laryngeal Vestibule Closure	71.7	71.7	71.7	71.7	65.6	56.6	29.4	62.6
			Average	70.9	67.5	62.5	55.0	46.1	31.6	15.8	49.9
Ruan *et al.* [1]	ActionMamba		Oral Transit	70.9	68.8	57.4	42.4	37.7	27.6	14.2	45.6
			Soft Palate Elevation	80.8	80.7	78.4	78.1	64.9	42.7	20.4	63.7
			Hyoid Motion	74.4	73.8	69.6	56.6	34.7	21.3	7.4	48.2
			UES Opening	80.5	80.5	79.7	79.7	79.1	76.1	63.8	77.1
			Swallow Initiation	71.4	55.4	37.3	17.6	12.9	6.7	2.2	29.1
			Pharyngeal Transit	84.2	83.7	82.2	80.4	76.8	71.6	41.7	74.4
			Laryngeal Vestibule Closure	83.2	82.1	82.0	80.7	77.9	70.5	52.6	75.6
			Average	77.9	75.0	69.5	62.2	54.8	45.2	28.9	59.1
Hyder *et al.* [15]	ActionMamba	✓	Oral Transit	71.7	69.6	60.1	46.2	40.7	29.1	13.7	47.3
			Soft Palate Elevation	83.6	83.6	80.0	78.0	72.7	64.8	46.6	72.3
			Hyoid Motion	81.4	78.9	76.9	54.6	31.4	13.9	1.6	48.4
			UES Opening	80.6	80.4	80.0	79.8	79.2	71.0	44.2	73.6
			Swallow Initiation	53.6	43.1	24.4	12.8	7.7	3.7	2.0	21.0
			Pharyngeal Transit	86.0	86.0	85.5	84.3	83.4	68.8	52.4	78.0
			Laryngeal Vestibule Closure	80.6	80.6	79.2	78.6	78.3	68.4	54.9	74.4
			Average	76.8	74.6	69.4	62.0	56.2	45.2	30.8	59.3
SG-Mamba (Ours)	ActionMamba	✓	Oral Transit	73.6	69.7	62.7	51.1	39.5	21.5	12.2	47.2
			Soft Palate Elevation	86.4	84.7	84.6	82.9	74.6	67.8	46.2	75.3
			Hyoid Motion	88.0	87.5	82.6	67.1	37.0	23.4	6.6	56.0
			UES Opening	85.9	85.9	85.5	85.2	85.0	83.6	77.8	84.1
			Swallow Initiation	73.0	47.9	34.0	18.0	10.9	5.1	2.2	27.3
			Pharyngeal Transit	90.7	89.8	89.3	89.3	86.5	75.2	54.6	82.2
			Laryngeal Vestibule Closure	84.2	84.0	83.4	82.1	79.3	73.5	60.9	78.2
			Average	83.1	78.5	74.6	67.9	59.0	50.0	37.2	64.3
SG-Mamba (Oracle)	ActionMamba	✓	Oral Transit	91.3	89.4	87.2	61.4	41.1	22.5	12.0	57.8
			Soft Palate Elevation	92.2	91.2	91.1	91.1	85.1	76.1	49.6	82.3
			Hyoid Motion	94.5	93.8	91.7	78.4	58.0	33.4	10.2	65.7
			UES Opening	93.2	93.2	93.2	93.2	92.1	91.4	88.8	92.1
			Swallow Initiation	80.2	49.7	35.4	19.6	11.0	5.5	2.6	29.2
			Pharyngeal Transit	93.6	93.6	93.6	93.6	92.0	85.6	58.6	87.2
			Laryngeal Vestibule Closure	93.6	93.6	93.6	92.6	88.8	82.5	73.9	88.4
			Average	91.2	86.3	83.7	75.7	66.9	56.7	42.2	71.8