

One-shot Learning Landmarks Detection

Zihao Wang, Clair Vandersteen, Charles Raffaelli, Nicolas Guevara, François Patou, Hervé Delingette

▶ **To cite this version:**

Zihao Wang, Clair Vandersteen, Charles Raffaelli, Nicolas Guevara, François Patou, et al.. One-shot Learning Landmarks Detection. MICCAI 2021 - Workshop on Data Augmentation, Labeling, and Imperfections, Oct 2021, strasbourg, France. pp.163-172, 10.1007/978-3-030-88210-5_15 . hal-03024759v2

HAL Id: hal-03024759

https://inria.hal.science/hal-03024759v2

Submitted on 11 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

One-shot Learning for Landmarks Detection

Zihao Wang[1], Clair Vandersteen[2], Charles Raffaelli[2], Nicolas Guevara[2], François Patou[3], and Hervé Delingette[1]

[1] Université Côte d'Azur, Inria, Epione Team, France
zihao.wang@inria.fr
[2] Université Côte d'Azur, Nice University Hospital, France
[3] Oticon Medical, France

Abstract. Landmark detection in medical images is important for many clinical applications. Learning-based landmark detection is successful at solving some problems but it usually requires a large number of the annotated datasets for the training stage. In addition, traditional methods usually fail for the landmark detection of fine objects. In this paper, we tackle the issue of automatic landmark annotation in 3D volumetric images from a single example based on a one-shot learning method. It involves the iterative training of a shallow convolutional neural network combined with a 3D registration algorithm in order to perform automatic organ localization and landmark matching. We investigated both qualitatively and quantitatively the performance of the proposed approach on clinical temporal bone CT volumes. The results show that our one-shot learning scheme converges well and leads to a good accuracy of the landmark positions.

Keywords: One-shot learning · Landmarks detection · Deep Learning

1 Introduction

Landmarks detection for target object localization plays a pivotal role in many imaging tasks. Automatic landmark detection can alleviate the challenges of image annotation by human experts and can also save time for many image processing tasks. The difficulty of landmark detection in clinical images may come from anatomical variability, or changes in body position which can lead to large differences of shape or appearance. The literature on automatic landmarks detection approaches can be roughly split into learning based versus non-learning based algorithms.

Non-Learning based landmarks detection in [1] is proposed the augmentation of the scale-invariant feature transform (SIFT) to arbitrary n dimensions (n-SIFT) for 3D-MRI volumes. However, the computation cost for 3D SIFT features is heavy as their complexity is a cubic function of the image size. Wörz *et al.* [2] leverage parametric intensity models for image landmarks detection. Ricardo *et al.* [3] use log-Gabor filters to extract frequency features for 3D Phase Congruency (PC) applied to detect head and neck landmarks.

This work was partially funded by the French government, by the National Research Agency: ANR-15-IDEX-01, and by the grant AAP Sante 06 2017-260 DGADSH.

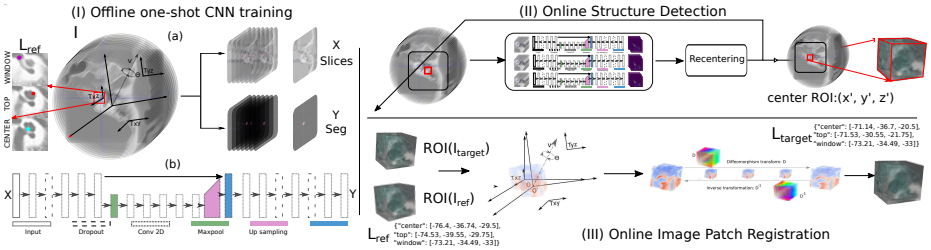


Fig. 1. Overview of the proposed framework.

Learning based landmarks detection Probabilistic graphical models were used for bones landmark labelling in [4] and [5]. Potesil *et al.* [6] use joint spatial priors and parts based graphical models to improve the landmarks detection accuracy of organs. Shouhei *et al.* [7] proposed a Bayesian inference of landmarks through a parametric stochastic landmark detector of the candidates. Donner *et al.* [8] applied random forest and Markov Random Field (MRF) for vertebral body landmarks detection. Mothes *et al.* [9] proposed a one-shot SVM based landmarks tracking method for X-Ray image landmark detection. Suzani *et al.* [10] proposed to train a convolutional neural network (CNN) with an annotated dataset for automatic vertebrae detection and localization. Liang *et al.* [11] proposed a two-step based residual neural network for landmarks detection. Deep reinforcement learning for landmarks detection was investigated by Ghesu *et al.* [12] where landmarks localization is considered as a navigation problem.

The main drawback of the above deep learning based landmarks detection methods is that the creation of manually annotated dataset with 3D landmarks is time consuming and in practice very difficult to collect. To tackle this problem, Zhang *et al.* [13] proposed a deep learning based landmarks detection method that can be used a limited number of annotated medical images. Their framework consists of two CNNs: one for regressing the patches and the second to predict the landmark positions. Yet, this method like the rest of the learning-based methods are not suited when only one annotated image is available. Another source of difficulties is to detect landmarks that are concentrated on a small part of the image. A typical example is the detection of cochlear landmarks from CT images since the human cochlea is a tiny structure.In this paper, we tackle the problem of automatic determination of 3D landmarks in a volumetric image from a single example consisting of a reference image with its landmarks. We propose a one-shot learning approach that first localizes a Structure Of Interest (SOI) (e.g. the cochlea in a CT image of the inner ear) which lies next to the landmarks. A 2D CNN is trained offline by generating arbitrary oriented slices of a reference image with the binary mask of the SOI. Given a target image, the location of the SOI is iteratively estimated by applying the 2D CNN on 3 orthogonal sets of slices. After aligning the orientations of the two SOI on the target and reference images, a non-rigid registration algorithm is applied to propagate the landmarks to the target image. We apply this approach on 200 CT images of the temporal bone to locate 3 cochlear landmarks and show that the positioning error is within

the intra-rater variability. To the best of our knowledge, this is the first one-shot learning method for landmark detection which makes it highly applicable for several clinical problems.

2 Method

2.1 Overview

The proposed approach is described in Fig.1. The algorithm requires as input a reference image I_{ref} where a set of landmarks L_{ref} are positioned. In addition, we require that a binary mask of a visible anatomical or pathological structure $S_{\text{ref}} \subset I_{\text{ref}}$ including the landmarks $L_{\text{ref}} \in S_{\text{ref}}$ be provided. Given a target image I_{target}, landmarks L_{target} are estimated by applying an image registration algorithm between an image patch $P_{\text{ref}} \subset I_{\text{ref}}$ centered on the reference landmarks and an image patch $P_{\text{target}} \subset I_{\text{target}}$ extracted on the target image. The main challenge is to automatically extract the target image patch P_{target} such that it is roughly aligned in position and orientation with the reference image patch in order to ease the non-rigid image registration task. To extract the centered target image patch, we first train a 2D CNN to segment the mask S_{ref} on random slices of the reference image. This stage is performed offline and also requires an additional validation image I_{val} where the same visible structure S_{val} has been segmented. Given a target image, the localization stage extracts the target image patch P_{target} by iteratively applying the segmentation network to find the center of mass of the structure and by aligning its axis of inertia. The last stage applies a registration algorithm to estimate the position of landmarks L_{target}.

2.2 Offline one-shot CNN training

The objective is to train an algorithm that can roughly segment the structure of interest $S_{\text{ref}} \subset I_{\text{ref}}$. That structure must include the landmarks or must lie in the vicinity of the landmarks L_{ref}. It should also be present in all target images and must be easy to detect int the image with some visible borders. One issue of one-shot learning is the limited amount of training data that can easily lead to overfitting [14, 15]. To this end, we chose to train a shallow 2D U-net f_ω segmentation network in order to segment the SOI that surrounds the landmarks. The training set consists of slices of the reference image I_{ref} along arbitrary rotations and translation offsets together with the associated binary masks created by slicing accordingly the reference segmentation S_{ref}. The 2D CNN is trained by minimizing the Binary Cross-Entropy (BCE) loss function. To limit the risk of overfitting, we use a validation set consisting of another volumetric image I_{val} and its segmentation S_{val}. The training is stopped when the segmentation performance of f_ω on the 3 orthogonal slices of I_{val} start to decrease. The details of the training procedure are provided in algorithm 1. The CNN training can be performed offline and the 2D random image slices are centered on the center of mass $\mathbf{C}_{\text{ref}}$ (for $T = 0$) of the segmented structure of

Algorithm 1: One-shot training of CNN

Inputs: image: $I_{\text{ref}}, I_{\text{val}}$, segmentation: $S_{\text{ref}}, S_{\text{ref}}$
Output: CNN parameters ω
Initialize: f_ω, $\Delta T, \Delta R$;
while L_{val} *decreases* **do**
  $T \leftarrow (U(-1,1)\Delta_T)^3$; ; // Uniform Random Translation
  $R \leftarrow (U(-1,1)\Delta_R)^3$; // Uniform Random Rotation
  $I_{trans} \leftarrow \text{Resample}(I_{\text{ref}}, R, T)$; // Transformed Image
  $S_{trans} \leftarrow \text{Resample}(S_{\text{ref}}, R, T)$; // Transformed Segmentation
  for $i = 1;\ i < K;\ i++$ **do**
    $f_\omega \xleftarrow{\omega} I_{trans}[i] | S_{trans}[i]$; // Train the CNN
  end
  $L_{val} \leftarrow loss(S_{\text{val}}, f_{cnn}(I_{\text{val}}))$; // Validation loss
end

interest S_{ref}. Furthermore, the 2D image size of the CNN input is chosen as to cope with the translation ΔT and rotation ΔR offsets such that random slices do not include any missing pixel values.

2.3 Online Structure Detection

Given an input image I_{target}, we seek to locate the structure of interest S_{target} with the proper translation and orientation offsets in order to ease the last image registration stage.

Translation offset estimation To determine the 3D translation offset between I_{target} and I_{ref}, we propose to align the centers of the mass corresponding to the structures of interest S_{target} and S_{ref}. We rely on the trained CNN $f_\omega()$ to determine S_{target} given I_{target}. However, with the limited training set of $f_\omega()$, we need to cope with its possible poor performance. To this end, we propose an iterative method described in algorithm 2 and Fig.2, where the estimation of the translation offset is progressively refined. We write as $f_\omega(I^x_{\text{target}}[k])[i,j]$ the output of the CNN applied on the slice k in the X direction of the volumetric image I_{target} which is a 2D probability map. We apply the CNN on the slices of I_{target} extracted along the X,Y,Z directions. To improve the robustness of the center of mass estimation of I_{target}, we combine their output by multiplying the 3 probabilities outputs for each voxel. The joint output of the network at voxel $[i,j,k]$ is then written as :

$$p[i,j,k] = f_\omega(I^Z_{\text{target}}[k])[i,j] \cdot f_\omega(I^Y_{\text{target}}[j])[k,i] \cdot f_\omega(I^x_{\text{target}}[i])[j,k] \qquad (1)$$

The product of the 3 probability maps favors the pixels where the 3 outputs agree. This helps to filter out the false positive pixels produced by the network that are not correlated on the 3 slice orientations. The center of mass $\mathbf{C}_{\text{target}}$

is then simply estimated as the barycenter of the image voxels weighted by the joint probability $p[i,j,k]$:

$$\mathbf{C}_{\text{target}} = \frac{\left(\sum_{i,j,k} x[i,j,k]*p[i,j,k], \sum_{i,j,k} y[i,j,k]*p[i,j,k], \sum_{i,j,k} z[i,j,k]*p[i,j,k]\right)^T}{\sum_{i,j,k} p[i,j,k]} \quad (2)$$

The target image is then cropped around the detected center $\mathbf{C}_{\text{target}}$ which is written as $\tilde{P}_{\text{target}}$. When the translation offset between the target and reference images is large, the CNN segmentation performances tend to degrade since it has been trained with slices roughly centered on the center of S_{ref}. This is why we propose to iteratively apply the same approach on I_{target} after being centered on $\mathbf{C}_{\text{target}}$. This way, we expect the centered image to be more and more accurately segmented by the neural network since it sees slices that resemble more and more to its training set. We stop the process when the changes in the detected center $\mathbf{C}_{\text{target}}$ become smaller than a threshold.

Algorithm 2: Iterative center of mass localization

Inputs: image: I_{target}, CNN: $f_\omega(\cdot)$
Output: Center of structure in target image $\mathbf{C}_{\text{target}}$
Initialize: ϵ;
$\mathbf{C}_{\text{target}} \leftarrow \mathbf{C}_{\text{ref}}$;
while $|\mathbf{C}_{\text{old}} - \mathbf{C}_{\text{target}}| < \epsilon$ **do**
  $\tilde{P}_{\text{target}} \leftarrow \text{Crop}(I_{\text{target}}, \mathbf{C}_{\text{target}})$; `// Patch centered on` $\mathbf{C}_{\text{target}}$
  while $o \in \{X, Y, Z\}$ **do**
    for $i = 1; i < K[o]; i++$ **do**
      $out[o][i] \leftarrow f_{omega}(\tilde{P}^o_{\text{target}}[i])$; `// apply CNN on slices`
    end
  end
  $p \leftarrow out[X] \cdot out[Y] \cdot out[Z]$; `// Combine probability maps as Eq.1`
  $\mathbf{C}_{\text{old}} \leftarrow \mathbf{C}_{\text{target}}$;
  $\mathbf{C}_{\text{target}} \leftarrow$ Eq. 2; `// Update center of mass`
end
$\tilde{P}_{\text{target}} \leftarrow \text{Crop}(I_{\text{target}}, \mathbf{C}_{\text{target}})$; `// Patch centered on` $\mathbf{C}_{\text{target}}$

Rotation offset estimation After having aligned the center of mass of the two structures of interest, the rotation offset is determined by aligning the moments of inertia of S_{ref} and S_{target}. More precisely, the matrix of inertia captures the ellipsoid appearance of each structure and it determines the structure orientation unambiguously if that structure does not have any axis of symmetry. Therefore the alignment of the matrices of inertia consists in applying a rotation to S_{target} such that the eigenvectors of the 2 matrices coincide [16,17] when they are sorted according to their eigenvalues. The moments of inertia of S_{target} are computed based on the combined probability $p[i,j,k]$ as computed in Eq.1. Thus, after performing the eigenvalue decomposition of the 2 matrices, the rotation matrix

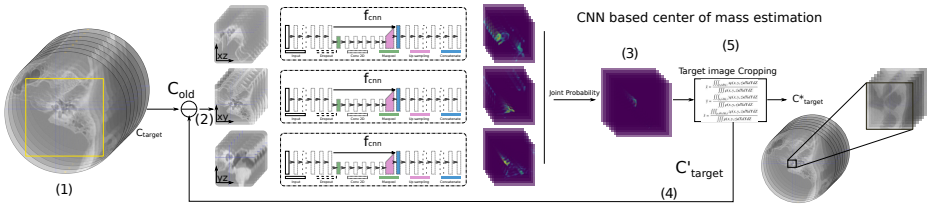


Fig. 2. Iterative determination of the center of mass of the structure of interest. Steps (1) - (2) show the 2D CNN segmentation of the structure of interest from the 3 set of orthogonal slices; (3) The probability maps of the 3 views are combined; (4) Update of the center of mass from the joint probability maps; (5) The target image is cropped around the center of mass.

centered on $\mathbf{C}_{\text{target}}$ is applied on the image patch $\tilde{P}_{\text{target}}$ to get the final target image patch P_{target}.

2.4 Online Image Patch Registration

After the two previous stages, the estimation of the landmarks L_{target} is achieved by performing a non-rigid registration of the reference image patch P_{ref} onto the target image patch P_{target}. The two image patches have the same size, are both centered on the structure of interest and their orientation roughly coincide. This is a good initialization for applying the standard diffeomorphic demons algorithm [18] as implemented in "itk::DiffeomorphicDemonsRegistrationFilter". This algorithm starts with a multi-resolution rigid registration followed by the non-rigid transformation parameterized by a stationary velocity field. It assumes that intensity distribution matches between the two images patches with a sum of square difference as similarity measure. The reference landmarks L_{ref} are then transported to the target image patch P_{target} through the estimated non-rigid transformation. Finally, the landmarks L_{target} on the original target image I_{target} are positioned by inverting the rigid transforms and cropping performed during the first two stages of the method.

3 Experiment

3.1 Dataset

The dataset consists of 200 volumetric CT images of the left temporal bones acquired by a GE LightSpeed CT scanner at the Nice University Center Hospital. The image dimensions are $(512, 512, 160)$ in 3D with corresponding spacing of $(0.25mm, 0.25mm, 0.24mm)$. In this case, the structure of interest is the cochlea, a relatively small bone having a spiral shape similar to a snail shell and without any axis of symmetry. The cochlea is easily visible on CT images. Two volumetric images were randomly selected to serve as reference and validation images and their cochlea was then segmented by an expert in a semi-automatic fashion. Three landmarks corresponding the cochlea top, center and round window were manually set on the reference image as shown in Fig. 1.

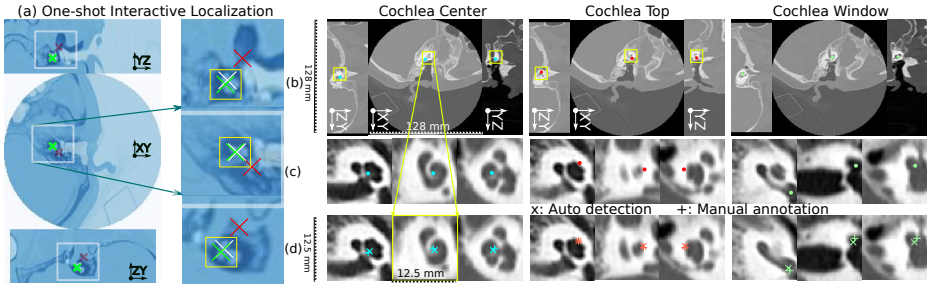


Fig. 3. (a) Positions of the center of mass of the cochlea during 3 iterations of the translation offset determination. The 3 cross marks in red, white, green correspond to the 1st, 2nd, 3rd iterations; Row (b) shows the result of the landmarks detection in the whole image I_{target}; Row (c) zooms on the detected landmarks before applying the last registration stage; Row (d) zooms on the generated landmarks ('x' marks) after the registration stage and the manually positioned landmarks ('+' marks) by an expert.

3.2 Network architecture and training details

We use a 2D U-net like network [19] for segmenting the cochlea in 2D images. The network structure is shown in Fig:1 and is relatively shallow in order to minimize its complexity. The network input size is $[\cdot, 100, 100, 1]$ followed with 4 convolutional layers (shape: $[\cdot, 100, 100, 64]$). Feature maps are convoluted with a group of halved size layers but doubled in channels (shape: $[\cdot, 50, 50, 128]$). Up-sampling layer applied to recover the size of the feature maps to merged with the jump concatenates feature maps (shape:$[\cdot, 100, 100, 64 + 128]$). Finally, 5 convolutional layers (shape:$[\cdot, 100, 100, 64]$, $chn = 64$ for middle layers, $chn = 1$ for the last layer) are used for generating the final feature map. An Adam optimizer is used with a learning rate initialized to $lr = 0.1$ and decreasing with the number of epochs. The neural network was implemented with Tensorflow 2.0 framework and trained on one NVIDIA 1080 Ti GPU. The offline stage of the CNN takes less than 1h for training and the online stages takes around 30s.

4 Results

The proposed approach was evaluated qualitatively and quantitatively. In Fig: 3(a), we show the position of the center of mass of the segmented cochlea $\mathbf{C}_{\text{target}}$ during three iterations of Algorithm 2. We see that the 3 points are getting closer to each other after each iteration thus demonstrating the convergence of the algorithm. In practice, we found that between 2 to 6 iterations are necessary to get a change of mass center position between two iterations less than 1mm.

For a quantitative assessment of performance, an expert positioned twice the 3 landmarks on 20 additional volumes in order to estimate the positioning error and the intra-rater variability. In addition, we also try to employ a naive registration-based landmarks detection method without the iterative localization. The setup of the naive method shares the same registration conditions as the registration steps in the proposed framework.

In Fig: 3(d) we show the 3 landmarks generated by our algorithm with the same landmarks positioned by the expert. Clearly those points are very close to

Table 1. Position errors of the 3 cochlear landmarks (centre, top and window) automatically generated landmarks (AUTO), a second set of manual (MANU) ones, and automatically generated landmarks of registration based naive method (REG).

Image ID	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	μ/σ
CEN AUTO	0.88	0.28	0.49	0.7	0.72	0.57	0.19	0.49	0.49	0.39	0.65	0.84	0.87	0.67	0.72	0.72	0.73	0.54	0.33	0.39	*0.58±0.19mm*
TOP AUTO	0.7	1.33	0.56	0.73	0.72	0.37	0.31	0.78	0.35	0.2	1.63	1.15	1	0.26	1.04	0.67	0.8	1.23	0.55	0.39	0.73±0.39mm
WIN AUTO	0.86	0.65	0.84	0.55	0.65	1.12	1.35	0.31	0.6	0.49	0.26	1.06	1.54	0.72	0.88	0.81	0.54	0.34	1.43	0.88	*0.79±0.36mm*
CEN MANU	0.28	0.56	0.53	1.06	0.65	0.59	0.45	0.57	0.25	1.09	0.94	0.84	1.09	0.53	0.37	0.5	0.25	0.54	0.59	0.3	0.60±0.27mm
TOP MANU	0.43	0.38	0.49	0.25	0.31	0.25	0.31	0.19	0.24	1.09	0	0.5	0.75	0.25	0.31	0.19	0.6	0.42	0.33	0.66	*0.40±0.24mm*
WIN MANU	0.69	0.62	1.11	1.1	0.31	1.07	0.31	0.77	0.43	0.57	0.79	1.22	0.91	0.77	0.97	0.75	0.9	1.01	1.18	1.25	0.84±0.29mm
CEN REG	4.42	10.95	15.78	16.49	12.83	13.04	14.28	15.09	9.66	16.21	6.82	12.91	11.06	6.96	22.69	6.16	2.22	11.68	17.79	9.74	11.84±4.97mm
TOP REG	1.11	8.85	13.73	14.12	10.47	11.25	12.69	12.90	7.55	14.29	4.30	9.56	7.28	4.84	20.27	3.77	0.25	12.88	15.33	13.65	9.95±5.18mm
WIN REG	2.21	2.82	8.62	9.73	4.15	5.67	5.37	7.44	2.51	7.89	1.09	6.42	3.46	1.45	15.33	1.77	4.91	16.20	9.40	16.77	6.66±4.85mm

each other on the 3 views. In Table 1(top), we list the average position error of the 3 landmarks on the 20 images with respect to one set of landmarks manually positioned by an expert, and in the bottom rows, we show the corresponding results obtained by the naive registration based method.

In average, the position error of L_{target} is around 0.6mm which corresponds to a difference of position of 2 to 3 voxels. This result is satisfactory when considering the small size of the cochlea (width: 6.53 $\pm 0.35mm$, height: 3.26 $\pm 0.24mm$ [20]) within the full CT volume ($128mm \times 128mm \times 55mm$). *In contrast, the naive method is almost unusable for cochlea landmarks detection as the relative error (on average* $9.48mm$*) is too large in comparison with the size of the cochlea.* For a better assessment, we also provide the intra-expert landmark position error in Table 1(middle). It shows that the algorithm error is similar to the intra-expert variability, with a lower error for two (the center and window landmarks) out of the three landmarks. When computing the landmark position error with the second set of landmarks made by the expert, or with the average of the 2 annotations, we also found that the algorithm was performing similarly to the expert. Since the intra-rater variability is in most cases lower than inter-rater variability, we believe that the proposed method is an effective way to automate landmark positioning around the cochlea on CT images. Note that the mean landmark position errors reported by Zhang *et al.* [13] also correspond between 2.5 to 3 times the voxel size whereas Grewal *et al.* [21] after training on 168 scans reports errors between 2 to 9 times the voxel size ($2 - 9mm$).

5 Conclusion

To the best of our knowledge, the proposed method is the first one-shot learning approach for 3D landmarks detection in volumetric images. We showed that the proposed approach was effective in localizing 3D landmarks in the cochlea from CT images of the inner ear. It relies on a segmentation stage and the registration of a single user-defined image patch which makes it easy explainable and interpretable. The approach is generic and could be applied to the detection of landmarks in CT imaging and other imaging modalities. In the future, we plan to use more complex image similarity measures in the final registration algorithm and to introduce more annotated data (few-shot learning) to address challenging landmark detection problems. Other network architectures proposed in the literature for one-shot deep learning such as [22–25] can be explored.

References

1. W. Cheung *et al.*. n-sift: n-dimensional scale invariant feature transform. *IEEE Transactions on Image Processing*, pages 2012 – 2021, 10 2009.
2. S. Wörz *et al.*. Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models. *MedIA*, 10(1):41–58, 2006.
3. F. R. José *et al.*. Detection of point landmarks in 3D medical images via phase congruency model. *JBCS*, 17:117–132, 2011.
4. S. Schmidt *et al.*. Spine detection and labeling using a parts-based graphical model. In *IPMI*, pages 122–133, 2007.
5. J. Corso *et al.*. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In *MICCAI*, pages 202–210, 2008.
6. V. Potesil *et al.*. Personalization of pictorial structures for anatomical landmark localization. In *IPMI*, pages 333–345, 2011.
7. H. Shouhei *et al.*. Automatic detection of over 100 anatomical landmarks in medical ct images. *MedIA*, 35:192–214, 2017.
8. R. Donner *et al.*. Global localization of 3D anatomical structures by prefiltered hough forests and discrete optimization. *MedIA*, 17:1304–1314, 2013.
9. O. Mothes *et al.*. One-shot learned priors in augmented active appearance models for anatomical landmark tracking. In *CVICG*, pages 85–104, 2019.
10. A. Suzani *et al.*. Fast automatic vertebrae detection and localization in pathological ct scans. In *MICCAI*, volume 9351, 2015.
11. X. Liang *et al.*. A deep learning framework for prostate localization in cone beam ct-guided radiotherapy. *Medical Physics*, 47(9):4233–4240, 2020.
12. F. Ghesu *et al.*. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in ct scans. *IEEE TPAMI*, 41(1):176–189, Jan 2019.
13. J. Zhang*et al.*. Detecting anatomical landmarks from limited medical imaging data using t2dl. *IEEE TIP*, 26(10):4753–4764, 2017.
14. D. Wu *et al.*. One shot learning gesture recognition from rgbd images. In *2012 IEEE CVPR Workshops*, pages 7–12, 2012.
15. V. Oriol *et al.*. Matching networks for one shot learning. In *NIPS*, pages 3630–3638. 2016.
16. A. Jaklic *et al.*. Moments of superellipsoids and their application to range image registration. *IEEE Transactions on Cybernetics*, 33(4):648–657, 2003.
17. J.J. Crisco *et al.*. Efficient calculation of mass moments of inertia for segmented homogenous 3D objects. *J Biomech*, 31(1):97–101, 1997.
18. T. Vercauteren *et al.*. Non-parametric diffeomorphic image registration with the demons algorithm. In *MICCAI 2007*, pages 319–326, 2007.
19. Ronneberger Olaf *et al.*. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI 2015*, pages 234–241, 2015.
20. Z. Devira *et al.*. Variations in Cochlear Size of Cochlear Implant Candidates. *International Archives of Otorhinolaryngology*, 23:184–190, 06 2019.
21. M. Grewal *et al.*. An end-to-end deep learning approach for landmark detection and matching in medical images. *PBOI*, 11313:1131–1328, 2020.
22. K. Gregory *et al.*. Siamese neural networks for one-shot image recognition. *ICML Deep Learning Workshop*, 2015.
23. S. Amirreza *et al.*. One-shot learning for semantic segmentation. 2017.
24. Z. Chen *et al.*. Image deformation meta-networks for one-shot learning. In *IEEE CVPR*, June 2019.
25. J. Shruti *et al.*. Improving siamese networks for one shot learning using kernel based activation functions. *ArXiv*, abs/1910.09798, 2019.