Communication

Reinforcement Learning Based Vocal Fold Localization in Preoperative Neck CT for Injection Laryngoplasty

Walid Abdullah Al [1], Wonjae Cha [2] and Il Dong Yun [3,†,*]

[1] Department of Computer Engineering, Hankuk University of Foreign Studies, Yongin, Korea; walid@hufs.ac.kr
[2] Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea; chawonjae@gmail.com
[3] Department of Computer Engineering, Hankuk University of Foreign Studies, Yongin, Korea; yun@hufs.ac.kr
* Correspondence: yun@hufs.ac.kr; Tel.: +82-31-330-4260
† Current address: Department of Computer Engineering, Hankuk University of Foreign Studies, 81 Oedae-ro, Yongin-si, 17035 Gyeonggi-do, Korea.

Abstract: Transcutaneous injection laryngoplasty is a well-known procedure for treating paralyzed vocal fold by injecting augmentation material to it. Hence, vocal fold localization plays a vital role in the preoperative planning as the fold location is required to determine the optimal injection route. In this communication, we propose a mirror environment based reinforcement learning (RL) algorithm for localizing the right and left vocal folds in preoperative neck CT. RL-based methods commonly showed noteworthy outcome in general anatomic landmark localization problem in the recent years. However, such methods suggest training individual agent for localizing each fold, though the right and left vocal folds are located in close proximity and have high feature-similarity. Utilizing the lateral symmetry between the right and left vocal folds, the proposed mirror environment allows for a single agent for localizing both the folds by treating the left fold as a flipped version of the right fold. Thus, localization of both folds can be trained using a single training session which utilizes the inter-fold correlation and avoids redundant feature learning. Experiment with 120 CT volumes showed improved localization performance and training efficiency of the proposed method compared with the standard RL method.

Keywords: injection laryngoplasty; neck CT; vocal fold localization; deep learning; reinforcement learning; mirror environment

1. Introduction

Vocal fold paresis is a common condition among older patients [1], which is characterized by the paralysis of any of the right (RVF) and left vocal folds (LVF) or both. Besides causing voice discomfort, breathing and swallowing difficulty, it can cause harm to the respiratory organs because of the glottal gap created by the paralyzed fold, therefore requiring immediate treatment [2]. One of the most common treatment procedure for this is the transcutaneous injection laryngoplasty (TIL) [3], where the glottal gap is filled by injecting augmentation material to the affected vocal fold. During the preoperative planning with neck CT, accurate localization of the vocal folds is required for estimating the optimal injection route [4]. Thus, an automatic vocal fold localization method using neck CT can be potentially useful for guiding and accelerating the preoperative planning process.

To the best of our knowledge, no computational approach currently exists for vocal fold localization or injection route identification. Some manual approaches can be found where vocal fold and needle routes are studied using neck CT across different patients [4] or a 3D printed larynx [5,6]. In general anatomical landmark localization problem, deep learning based heatmap regression methods are widely used where spatial heatmaps around the landmarks are usually regressed [7]. However, such heatmaps constitute a negligible foreground region compared to the huge 3D background causing sample bias [8]. On the other hand, deep reinforcement learning (RL) based localization approaches

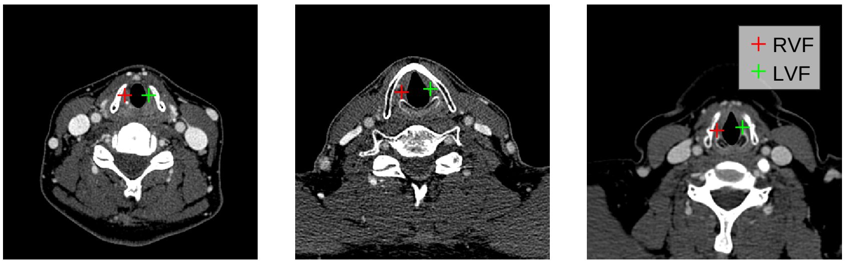

Figure 1. RVF and LVF shown in axial slices of three neck CT volumes.

suggest a sequential decision process involving a finer pixel-to-pixel navigation combined with deep feature learning, thereby producing improved predictions [9–11].

In this communication, we propose a RL-based method to localize the right and left folds in preoperative neck CT. Owing to the existing RL formulation [8,10,11], the proposed agent takes sequential steps to navigate through the voxel space to finally converge to the target fold location. Starting from a random initial position, the steps are decided based on local features learnt by a deep convolutional neural network (CNN). However, existing RL-based localization methods commonly require an independent agent to be separately trained for each target landmark. On the other hand, the vocal folds have high feature-similarity as a result of the laryngeal symmetry (see Fig 1). Moreover, they are located close to each other in the same area of the larynx. Therefore, redundant features would be learnt during the separate training sessions of the two agents. Besides, such training would also fail to utilize the high feature correlation between the folds.

To benefit from the inter-fold similarity, we propose a mirror environment which allows for a single agent to localize both the RVF and LVF position. The agent localizes the RVF in the original volume and the LVF in the laterally flipped volume, essentially considering the flipped LVF to be the same target as the RVF due to the laryngeal symmetry. In this mirror environment, the agent can be efficiently be trained using a single session avoiding redundant feature learning, while improving localization performance by sharing correlated features from the both the right and left folds during training.

2. Methodology

In our deep RL-based method, we automatically localize the RVF and LVF in preoperative neck CT. Neck CT can be represented by a 3D volume V with the three cartesian axes X, Y, and Z indicating towards *right*-to-*left*, *anterior*-to-*posterior*, and *superior*-to-*inferior*, respectively. Inside this volume, we denote the right and left fold positions by $\boldsymbol{p}_r$ and $\boldsymbol{p}_l$. Specifically, these positions are identified as the posterior vocal fold landmarks.

The basic agent interaction in the proposed RL environment is similar to the standard formulation commonly found in all existing RL-based studies [8,10,11]. Usually, an agent localizes a target landmark $\boldsymbol{p}_{\text{goal}}$ by taking an episode of sequential steps starting from a random initial voxel inside the environment (i.e., input CT volume). At any step t, it observes the current state $\boldsymbol{s}_t$ as the local patch centered at the current voxel $\boldsymbol{p}_t$ and takes an action $a_t \in \{x^+, x^-, y^+, y^-, z^+, z^-\}$ to move one voxel forward or backward along any of the axes. Thus, it updates to a neighboring voxel $\boldsymbol{p}_{t+1}$ and transitions to a new state $\boldsymbol{s}_{t+1}$. A policy $\pi(\boldsymbol{s}_t, a_t)$ gives the optimal probability for choosing an action a_t in a given state $\boldsymbol{s}_t$. Fig 2 illustrates the agent-interaction with the volume. Similar to the existing studies [8,10,11], the local patch size of the state was set to 64 × 64 × 64 voxels.

During training, a scalar reward r_t is provided after each step where a positive reward is assigned for a step towards the target and a negative reward is assigned otherwise. Specifically, the reward for transition from position $\boldsymbol{p}_t$ to $\boldsymbol{p}_{t+1}$ by action a_t can be indicated as follows:

$$r_t = \begin{cases} -1 & \text{if } |\boldsymbol{p}_{\text{goal}} - \boldsymbol{p}_{t+1}| > |\boldsymbol{p}_{\text{goal}} - \boldsymbol{p}_t| \\ +1 & \text{if } |\boldsymbol{p}_{\text{goal}} - \boldsymbol{p}_{t+1}| \leq \max(|\boldsymbol{p}_{\text{goal}} - \boldsymbol{p}_t|, \tau) \end{cases} \tag{1}$$

doi:10.20944/preprints202211.0234.v1

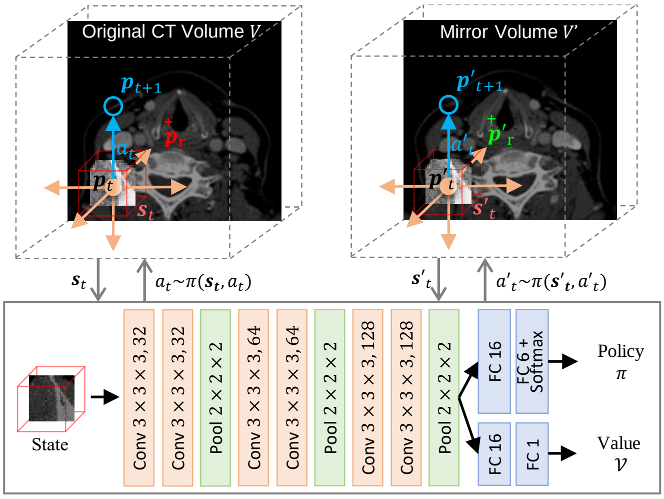

Figure 2. Proposed mirror environment. A single RL agent localizes the right fold in the original volume and left fold in the flipped volume.

where τ represents a small distance within which the agent constantly receives positive rewards, indicating to the convergence state. For training, the agent-trajectories after multiple episodes of transitions is recorded, where each transition can be denoted by $(\boldsymbol{s}, a, \tilde{\boldsymbol{s}}, r)$ where $\boldsymbol{s} = \boldsymbol{s}_t, a = a_t, \tilde{\boldsymbol{s}} = \boldsymbol{s}_{t+1}$, and $r = r_t$. The goal of training is to optimize the policy so that the expected cumulative reward over such trajectories is maximized.

In multi-landmark situation, individual agent-policies are usually trained on different trajectories gather for different landmarks [8,10,11]. However, this is sub-optimal and inefficient because RVF and LVF in our problem have high similarity in feature and proximity in location. In the following, we describe the mirror environment that allows for efficient and improved training of a single agent-policy to localize both the RVF and LVF.

2.1. Mirror Environment

The proposed mirror environment is solely based on the the laryngeal symmetry where each of the two vocal folds is almost a mirrored version of the other. We first model the left vocal fold $\boldsymbol{p}_l$ in volume V as the right fold $\boldsymbol{p}'_r$ in volume V', a laterally (along X-axis) flipped version of the original volume V. Then, we propose a single RL agent to localize both the right vocal fold (as $\boldsymbol{p}_r$) and left vocal fold (as $\boldsymbol{p}'_r$), essentially arguing $\boldsymbol{p}_r$ and $\boldsymbol{p}'_r$ to be the same target landmark with similar feature.

With this new model, agent interaction with the environment during localization remains the same. Except, the environment now can either be the original volume V or the mirrored volume V', depending on the target being the right fold $\boldsymbol{p}_{\text{goal}} = \boldsymbol{p}_r$ or the (mirrored) left fold $\boldsymbol{p}_{\text{goal}} = \boldsymbol{p}'_r$, respectively. Consequently, the reward in (1) is also calculated as r_t or r'_t based on the target.

Similar to the previous RL frameworks [8,10,11], we represent our policy function by a 3D CNN which outputs optimal action-probabilities $\pi(\boldsymbol{s}, a)$ for an input state $\boldsymbol{s}$. The CNN consists of three convolutional blocks each having two convolutional layers followed by a max-pooling layer (see Fig 2). The last convolutional block is connected to two fully connected layers to produce the action-probabilities. Algorithm 1 summarizes the policy training process with the mirror environment.

At each epoch, we conduct a number of localization episodes. Each episode is initiated by sampling a training volume V and a target landmark $\boldsymbol{p}_{\text{goal}}$ between the right and left folds. Based on the sampled target, the environment is set with either the original volume V or mirrored volume V'. Then, subsequent steps are performed applying the current policy. Running multiple episodes, we record the trajectories for the original and mirror volumes as $\mathbb{T}$ and $\mathbb{T}'$, respectively.

Algorithm 1: RL Policy Training with Mirror Environment

```
Input: Training CT volumes $\boldsymbol{V}_{\text{train}}$, episodes per epoch $E$
Output: Optimized policy $\pi$
$\pi \leftarrow$ random policy
repeat
    Trajectories $\mathbb{T} = \emptyset$, mirror trajectories $\mathbb{T}' = \emptyset$
    for episode $= 1$ to $E$ do
        Sample a volume $V \in \boldsymbol{V}_{\text{train}}$ and a target $\boldsymbol{p}_{\text{goal}} \in \{\boldsymbol{p}_r, \boldsymbol{p}_l\}$
        if $\boldsymbol{p}_{\text{goal}} = \boldsymbol{p}_l$ then
            $V = V' \leftarrow$ laterally flipped $V$
            $\boldsymbol{p}_{\text{goal}} = \boldsymbol{p}'_r \leftarrow$ laterally flipped $\boldsymbol{p}_l$
            Gather trajectories $(\boldsymbol{s}', a', \tilde{\boldsymbol{s}}', r')$into $\mathbb{T}'$ applying $\pi$
        else
            Gather trajectories $(\boldsymbol{s}, a, \tilde{\boldsymbol{s}}, r)$ into $\mathbb{T}$ applying $\pi$
    Optimize $\pi$ for the PPO objective in (3) on $\mathbb{T}$ and $\mathbb{T}'$
until convergence
```

Now, the policy is updated so that expected cumulative reward on both $\mathbb{T}$ and $\mathbb{T}'$ is maximized. We follow the widely used proximal policy optimization (PPO) framework [12] where *advantage* (i.e., the improvement of the current policy over the previous) is maximized to reduce variance during training. The improvement is estimated using a value function $\mathcal{V}(\boldsymbol{s}; \pi)$ that evaluates the discounted cumulative reward of the current policy π over a transition $(\boldsymbol{s}, a, \tilde{\boldsymbol{s}}, r)$ as: $r + \gamma\mathcal{V}(\tilde{\boldsymbol{s}}; \pi_{\text{prev}})$, with $0 < \gamma < 1$ being the discount factor. This value network is also built on top of the final convolutional layer of the policy (see Fig 2). Value is also updated at each epoch to minimize the difference between the predicted value and the observed discounted reward in the gathered trajectories. Now the advantage is computed as follows:

$$A(\boldsymbol{s}, \tilde{\boldsymbol{s}}, r) = \big(r + \gamma\mathcal{V}(\tilde{\boldsymbol{s}}; \pi_{\text{prev}})\big) - \mathcal{V}(\tilde{\boldsymbol{s}}; \pi_{\text{prev}}) \tag{2}$$

Now, the PPO training objective for maximizing advantage in our mirror environment can be expressed as follows:

$$\begin{aligned} \max_{\pi} \mathbb{E}_{(\boldsymbol{s},a,\tilde{\boldsymbol{s}},r)\in\mathbb{T}} \Big[\frac{\pi(\boldsymbol{s}, a)}{\pi_{\text{prev}}(\boldsymbol{s}, a)} A(\boldsymbol{s}, \tilde{\boldsymbol{s}}, r)\Big] \\ + \mathbb{E}_{(\boldsymbol{s}',a',\tilde{\boldsymbol{s}}',r')\in\mathbb{T}'} \Big[\frac{\pi(\boldsymbol{s}', a')}{\pi_{\text{prev}}(\boldsymbol{s}', a')} A(\boldsymbol{s}', \tilde{\boldsymbol{s}}', r')\Big] \end{aligned} \tag{3}$$

The above objective suggests that the policy for an action is increased if the advantage is positive and decreased if the advantage is negative. Usually, the policy ratio (π to π_{prev}) in (3) is clipped between $[0.8, 1.2]$ to avoid large policy updates. Thus, a single policy is optimized to localize both the RVF and the flipped LVF.

3. Results

3.1. Data

To evaluate the performance of the proposed method, we collected 120 neck CT volumes from Seoul National University Bundang Hospital, South Korea. Among the 120 patients (62 ± 14 years old, 48% femle), 27 patients had vocal fold paresis. The axial slice (X-Y axes) dimension was consistently 512×512 voxels, whereas the number axial slices per volume (Z-axis dimension) was 172 on average. The average voxel-spacing was 0.44 mm × 0.44 mm × 0.95 mm. The vocal fold locations were annotated by an expert with more than 10 years experience in injection laryngoplasty.

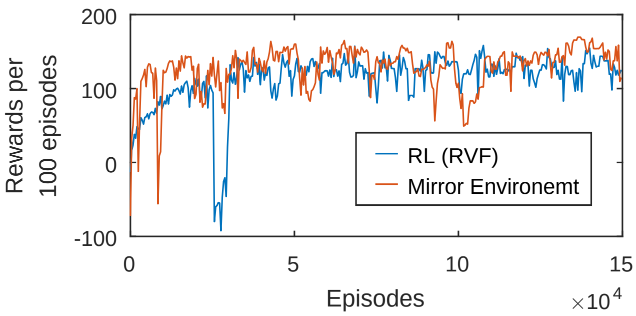

Figure 3. Reward plot over the training episodes. Proposed agent can learn to localize both folds in a similar time the usual agent learn a single fold.

Table 1. Number of Required Episodes and Final Reward

Methods	Required Episodes	Final Reward
Standard RL (RVF)	110k ± 5k	126.495 ± 14.257
Standard RL (LVF)	96k ± 4k	137.152 ± 17.309
Standard RL (RVF and LVF)	206k ± 8k	131.823 ± 15.783
Mirror Environment	**122**k ± 5k	**147.296 ± 13.679**

3.2. Evaluation Method and Performance Metric

We performed a 4-fold cross validation on the 120 volumes to assess the average performance of our algorithm. We hypothesized that the proposed algorithm can result in improved localization and training because of its utilization of the feature similarity between the two folds. Therefore, we use two metrics to validate our proposed contributions: (i) training efficiency (number of episodes explored until convergence) and (ii) localization performance (localization error and accuracy). Localization error is measured by the Euclidean distance of the localized landmark from the expert annotation. Localization accuracy is measured by the percentage of acceptable localization results as evaluated by the expert.

For comparison, we also applied the standard RL-based localization [11] (with individual training method) as our main baseline method. Furthermore, we applied the widely used deep learning based end-to-end localization approaches, e.g. direct location/coordinate regression and heatmap regression [7], to compare the localization performance.

3.3. Training Efficiency

We plot the average episodic rewards over the training epochs of the standard RL agents along with the proposed one in Fig 3. For all the agents, 400 episodes were explored at each epoch where each episode consisted of 200 steps. All other hyperparameters were also kept identical between all the agents for fair comparison. The learning rate and the discount factor γ were set to $1e-4$ and 0.99, following the convention of the previous works [8,10,11]. The average number of episodes required for training the independent agents was about 206k (110k for RVF and 96k for LVF). On the other hand, the proposed RL agent only required about 122k episodes to successfully learn to localize both the folds. Thus, we could achieve almost two times faster training by the simultaneous training.

In Table 1, we also report the average final reward when the training converges in different sessions of the 4-fold cross validation process. The final reward achieved by the individual agent method was 131.823 ± 15.783, whereas the proposed agent could achieve a significantly higher reward of 147.296 ± 13.679 (p-value < 0.01).

3.4. Localization Performance

Besides the higher training efficiency, the proposed method also outperformed the standard RL-based localization approach significantly (p-value < 0.03). The individual

doi:10.20944/preprints202211.0234.v1

Table 2. Localization Error and Accuracy

Methods	Loc. Error (mm)	Loc. Accuracy
Deep Heatmap Regression [7]	2.92 ± 1.88	82.50%
Deep Coordinate Regression [13]	3.70 ± 2.02	77.08%
Standard RL [11]	2.64 ± 1.38	91.25%
Mirror Environment	$\mathbf{2.33 \pm 1.17}$	**95.83%**

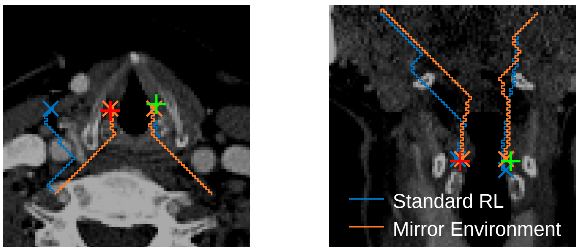

Figure 4. Agent-trajectories for localizing the vocal folds in two CT volumes (left: axial view, right: coronal view). Localization failure (left) or higher localization error (right) for one fold despite correct localization of the other fold by the standard RL. Mirror environment showed improvement by employing similar policies for both.

agents in the usual RL method [11] could give an average localization error of 2.64 ± 1.38 mm yielding improvement over the end-to-end deep learning methods [7,13] (see Table 2). The proposed mirror environment based RL could further lower the error to 2.33 ± 1.17 mm with its single agent utilizing the inter-fold similarity. Table 2 also presents the corresponding localization accuracy for each method.

In Fig 4, we present two examples of improvements by the proposed method, where the localization trajectories of the standard RL agents and the mirror environment agent are plotted. Being ignorant of the inter-fold symmetry and similarity, The standard RL agents gave localization failure (Fig 4-*left*) or higher error (Fig 4-*right*) for one fold despite correctly localizing the other. To the contrary, the proposed agent could improve on both the cases by enforcing a similar and more general policy for both the folds, resulted from the implicit sharing of RVF and LVF features during the simultaneous training.

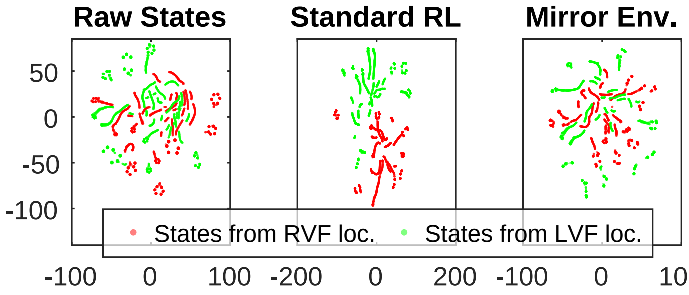

Figure 5. Manifold representation of the learnt features by different agents. Inter-fold similarity was better retained in features of the proposed agent.

To further illustrate the improved generalization and feature-similarity utilization of the proposed method, we compared the learnt feature representation of different agents (Fig 5). We collect about 6000 states from multiple episodes employing the standard RL agents to localize the right and left folds on different volumes. For these states, we then extract the corresponding features (outputs of the final convolution layer) of the policy networks of the standard and proposed agents. Manifold representation [14] is used to effectively plot these high dimensional features into 2-D space.

State similarity between right and left folds can clearly be observed in Fig 5-*left*, where raw voxel array of the state is represented. However, the standard RL agents learnt different representations for the two folds despite their general similarity. On the other hand, feature representation of the proposed agent better retained the general trend of the original state-similarity, resulting in improved localization policies.

4. Conclusion

We proposed a mirror environment for localizing the RVF and LVF in neck CT using RL. Modeling the LVF as a flipped RVF, the proposed method could train a single agent for both folds. Compared to the individual agents in existing RL methods, the proposed method could give higher training efficiency and localization accuracy by effectively utilizing the inter-fold similarity and anatomical symmetry. For our future work, we plan to collect datasets from different sites and expand the method for actual injection route planning based on the localized folds.

Author Contributions: Conceptualization, methodology, Walid Abdullah Al and Il Dong Yun; software, Walid Abdullah Al; validation, Wonjae Cha; formal analysis, Walid Abdullah Al; investigation, Wonjae Cha and Il Dong Yun; resources, data curation, Wonjae Cha; writing—original draft preparation, Walid Abdullah Al; writing—review and editing, Wonjae Cha and Il Dong Yun; visualization, Walid Abdullah Al; supervision, Il Dong Yun and Wonjae Cha; project administration, Il Dong Yun; funding acquisition, Il Dong Yun. All authors have read and agreed to the published version of the manuscript

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science, Technology (No. 2019R1A2C1085113)

Institutional Review Board Statement: Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of Seoul National University Bundang Hospital (IRB approval number: B-2202-738-105

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

LVF	Left vocal fold
RVF	Right vocal fold
RL	Reinforcement learning

References

1. Ahmad, S.; Muzamil, A.; Lateef, M. A study of incidence and etiopathology of vocal cord paralysis. *Indian Journal of Otolaryngology and Head and Neck Surgery* **2002**, *54*, 294–296.
2. Tsai, M.S.; Yang, Y.H.; Liu, C.Y.; Lin, M.H.; Chang, G.H.; Tsai, Y.T.; Li, H.Y.; Tsai, Y.H.; Hsu, C.M. Unilateral vocal fold paralysis and risk of pneumonia: a nationwide population-based cohort study. *Otolaryngology–Head and Neck Surgery* **2018**, *158*, 896–903.
3. Chhetri, D.K.; Jamal, N. Percutaneous injection laryngoplasty. *The Laryngoscope* **2014**, *124*, 742.
4. Nasir, Z.M.; Azman, M.; Baki, M.M.; Mohamed, A.S.; Kew, T.Y.; Zaki, F.M. A proposal for needle projections in transcutaneous injection laryngoplasty using three-dimensionally reconstructed CT scans. *Surgical and Radiologic Anatomy* **2021**, pp. 1–9.
5. Lee, M.; Ang, C.; Andreadis, K.; Shin, J.; Rameau, A. An Open-Source Three-Dimensionally Printed Laryngeal Model for Injection Laryngoplasty Training. *The Laryngoscope* **2021**, *131*, E890–E895.
6. Hamdan, A.L.; Haddad, G.; Haydar, A.; Hamade, R. The 3D printing of the paralyzed vocal fold: added value in injection laryngoplasty. *Journal of Voice* **2018**, *32*, 499–501.
7. Payer, C.; Štern, D.; Bischof, H.; Urschler, M. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Medical image analysis* **2019**, *54*, 207–219.

doi:10.20944/preprints202211.0234.v1

8. Abdullah Al, W.; Yun, I.D. Partial Policy-Based Reinforcement Learning for Anatomical Landmark Localization in 3D Medical Images. *IEEE Transactions on Medical Imaging* **2020**, *39*, 1245–1255. https://doi.org/10.1109/TMI.2019.2946345.
9. Ghesu, F.C.; Georgescu, B.; Mansi, T.; Neumann, D.; Hornegger, J.; Comaniciu, D. An artificial agent for anatomical landmark detection in medical images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 229–237.
10. Ghesu, F.C.; Georgescu, B.; Grbic, S.; Maier, A.; Hornegger, J.; Comaniciu, D. Towards intelligent robust detection of anatomical structures in incomplete volumetric data. *Medical image analysis* **2018**, *48*, 203–213.
11. Alansary, A.; Oktay, O.; Li, Y.; Le Folgoc, L.; Hou, B.; Vaillant, G.; Kamnitsas, K.; Vlontzos, A.; Glocker, B.; Kainz, B.; et al. Evaluating reinforcement learning agents for anatomical landmark detection. *Medical image analysis* **2019**, *53*, 156–164.
12. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* **2017**.
13. Lv, J.; Shao, X.; Xing, J.; Cheng, C.; Zhou, X. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3317–3326.
14. LJPvd, M.; Hinton, G. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* **2008**, *9*, 9.