A New Cine-MRI Segmentation Method of Tongue Dorsum for Postoperative Swallowing Function Analysis

Minghao Sun[1], Tian Zhou[2], Chenghui Jiang[2], Xiaodan Lv[3] and Han Yu[1]

[1] School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China
[2] Nanjing Medical University, Nanjing, China
[3] The affiliated Shandong Public Health Clinical Center of Shandong University, Jinan, China
han. yu@njupt.edu.cn

Abstract. Advantages of cine-MRI include high spatial-temporal resolution and free radiation, and the technique has become a new method for analyzing and assessing the swallowing function of patients with head and neck tumors. To reduce the labor work of physicians and improve the robustness of labeling the cine-MRI images, we propose a new swallowing analysis method based on a revised cine-MRI segmentation model. This method aims to automate the calculation of tongue dorsum motion parameters in the oral and pharyngeal phases of swallowing, followed by a quantitative analysis. Firstly, based on manually annotated swallowing structures, we propose a method for calculating tongue dorsum motion parameters, which enables the quantitative analysis of swallowing capability. Secondly, a spatial-temporal hybrid model composed of convolution and temporal transformer is proposed to extract the tongue dorsum mask sequence from a swallowing cycle MRI sequence. Finally, to fully exploit the advantages of cine-MRI, a Multi-head Temporal Self-Attention (MTSA) mechanism is introduced, which establishes connections among frames and enhances the segmentation results of individual frames. A Temporal Relative Positional Encoding (TRPE) is designed to incorporate the temporal information of different swallowing stages into the network, which enhances the network's understanding of the swallowing process. Experimental results show that the proposed segmentation model achieves a 1.45% improvement in Dice Score compared to the state-of-the-art methods, and the interclass correlation coefficient (ICC) of the displacement data of swallowing feature points obtained respectively from the model mask and physician annotation exceeds 90%. Our code is available at: https://github.com/MinghaoSam/SwallowingFunctionAnalysis.

Keywords: Swallowing function, Quantitative analysis, Cine-MRI, Head and neck tumor, Temporal attention.

1 Introduction

The rapid rise in the incidence of tongue cancer is a significant contributor to mortality in oral cancer. Various treatments can lead to different degrees of swallowing abnormalities [1]. Postoperative dysphagia occurs in approximately 40-60% of cases [2] and is a common sequela in tongue cancer patients, greatly affecting their quality of living [3]. Despite increasing attention to postoperative dysphagia, there is a lack of research on the mechanism of dysphagia [4].

Videofluoroscopic swallowing study (VFSS) is the most commonly used method for assessing swallowing capacity and is considered the gold standard tool for diagnosing swallowing disorders [5]. However, VFSS carries the risk of radiation exposure and cannot be frequently used. Additionally, due to incomplete visualization in the coronal plane, VFSS cannot provide comprehensive information on swallowing movements. Cine-magnetic resonance imaging (MRI), with its non-invasive, radiation-free, contrast-agent-free, and high temporal and spatial resolution characteristics, has become an increasingly adopted paradigm for evaluating swallowing capacity. During the swallowing process, cine-MRI continuously acquires multiple frames of images, visualizing the dynamic structures of the oral cavity and pharynx. However, previous studies utilizing cine-MRI to investigate swallowing-related structures have primarily relied on qualitative description [6, 7], semi-quantitative analysis[8–10], and quantitative analysis of local anatomical landmarks[11–13]. There is still a lack of quantitative research focusing on the motion trajectories of swallowing-related structures.

Additionally, the aforementioned quantitative analysis methods largely depend on manual annotations of the swallowing structures and personal calculations of swallowing parameters based on anatomical landmarks with experts' experience. This process is time-consuming and labor-intensive. Due to the complexity of the organs and limitations in the annotator's experience, there are drawbacks including subjectivity and susceptibility to errors.

Previously, deformable registration-based methods were used for swallowing motion estimation. For instance, Yang et al.[14] studied the characteristics of tongue root during swallowing with an improved deformable registration algorithm to track its motion in four directions with deformation vectors. However, local tongue root movement in these few directions cannot fully reflect the entire tongue motion.

In this study, to address the limitations of previous quantitative analyses based on cine-MRI, in Section 2, we propose a method for automatically extracting ten feature points on the tongue dorsum and computing their motion parameters based on physician-annotated MRI swallowing sequences. This method can estimate parameters involving displacement, velocity, and acceleration for each feature point, which are used later for quantitative analysis of swallowing function, and its accuracy depends solely on the ROI (tongue dorsum) precision, eliminating further manual intervention. Secondly, to alleviate the burden of physician annotation and mitigate the inconsistency among annotators, we propose a spatial-temporal hybrid tongue dorsum segmentation model. In this model, convolution is utilized to extract spatial features, while the temporal self-attention mechanism to extract temporal features. Moreover, to explore the characteristics of different swallowing phases, we design a temporal relative positional encoding in the temporal transformer, encouraging the model to learn the temporal patterns of the swallowing cycle. Section 3 presents the numerical experiments of our proposed method, and Section 4 is the conclusion part.

2 Methodology

2.1 Tongue Dorsum Motion Computation Method

The method for obtaining tongue dorsum motion parameters is based on morphological and geometric techniques, extracting feature points from multiple frames of annotated images, and calculating the motion parameters of each feature point. The detailed process is illustrated in Fig. 1.

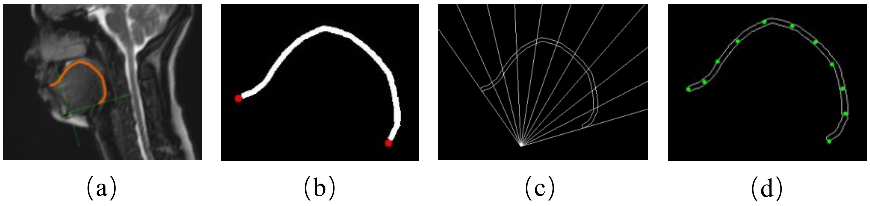


Fig. 1. Illustration of tongue dorsum motion computation method, including (a) the annotated image composes of the tongue dorsum and the registered coordinate axes, (b) the tongue dorsum mask with the endpoints of the tip and the base of the tongue, (c) the intersection between rays and the tongue dorsum contour, and (d) the extracted tongue dorsum feature points.

This tongue motion computation method can be divided into four steps as follows. (1) We establish a Cartesian coordinate system for frame alignment from i to iv. (i) Considering the mandible as the primary support structure for tongue motion [11], we set the lower margin of the attachment point of the genioglossus muscle to the inner side of the mandible as the origin. (ii) The X-axis is defined by the second cervical intervertebral disc due to its fixed position, hence establishing the coordinate system. (iii) We then mark the upper border of the tongue dorsum with integrity and accuracy. (iv) A set of rays is then constructed with the origin as the common starting point. The first and last rays pass through the tip and base of the annotated tongue dorsum, respectively, ensuring that the angles between adjacent rays are equal. (2) The pixel coordinates of the overlap between each ray and the tongue dorsum contour are recorded, and the average coordinates of the overlap pixel belonging to each ray are calculated to obtain the coordinates of the feature points on the tongue dorsum. (3) A spatial transformation is applied to these feature point coordinates to obtain affined coordinates by mapping the pixel coordinates to the registered coordinates system. (4) During one swallowing cycle, the coordinates of ten feature points are extracted from 8 frames of MRI images. The displacement of each feature point is calculated by analyzing the coordinate changes between each pair of adjacent time steps. Subsequently, seven velocity values and six acceleration values are computed based on the displacement data.

2.2 Network Architecture

Medical image segmentation methods mainly focus on improving spatial feature extraction ability. For the cine-MRI data, the temporal corrections between multiple frames are not fully utilized in previous literatures. Therefore, we hypothesize the existence of correlations among multiple frames of images within a swallowing cycle and will introduce it into the neural network. During the oral and pharyngeal phases of swallowing, the predominant movement involves the tongue dorsum and its surrounding structures, with minimal movement observed in other regions. Moreover, there is a coherent trajectory in the movement of the tongue dorsum and its surrounding structures across multiple frames. By integrating features from multiple frames, we can enhance the segmentation results of individual frames. Therefore, we propose the Convolution and Temporal Transformer Hybrid Network (CTTH-Net) for tongue dorsum segmentation. Fig. 2 illustrates an overview of our CTTH-Net architecture.

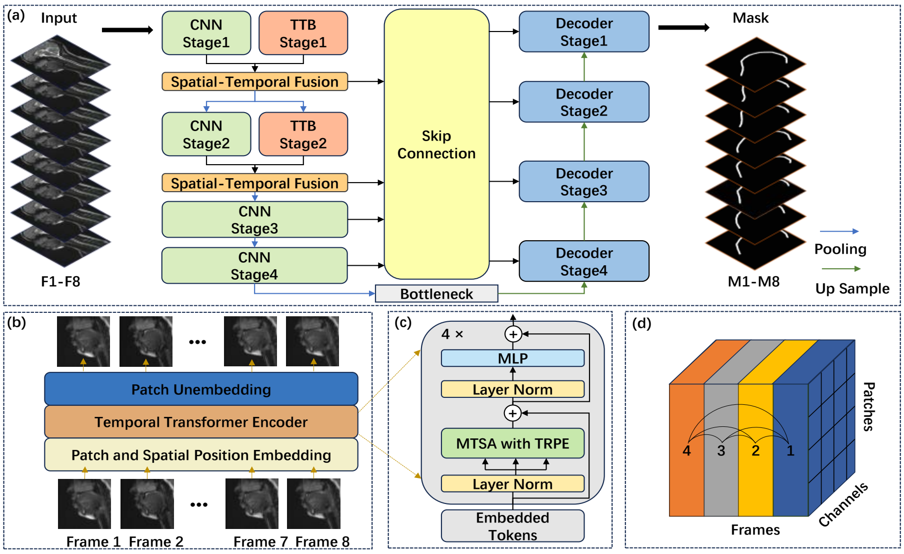


Fig. 2. An overview of the proposed tongue dorsum segmentation model. (a) The Convolution and Temporal Transformer Hybrid Network (CTTH-Net), (b) the data flow of the Temporal Transformer Block (TTB), (c) the architecture of the transformer block inside the TTB, and (d) the proposed temporal attention.

Temporal Transformer Block (TTB). The data flow of TTB is shown in Fig. 2b. We first perform tokenization [15] by reshaping the features into sequences of flattened 2D patches with patch sizes of 16 and 8 in stages 1 and 2, respectively, so that the patches can be mapped to the same areas of the encoder features. The tokenization is involved with convolution and spatial position encoding, and permute operation is performed making tokens $T \in \mathbb{R}^{n_{patches} \times n_{frames} \times C_i}$, where $C_i (i = 1,2)$ denotes channel dimensions of stage i, in our implementation $C_1 = 64, C_2 = 128$. The tokens are then fed into the temporal transformer block, where the operations within the block are repeated four times. Each operation involves a **M**ulti-head **T**emporal **S**elf-**A**ttention module (MTSA) with **T**emporal **R**elative **P**ositional **E**ncoding (TRPE), followed by a **M**ulti-**l**ayer

Perceptron (MLP) with residual structure[16], as shown in Fig. 2c. The MTSA is guided by the following formula:

$$H_i = \text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}} + B\right) V_i, \tag{1}$$

where H_i represents the output of head $i \in [N_h]$, $Q_i = TW_i^Q$, $K_i = TW_i^K$, $V_i = TW_i^V$, T for the input tokens. W_i^Q, W_i^K, W_i^V are trained for each head, and d_k for the dimension of each head, B for temporal relative position bias. We set $N_h = 4$ in implementation.

The major difference of MTSA with respect to the original self-attention [17] is that we conduct the attention operation along the temporal axis (batch axis) rather than the patch-axis, which is shown in Fig. 2 (d). In a N-head attention situation [18], the output after MTSA is calculated by:

$$\text{MTSA} = \underset{i \in N_h}{\text{Concat}}[H_i]W_O, \tag{2}$$

where W_O is an extra parameter matrix that projects the concatenation of the N_h head outputs to the output space. Hereinafter, applying a MLP and residual operators, the output can be obtained by:

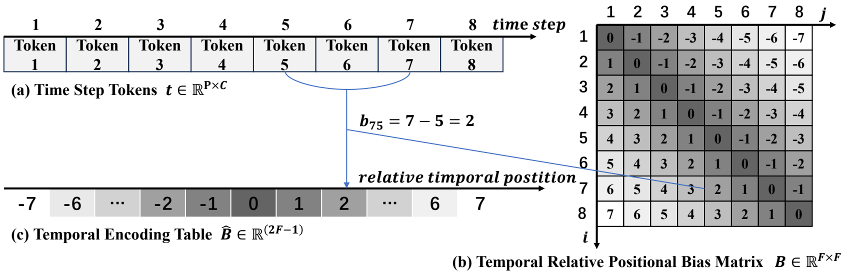
$$\text{TTrans} = \text{MTSA} + \text{MLP}\big(\text{LN}(T + \text{MTSA})\big), \tag{3}$$

where LN represents the layer normalization operation, and TTrans for the output of the temporal transformer block.

Spatial-temporal Fusion. The fusion process is implemented using 1 × 1 convolution. Initially, the feature maps obtained from the convolutional layers and temporal transformer are concatenated along the channel dimension. Subsequently, they undergo dimension reduction through 1 × 1 convolution to obtain feature maps with the same shape as the input of the fusion. This is followed by batch normalization and application of the GELU activation function.

2.3 Temporal Relative Positional Encoding (TRPE)

Unlike CNNs, Transformer models lack an inherent understanding of positional information, thus requiring the introduction of positional encoding [19]. Conventional methods for positional encoding include absolute positional encoding and trainable positional encoding. However, these approaches are not suitable for swallow-related cine-MRI data, which exhibit continuous temporal correlations. Therefore, we propose a temporal encoding technique with relative positions. Similar to the concept of Markov chains, temporal relative positional encoding enables the transformer to better capture the temporal relationships among frames within a swallowing cycle.


1 2 3 4 5 6 7 8 *time step*
Token 1
Token 2
Token 3
Token 4
Token 5
Token 6
Token 7
Token 8
(a) Time Step Tokens $t \in \mathbb{R}^{P \times C}$
$b_{75} = 7 - 5 = 2$
relative timporal postition
-7 -6 ··· -2 -1 0 1 2 ··· 6 7
(c) Temporal Encoding Table $\hat{B} \in \mathbb{R}^{(2F-1)}$
j
i
(b) Temporal Relative Positional Bias Matrix $B \in \mathbb{R}^{F \times F}$

	1	2	3	4	5	6	7	8
1	0	-1	-2	-3	-4	-5	-6	-7
2	1	0	-1	-2	-3	-4	-5	-6
3	2	1	0	-1	-2	-3	-4	-5
4	3	2	1	0	-1	-2	-3	-4
5	4	3	2	1	0	-1	-2	-3
6	5	4	3	2	1	0	-1	-2
7	6	5	4	3	2	1	0	-1
8	7	6	5	4	3	2	1	0

Fig. 3. The temporal relative positional encoding.

A swallowing cycle of 8 frames is mapped to 8 tokens, where each token $t \in \mathbb{R}^{P \times C}$ (refer to Fig. 3a), with P and C denoting the number of patches and channels, respectively. During the temporal transformer process, the attention score is denoted by $QK^T \in \mathbb{R}^{P \times F \times F}$, where F represents the number of frames. Therefore, we introduce the position bias matrix $B \in \mathbb{R}^{F \times F}$ (refer to Fig. 3b), and its elementary values are derived from the temporal encoding table (see Fig. 3c). In this table, temporal relative positions are represented as $\hat{B} \in \mathbb{R}^{2F-1}$. The overall operation can be formulated as follows:

$$b_{ij} = \text{BiasMatrix}[i, j] = \text{EncodingTable}[i - j]. \quad (4)$$

3 Experiments

3.1 Image Acquisition and Dataset

Each swallowing cycle involves the acquisition of 8 MRI images. Initially, cine-MRI images of swallowing were obtained from 9 individuals, comprising 5 healthy subjects and 4 subjects with post-reconstruction surgery. Among them, 3 individuals underwent 4 swallowing cycles, 1 individual underwent 5 cycles, and 5 individuals underwent 3 cycles, resulting in a total of 32 swallowing cycles and 256 captured images.

The dataset is divided into four parts: (1-2) For participants with 4 and 5 collected swallowing cycles, they are grouped into training, validation, and testing sets by the ratios of 2:1:1 and 3:1:1, respectively, on a per-swallowing-cycle basis. (3) For participants with only three collected swallowing cycles, data from two participants are grouped into training, validation, and testing sets by a ratio of 1:1:1 on a per-swallowing-cycle basis. (4) For the remaining three participants, their data are grouped and then split into training, validation, and testing sets by a ratio of 1:1:1 for each participant.

3.2 Implementation Details and Evaluation Metrics

The segmentation model is implemented with Pytorch 2.0.1 using an NVIDIA RTX 3090 GPU. We train the target model for 300 epochs with a batch size of 8 (8 frames of cine-MRI in one swallowing cycle). Adam optimizer is adopted with the learning rate equal to 0.001. To prevent overfitting, we used L2 regularization and set the weight attenuation to 0.001. We adopt a weight binary cross-entropy (BCE) and Dice loss functions, with both weights set to 0.5.

We chose the Dice score and mean intersection over union (mIoU) to evaluate the performance of this method in the tongue dorsum segmentation task. During training, an early stopping strategy is implemented, where training will be stopped if the Dice score fails to exceed the current best model's Dice score for over 100 epochs.

3.3 Comparison with State-of-the-Arts and Ablation Study

We compare our model to several medical image segmentation models, including U-Net [20], ACC-UNet [21], UCTransNet [22] and DSCNet [23]. UCTransNet leverages multi-scale channel-wise cross-attention to mitigate the semantic gap across different stages of the encoder, achieving state-of-the-art performance on multiple publicly available medical image datasets. DSCNet is specifically designed to capture topological tubular structures such as blood vessels and roads. Therefore, its dynamic snake convolution can adaptively focus on slender and tortuous local structures, making it intuitively suitable for tongue dorsum segmentation tasks. During training and testing, the batch size for all models is set to 8, representing eight cine-MRI images of a single swallowing cycle. Comparisons of evaluation metrics of different models are presented in Table 1.

As presented in the bottom rows of Table 1, our model outperforms the second-ranked model, DSCNet, by 1.45% in terms of Dice score and surpasses the second-ranked model, UCTransNet, by 1.29% in terms of mIoU, achieving the best performance. Regarding model efficiency, it is noteworthy that our model maintains comparable parameter and FLOPs counts to the baseline U-Net while achieving superior segmentation performance. The ablation study demonstrates that the inclusion of TTB in the baseline model resulted in a 0.9% and 0.2% improvement in Dice score and mIoU, respectively. Subsequently, the addition of TRPE leads to a further 1.41% and 1.33% in Dice score and mIoU, respectively.

Table 1. Comparisons of different models, where (w/o) and (w) represent models without and with TRPE, respectively. FLOPs are tested with batch size of 8.

Model	Params(M)	FLOPs(T)	Dice	mIoU
U-Net (baseline)	**14.751**	**0.393**	62.91	47.84
ACC-UNet	16.771	0.708	59.93	44.10
UCTransNet	63.096	0.953	63.73	*48.08*
DSCNet	23.924	1.613	63.77	47.98
Ours (w/o)	*14.845*	*0.414*	*63.81*	48.04
Ours (w)	*14.845*	*0.414*	**65.22**	**49.37**

3.4 Results

Fig. 4 illustrates the segmentation results of different models for tongue dorsum. It can be observed that our method outperforms other models in extracting tongue dorsum masks. Benefitted from the application of TTB and TRPE, long-range temporal dependencies between frames are established, enhancing the segmentation performance of individual frames, thus preserving target structures and better continuity.

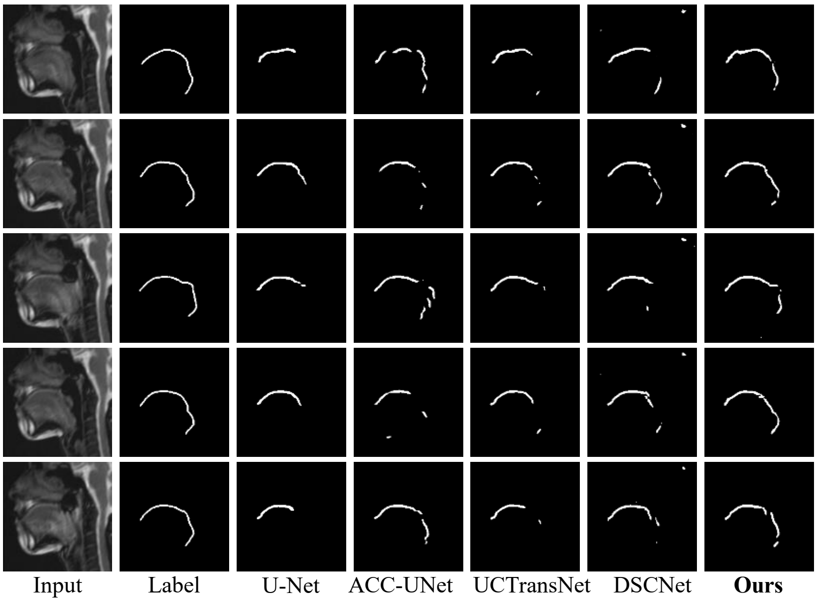


Fig. 4. Output comparisons of different models. In the outputs of U-Net and UCTransNet, incomplete segments are observed in the tongue dorsum structure. Compared to other models, the segmentation accuracy of ACC-UNet is unsatisfactory. Although DSCNet demonstrates improved continuity in the tongue dorsum predictions, it is susceptible to intensity interference from liquid during later stages of swallowing. Leveraging the temporal transformer, our model establishes correlations between multiple frames, resulting in superior segmentation integrity.

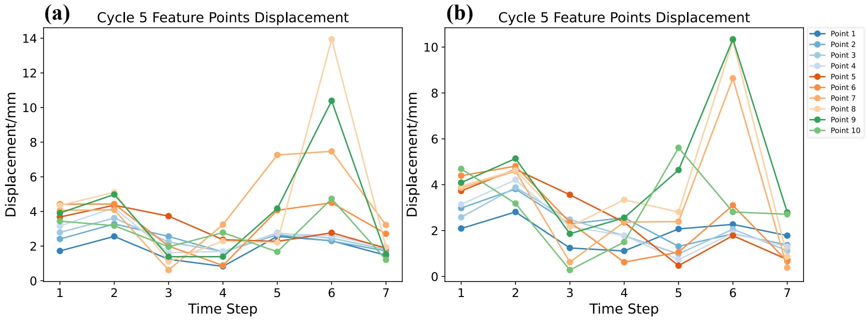


Fig. 5. Displacements of feature points in one swallowing cycle, calculated from (a) physician's annotations, and (b) our model-generated masks.

In our study, intraclass correlation coefficient (ICC) [24] was employed to evaluate the consistency between displacement data calculated from the masks extracted from physician's annotation (refer to Fig. 5a) and the masks generated by our model (refer to Fig. 5b). The ICC results exceeded 90%, indicating the effectiveness and accuracy of our approach in capturing the tongue dorsum motion.

4 Conclusion

This study introduces a novel swallowing analysis method based on a tongue dorsum segmentation neural network. Firstly, we propose a method for calculating tongue dorsum motion parameters from tongue dorsum masks for quantitative analysis. Secondly, we present a tongue dorsum segmentation model with temporal transformer. Experimental results demonstrate the robustness and effectiveness of the model's segmentation ability, and the motion parameters calculated from model extracted masks demonstrate the consistency with those obtained from physician annotation. Future work will focus on further refining the model and exploring its application to other planes of swallowing-related cine-MRI.

Acknowledgments. This work was funded in part by the Natural Science Foundation of China under Grants 12371440, and in part by the Natural Science Foundation of Nanjing University of Posts and Telecommunications under Grant NY222140.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Pauloski, B.R.: Rehabilitation of Dysphagia Following Head and Neck Cancer. Physical Medicine and Rehabilitation Clinics of North America. **19**(4), 889–928 (2008)
2. Huang, Z., Chen, W., Huang, Z., Yang, Z.: Dysphagia in Tongue Cancer Patients Before and After Surgery. Journal of Oral and Maxillofacial Surgery. **74**(10), 2067–2072 (2016)
3. García-Peris, P., Parón, L., Velasco, C., de la Cuerda, C., Camblor, M., Bretón, I., Herencia, H., Verdaguer, J., Navarro, C., Clave, P.: Long-term prevalence of oropharyngeal dysphagia in head and neck cancer patients: Impact on quality of life. Clinical Nutrition. **26**(6), 710–717 (2007)
4. Deng, W., Zhao, G., Li, Z., Yang, L., Xiao, Y., Zhang, S., Guo, K., Xie, C., Liang, Y., Liao, G.: Recovery pattern analysis of swallowing function in patients undergoing total glossectomy and hemiglossectomy. Oral Oncology. **132**, 105981 (2022)
5. East, L., Nettles, K., Vansant, A., Daniels, S.K.: Evaluation of oropharyngeal dysphagia with the videofluoroscopic swallowing study. Journal of Radiology Nursing. **33**(1), 9–13 (2014)
6. Hartl, D.M., Kolb, F., Bretagne, E., Bidault, F., Sigal, R.: Cine-MRI swallowing evaluation after tongue reconstruction. European Journal of Radiology. **73**(1), 108–113 (2010)
7. Hartl, D.M., Albiter, M., Kolb, F., Luboinski, B., Sigal, R.: Morphologic parameters of normal swallowing events using single-shot fast spin echo dynamic MRI. Dysphagia. **18**, 255–262 (2003)
8. Kreeft, A.M., Rasch, C.R., Muller, S.H., Pameijer, F.A., Hallo, E., Balm, A.J.: Cine MRI of swallowing in patients with advanced oral or oropharyngeal carcinoma: a feasibility study. European Archives of Oto-Rhino-Laryngology. **269**, 1703–1711 (2012)

9. Nishimura, S., Tanaka, T., Oda, M., Habu, M., Kodama, M., Yoshiga, D., Osawa, K., Kokuryo, S., Miyamoto, I., Kito, S.: Functional evaluation of swallowing in patients with tongue cancer before and after surgery using high-speed continuous magnetic resonance imaging based on T2-weighted sequences. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. **125**(1), 88–98 (2018)
10. Joujima, T., Oda, M., Sasaguri, M., Habu, M., Kataoka, S., Miyamura, Y., Wakasugi-Sato, N., Matsumoto-Takeda, S., Takahashi, O., Kokuryo, S.: Evaluation of velopharyngeal function using high-speed cine-magnetic resonance imaging based on T2-weighted sequences: a preliminary study. International Journal of Oral and Maxillofacial Surgery. **49**(4), 432–441 (2020)
11. Kim, Y.C., Lee, S.J., Park, H., Choi, Y.J., Jeong, W.S., Lee, Y.S., Choi, K.H., Oh, T.S., Choi, J.W.: Swallowing analysis in hemi-tongue reconstruction using motor-innervated free flaps: A cine-magnetic resonance imaging study. Head & Neck. **45**(5), 1097–1112 (2023).
12. Kitano, H., Asada, Y., Hayashi, K., Inoue, H., Kitajima, K.: The evaluation of dysphagia following radical surgery for oral and pharyngeal carcinomas by cine-magnetic resonance imaging (Cine-MRI). Dysphagia. **17**, 187–191 (2002)
13. Ha, J., Sung, I., Son, J., Stone, M., Ord, R., Cho, Y.: Analysis of speech and tongue motion in normal and post-glossectomy speaker using cine MRI. Journal of Applied Oral Science. **24**, 472–480 (2016)
14. Yang J, Mohamed A S R, Bahig H, et al.: Automatic registration of 2D MR cine images for swallowing motion estimation. PLoS One. **15**(2): e0228652 (2020)
15. Dosovitskiy A, Beyer L, Kolesnikov A, et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. IEEE (2016)
17. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
18. Cordonnier, Jean-Baptiste, Andreas Loukas, and Martin Jaggi.: Multi-head attention: Collaborate instead of concatenate. arXiv preprint arXiv:2006.16362 (2020)
19. Wu, K., Peng, H., Chen, M., Fu, J., Chao, H.: Rethinking and Improving Relative Position Encoding for Vision Transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10013–10021. IEEE (2021)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., and Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. Springer International Publishing, Cham (2015)
21. Ibtehaz, N., Kihara, D.: ACC-UNet: A Completely Convolutional UNet Model for the 2020s. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 692–702. Springer Nature Switzerland, Cham (2023)
22. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. **36**(3), 2441–2449 (2022)
23. Qi Y, He Y, Qi X, et al.: Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6070-6079. IEEE, (2023)
24. Koo, T.K., Li, M.Y.: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine. **15**, 155–163 (2016)