Keypoint Transfer Segmentation

C. Wachinger[1,2(✉)], M. Toews[3], G. Langs[1,4], W. Wells[1,3], and P. Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT, Cambridge, USA
wachinger@csail.mit.edu
[2] MGH, Harvard Medical School, Boston, USA
[3] BWH, Harvard Medical School, Boston, USA
[4] CIR, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria

Abstract. We present an image segmentation method that transfers label maps of entire organs from the training images to the novel image to be segmented. The transfer is based on sparse correspondences between keypoints that represent automatically identified distinctive image locations. Our segmentation algorithm consists of three steps: (i) keypoint matching, (ii) voting-based keypoint labeling, and (iii) keypoint-based probabilistic transfer of organ label maps. We introduce generative models for the inference of keypoint labels and for image segmentation, where keypoint matches are treated as a latent random variable and are marginalized out as part of the algorithm. We report segmentation results for abdominal organs in whole-body CT and in contrast-enhanced CT images. The accuracy of our method compares favorably to common multi-atlas segmentation while offering a speed-up of about three orders of magnitude. Furthermore, keypoint transfer requires no training phase or registration to an atlas. The algorithm's robustness enables the segmentation of scans with highly variable field-of-view.

1 Introduction

Is atlas-based segmentation without dense correspondences possible? Typical registration- and patch-based segmentation methods [3,7,15,16] compute correspondences for each location in the novel image to be segmented to the training images. These correspondences are either obtained from dense deformation fields or from the retrieval of similar patches. For scans with a large field-of-view, such approaches become computationally intense. We propose a segmentation method based on distinctive locations in the image - *keypoints*. In contrast to manually selected landmarks [14], keypoints are automatically extracted as local optima of a saliency function [12]. Matches between keypoints in test and training images provide correspondences for a sparse set of image locations, which we use to transfer entire organ segmentations. Working with sparse correspondences and transferring whole organ maps makes our method computationally efficient. The probabilistic fusion of organ maps yields a segmentation accuracy comparable to that of state-of-the-art methods, while offering orders of magnitude of speed-up.

S. Ourselin et al. (Eds.): IPMI 2015, LNCS 9123, pp. 233–245, 2015.
DOI: 10.1007/978-3-319-19992-4_18

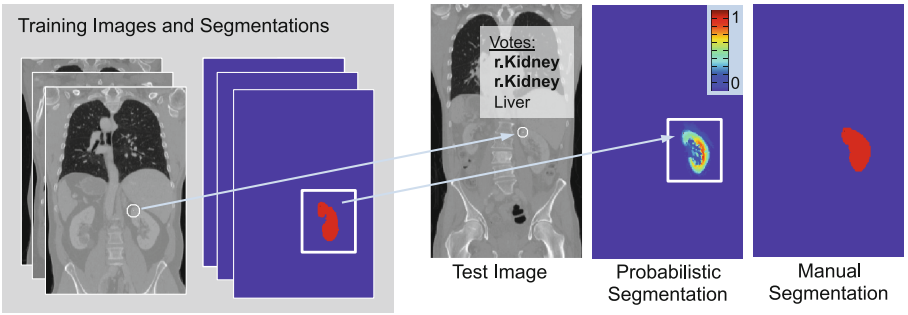


Fig. 1. Illustration of keypoint transfer segmentation. First, keypoints (white circles) in training and test images are matched (arrow). Second, voting assigns an organ label to the test keypoint (r.Kidney). Third, matches from the training images with r.Kidney as labels are transferred to the test image, creating a probabilistic segmentation. We show the manual segmentation for comparison.

Keypoint matching offers the additional advantage of robustness in establishing correspondences between images with varying field-of-view. This property is important when using manually annotated whole-body scans to segment clinical scans with a limited field-of-view. In clinical practice, the diagnostic focus is commonly on a specific anatomical region. To minimize radiation dose to the patient and scanning time, only the region of interest is scanned. The alignment of scans with a limited field-of-view to full abdominal scans is challenging with intensity-based registration, especially when the initial transformation does not roughly align anatomical structures. The efficient and robust segmentation through keypoint transfer offers a practical tool to handle the growing number of clinical scans.

Figure 1 illustrates the keypoint transfer segmentation. Keypoints are identified at salient image regions invariant to scale. Each keypoint is characterized by its geometry and a descriptor based on a local gradient histogram. After keypoint extraction, we obtain the segmentation in three steps. First, keypoints in the test image are matched to keypoints in the training images based on the geometry and the descriptor. Second, reliable matches vote on the organ label of the keypoint in the test image. In the example, two matches vote for right kidney and one for liver, resulting in a majority vote for right kidney. Third, we transfer the segmentation mask for the entire organ for each match that is consistent with the majority label vote; this potentially transfers the organ map from one training image multiple times if more than one match is identified for this training image. The algorithm also considers the confidence of the match in the keypoint label voting. Keypoint transfer does not require a training stage. Its ability to approximate the organ shape improves as the number of manually labeled images grows.

1.1 Related Work

Several methods have been previously demonstrated for segmenting large field-of-view scans. Entangled decision forests [13] and a combination of discriminative

and generative models [8] have been proposed for the segmentation of CT scans. A combination of local and global context for simultaneous segmentation of multiple organs has been explored [11]. Organ detection based on marginal space learning was proposed in [20]. The application of regression forests for efficient anatomy detection and localization was described in [4]. In contrast to previously demonstrated methods, our algorithm does not require extensive training on a set of manually labeled images.

We evaluate our method on the publicly available Visceral dataset [10,19]. Multi-atlas segmentation on the Visceral data was proposed in [6,9], which we use as a baseline method in our experiments. Our work builds on the identification of keypoints, defined as a 3D extension [18] of the popular scale invariant feature transform (SIFT) [12]. In addition to image alignment, 3D SIFT features were also applied to study questions related to neuroimaging [17]. In contrast to previous uses of the 3D SIFT descriptor, we use it to transfer information across images.

2 Method

In atlas-based segmentation, the training set includes images $\mathcal{I} = \{I_1, \ldots, I_n\}$ and corresponding segmentations $\mathcal{S} = \{S_1, \ldots, S_n\}$, where $S_i(x) \in \{1, \ldots, \eta\}$ for η labels. The objective is to infer segmentation S for test image I. Instead of aligning training images to the test image with deformable registration, we automatically extract anatomical features from the images and use them to establish sparse correspondences. We identify keypoints that locally maximize a saliency function. In the case of SIFT, it is the difference-of-Gaussians [12]

$$\{(x_i, \sigma_i)\} = \text{local arg}\max_{x,\sigma} |f(x, \kappa\sigma) - f(x, \sigma)|, \tag{1}$$

where x_i and σ_i are the location and scale of keypoint i, $f(\cdot, \sigma)$ is the convolution of the image I with a Gaussian kernel of variance σ^2, and κ is a multiplicative scale sampling rate. The identified local extrema in scale-space correspond to distinctive spherical image regions. We characterize the keypoint by a descriptor F^D computed in a local neighborhood whose size depends on the scale of the keypoint. We work with a 3D extension of the image gradient orientation histogram [18] with 8 orientation and 8 spatial bins. This description is scale and rotation invariant and further robust to small deformations. Constructing the descriptors from image gradients instead of intensity values facilitates comparisons across subjects.

We combine the 64-dimensional histogram F^D with the location $F^x \in \mathbb{R}^3$ and scale $F^\sigma \in \mathbb{R}$ to create a compact 68-dimensional representation F for each salient image region. We let F_I denote the set of keypoints extracted from the test image I and $\mathcal{F}_\mathcal{I} = \{\mathcal{F}_{I_1}, \ldots, \mathcal{F}_{I_n}\}$ denote the set of keypoints extracted from the training images $\mathcal{I}$. We assign a label to each keypoint in $\mathcal{F}_{I_i}$ according to the organ that contains it, $\mathcal{L} = S_i(\mathcal{F}^x)$ for $\mathcal{F} \in \mathcal{F}_{I_i}$. We only keep keypoints within the segmented organs and discard those in the background. The organ label L is unknown for the keypoints in the test image and is inferred with a voting algorithm as described later in this section.

2.1 Keypoint Matching

The first step in the keypoint-based segmentation is to match each keypoint in the test image with keypoints in the training images. Some of these initial matches might be incorrect. We employ a two-stage matching procedure with additional constraints to improve the reliability of the matches. First, we compute a match $\mathcal{M}(F)_i$ for a test keypoint $F \in F_I$ to keypoints in a training image $\mathcal{F}_{I_i}$ by identifying the nearest neighbor based on the descriptor and scale constraints

$$\mathcal{M}(F)_i = \underset{\mathcal{F} \in \mathcal{F}_{I_i}}{\arg\min} \|F^D - \mathcal{F}^D\|, \quad \text{s.t.} \;\; \varepsilon_\sigma^{-1} \leq \frac{F^\sigma}{\mathcal{F}^\sigma} \leq \varepsilon_\sigma, \tag{2}$$

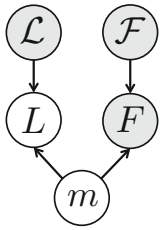
where we set a loose threshold on the scale allowing for variations up to a factor of $\varepsilon_\sigma = 2$. We use the distance ratio test to discard keypoint matches that are not reliable [12]. The distance ratio is computed between the descriptors of the closest and second-closest neighbor. We reject all matches with a distance ratio of greater than 0.9.

To further improve the matches, we impose loose spatial constraints on the matches, which requires a rough alignment. For our dataset, accounting for translation was sufficient at this stage; alternatively a keypoint-based pre-alignment could be performed [18]. We estimate the mode of the translations t_i proposed by the matches $\mathcal{M}_i$ from training image I_i with the Hough transform [2]. Mapping the training keypoints with t_i yields a rough alignment of the keypoints and enables an updated set of matches with an additional spatial constraint

$$\mathcal{M}(F)_i = \underset{\mathcal{F} \in \mathcal{F}_{I_i}}{\arg\min} \|F^D - \mathcal{F}^D\|, \quad \text{s.t.} \;\; \varepsilon_\sigma^{-1} \leq \frac{F^\sigma}{\mathcal{F}^\sigma} \leq \varepsilon_\sigma, \;\; \|F^x - \mathcal{F}^{x+t_i}\|_2 < \varepsilon_x,$$

where we set the spatial threshold ε_x to keep 10 % of the closest matches. As before, we discard matches that do not fulfill the distance ratio test.

We define a distribution $p(m)$ over matches, where a match m associates keypoints in the test image I and training images I_i. We use kernel density estimation on translations proposed by all matches $\mathcal{M}_i$ between keypoints in the test image and those in the i-th training image. For a match $m \in \mathcal{M}_i$, the probability $p(m)$ expresses the translational consistency of the match m with respect to all other matches in $\mathcal{M}_i$. This non-parametric representation accepts multi-modal distributions, where the keypoints in the upper abdomen may suggest a different transformation than those in the lower abdomen.



2.2 Keypoint Voting

After establishing matches for keypoints in the test image, we estimate an organ label L for each keypoint in the test image based on the generative model illustrated above. The latent variable m represents the keypoint matches found in the

previous step. Keypoint labeling is helpful to obtain a coarse representation of the image, including rough location of organs. Additionally, we use the keypoint labels to guide the image segmentation as described in the next section. For inference of keypoint labels, we marginalize over the latent random variable m and use the factorization from the graphical model to obtain

$$p(L, F, \mathcal{L}, \mathcal{F}) = \sum_{m \in \mathcal{M}(F)} p(L, F, \mathcal{L}, \mathcal{F}, m) \tag{3}$$

$$= \sum_{m \in \mathcal{M}(F)} p(L|\mathcal{L}, m) \cdot p(F|\mathcal{F}, m) \cdot p(m), \tag{4}$$

where $\mathcal{M}(F)$ contains matches for keypoint F across all training images. The marginalization is computationally efficient, since we only compute and evaluate a sparse set of matches. We define the label likelihood

$$p(L = l|\mathcal{L}, m) = \begin{cases} 1 & \text{if } \mathcal{L}_{m(F)} = l, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathcal{L}_{m(F)}$ is the label of a training keypoint that the match m assigns to the test keypoint F. The keypoint likelihood is based on the descriptor of the keypoint

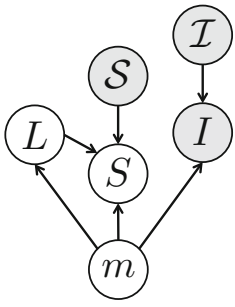
$$p(F|\mathcal{F}, m) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{\|F^D - \mathcal{F}^D_{m(F)}\|_2^2}{2\tau^2}\right), \tag{6}$$

where we set $\tau^2 = \max_m \|F^D - \mathcal{F}^D_{m(F)}\|_2^2$. We assign the most likely organ label to the keypoint

$$\hat{L} = \arg\max_{l \in \{1,\ldots,\eta\}} p(L = l|F, \mathcal{L}, \mathcal{F}) = \arg\max_{l \in \{1,\ldots,\eta\}} p(L = l, F, \mathcal{L}, \mathcal{F}). \tag{7}$$

2.3 Keypoint Segmentation

Here, we introduce a generative model for image segmentation based on keypoint matches and keypoint voting. The latent image segmentation S depends on the keypoint label L and the training segmentations $\mathcal{S}$. A further dependency exists between the test image I and the training images $\mathcal{I}$. All relations between test and training images or keypoints depend on the matches, which bring them into correspondence. We let $\mathcal{I}_m$ denote the training image identified with match m after the transformation implied by the match has been applied. $\mathcal{S}_m$ is similarly defined to be the selected and transformed segmentation map. We infer the segmentation S by marginalizing over the latent random variables and using the factorization from the graphical model



$$p(S, I, \mathcal{S}, \mathcal{I}, \mathcal{L}) = \sum_{m \in \mathcal{M}} \sum_{L} p(S, I, \mathcal{S}, \mathcal{I}, \mathcal{L}, L, m) \tag{8}$$

$$= \sum_{m \in \mathcal{M}} \sum_{L} p(S|L, \mathcal{S}, m) \cdot p(I|\mathcal{I}, m) \cdot p(L|m) \cdot p(m). \tag{9}$$

The likelihood of image segmentation causes keypoints to transfer entire organ label maps

$$p(S|L, \mathcal{S}, m) \propto \begin{cases} 1 & \text{if } S^L = \mathcal{S}_m^L, \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where S^L and $\mathcal{S}_m^L$ are the regions with label L in the test and training segmentations, respectively. This likelihood further restricts keypoints to only transfer segmentations with the same label. We also investigate the transfer of organ segmentations that are different from the keypoint labels in our experimental evaluation.

For the label likelihood we consider $p(L|m) \propto p(L) \cdot \delta(\mathcal{L}_m, \hat{L})$. The Kronecker delta δ only allows training keypoints to transfer their votes that are consistent with the majority vote in Eq. (7). This improves the robustness of the method because even if single matches propose to assign the wrong label to the test keypoint, such matches are discarded for the segmentation, as long as they do not reach the majority. The probability $p(L)$ models the certainty of the label voting for the keypoint in Eq. (7).

The image likelihood assumes conditional independence of the locations x on the image grid Ω and models the local similarity between test and training image

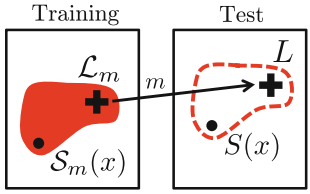
$$p(I(x)|\mathcal{I}, m) = \frac{1}{\sqrt{2\pi}\nu} \exp\left(-\frac{(I(x) - \mathcal{I}_m(x))^2}{2\nu^2}\right), \tag{11}$$

where ν^2 is the intensity noise variance. We obtain the final segmentation $\hat{S}(x)$ by selecting the most likely label

$$\hat{S}(x) = \underset{l \in \{1,\ldots,\eta\}}{\arg\max}\, p(S(x) = l|I(x), \mathcal{S}, \mathcal{I}, \mathcal{L}) = \underset{l \in \{1,\ldots,\eta\}}{\arg\max}\, p(S(x) = l, I(x), \mathcal{S}, \mathcal{I}, \mathcal{L}).$$

We account for not transferring the background surrounding the organ by assigning $\hat{S}(x)$ to the background label if the maximal probability in the voting is below 15 %.

We illustrate the mechanism for computing the segmentation likelihood $p(S(x) = \text{liver})$ on an example of liver. We sum across all matches to all the training images. Only matches that involve training keypoints with the label of liver are considered, identified by $\delta(\mathcal{L}_m, \hat{L})$. Further, the label of liver must be assigned to the test keypoint L.



If the match satisfies these requirements, the entire liver label map is transferred with the transformation proposed by the match m; this step is modeled by $P(S|L, \mathcal{S}, m)$. The transfer affects the segmentation likelihood $p(S(x))$ only if location x is within the spatial extent of the transferred liver label map. To increase the robustness and accuracy of the segmentation, we weigh the transferred segmentation according to the certainty in the keypoint label voting $p(L)$, in the match $p(m)$, and in the local intensity similarity of the test and training image $p(I(x)|\mathcal{I}, m)$.

We also investigate the potential improvement of the segmentation by accounting for affine organ variations across subjects. If there are at least three matches for an organ between one training image and the test image, we estimate an organ-specific affine transformation. We apply the random sample consensus (RANSAC) algorithm [5] to determine the transformation parameters with the highest number of inliers. In our experimental evaluation, the organ-wide affine transformation did not achieve a robust improvement of segmentation accuracy and is therefore not reported in the results. The affine transformation may not improve results because we transfer organ labels multiple times per scan for different translations, which already accounts for organ variability in combination with the weighted voting across subjects.

Table 1. Keypoint voting statistics per organ for ceCT (top) and wbCT (bottom): the average number of keypoints per organ, the average fraction of keypoints that get labeled, and the average fraction of correct keypoints labels. Keypoints are not assigned labels if there exists no reliable match. We omit background keypoints from the training images. Only about one third of the background keypoints are labeled.

Organs	Liver	Spleen	Aorta	Trachea	R.Lung	l.Lung	r.Kid	l.Kid	r.PM	l.PM	Bckgrnd
# Keypts	13.6	4.0	7.6	3.0	29.7	24.7	12.1	12.2	2.5	3.0	526.0
% Labeled	0.73	0.89	0.98	1.00	0.95	0.92	0.98	0.99	0.94	0.92	0.33
% Correct	0.87	0.91	0.97	0.99	1.00	1.00	0.98	1.00	0.99	0.93	0.00
# Keypts	6.0	2.6	5.6	4.4	28.2	24.0	6.7	9.0	2.5	2.5	637.2
% Labeled	0.93	0.98	1.00	1.00	0.98	0.98	0.98	0.99	0.98	1.00	0.35
% Correct	0.82	0.87	0.92	1.00	0.99	0.99	0.98	0.96	1.00	0.93	0.00

3 Results

We perform experiments on 20 contrast-enhanced CT (ceCT) scans and on 20 whole-body CT (wbCT) scans from the Visceral dataset re-sampled to 2 mm isotropic voxels [10]. We segment 10 anatomical structures (liver, spleen, aorta, trachea, left/right lung, left/right kidney, left/right psoas major muscle (PM)). Image dimensions are roughly $217 \times 217 \times 695$ for wbCT and $200 \times 200 \times 349$ for ceCT. We set $\nu = 300$ for lungs and trachea and $\nu = 50$ for all other organs. We

perform leave-one-out experiments by using one image as test and the remaining 19 images as training images. We compare our method to multi-atlas segmentation with majority voting (MV) [7,15] and locally-weighted label fusion (LW) [16] using ANTS [1] for deformable registration. We quantify the segmentation accuracy with the Dice volume overlap between manual and automatic segmentation.

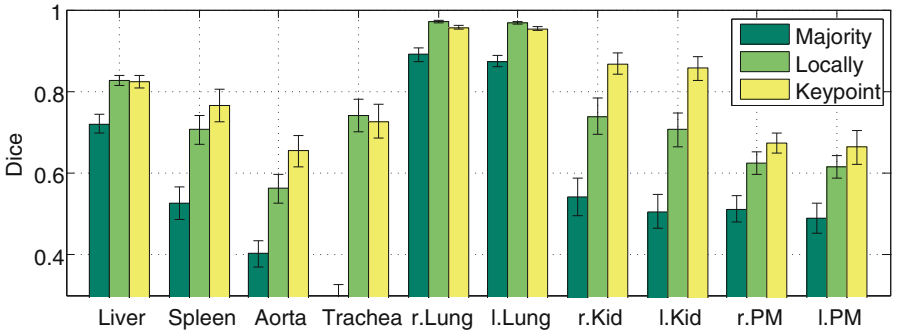


Fig. 2. Segmentation accuracy for ten organs on ceCT images for majority voting, locally-weighted voting, and keypoint transfer. Bars indicate the mean Dice and error bars correspond to standard error.

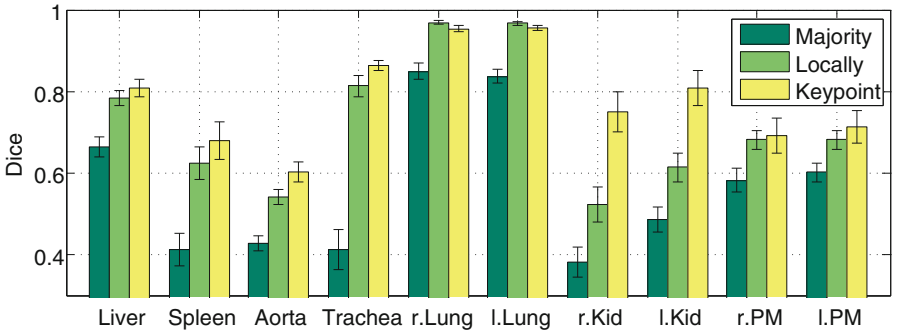


Fig. 3. Segmentation accuracy for ten organs on wbCT images for majority voting, locally-weighted voting, and keypoint transfer. Bars indicate the mean Dice and error bars correspond to standard error.

Table 1 reports statistics for the voting on keypoint labels. The average number of keypoints varies across organs. Keypoints are not labeled if they do not receive reliable matches that pass the spatial constraint and the distance ratio test. Focusing on reliable keypoints improves the performance of the algorithm because it is possible that certain keypoints in the test image do not appear in the training set. For the keypoints that are labeled, the voting accuracy is high. All of the votes on background keypoints in the test image are incorrect, since

we do not include background keypoints in the training set. However, only about one third of the background keypoints receives labels. The remaining background keypoints have limited impact on the segmentation as long as there is no bias in transferring organ maps to a specific location.

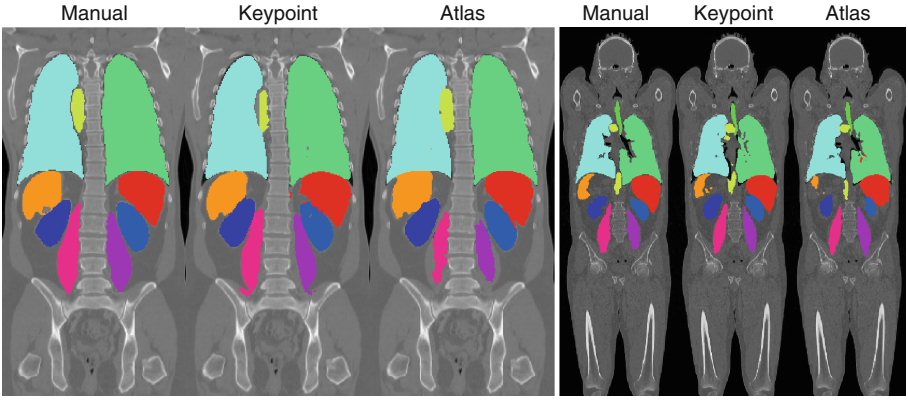


Fig. 4. Coronal views of example segmentation results for ceCT (left) and wbCT (right) overlaid on the intensity images. Each series reports segmentations in the following order: manual, keypoint transfer, locally-weighted multi-atlas.

Figures 2 and 3 report segmentation results for ceCT and wbCT scans, respectively, comparing keypoint transfer to multi-atlas segmentation. Locally-weighted voting outperforms majority voting for all anatomical structures. Keypoint transfer segmentation yields segmentation accuracy comparable to that of locally-weighted voting for most structures and better accuracy for the segmentation of kidneys; the increase in Dice for kidneys is about 0.15 on ceCT and about 0.2 on wbCT. In these experiments, the transfer of segmentations that are different from the keypoint label did not achieve a robust improvement and are therefore not reported. Figure 4 illustrates segmentation results for ceCT and wbCT.

Figure 5 reports the average segmentation result for ceCT scans when varying the number of training scans from 5 to 15; the evaluation is on the five images not included in the training set. The segmentation accuracy generally increases with the number of training scans. This result suggests that averaging over segmentations of a larger number of subjects helps in recovering the true shape of the organ. The availability of larger datasets in the future may therefore further improve the segmentation results. An atlas selection scheme that only transfers organs from overall similar subjects may be helpful, which could be efficiently implemented based on keypoints.

Figure 6 reports the runtime of keypoint transfer segmentation and multi-atlas label fusion. The segmentation with keypoint transfer is about three orders of magnitude faster. On ceCT scans, the extraction of keypoints takes about

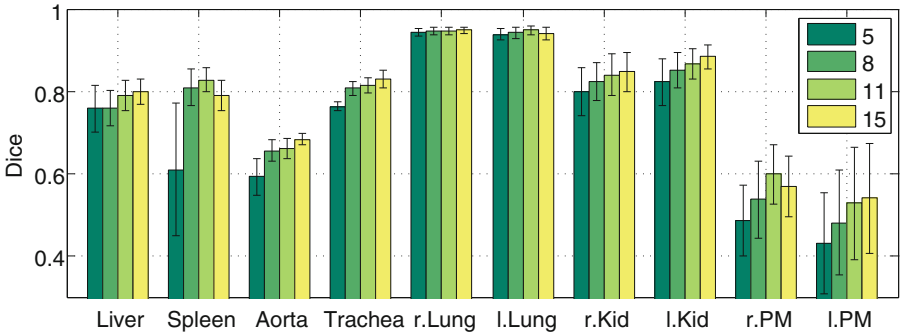


Fig. 5. Segmentation accuracy for ten organs on ceCT images with keypoint transfer with the number of training images ranging from 5 to 15. Bars indicate the mean Dice over five test images and error bars correspond to standard error.

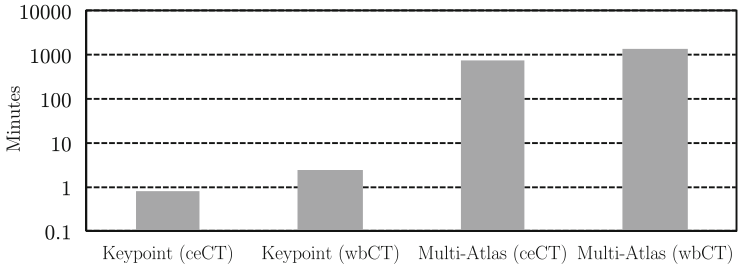


Fig. 6. Average runtimes (in minutes) of the segmentation of ten organs in one image with keypoint transfer and multi-atlas label fusion for ceCT and wbCT. The time is displayed on the logarithmic scale.

30 s and the segmentation transfer takes 16 s, yielding a segmentation time for ten organs that is below one minute. The segmentation transfer is implemented in Matlab without parallelization. For multi-atlas segmentation, the pairwise deformable registration consumes most of the runtime. We also experimented with creating a probabilistic atlas, which reduces computational costs. However, the iterative estimation of the atlas is also expensive and the high anatomical variability of the abdomen makes the summarization challenging.

In addition to the segmentation of abdominal and whole-body scans, we also evaluated the segmentation of scans with limited field-of-view. In clinical practice, such partial scans frequently occur because of a specific diagnostic focus. To test the performance of the algorithm, we crop ceCT and wbCT images around the kidneys and the spleen, as shown in Fig. 7. For spleen images, we found a substantial improvement by transferring organ segmentations that are different from the keypoint label. Figure 7 reports results for segmenting the spleen byonly using spleen keypoints and by also using lung and liver keypoints. In the

partial scans, we notice a slight decrease in segmentation accuracy, compared to working on the full scans. However, the keypoint transfer is overall robust to variations in the field-of-view and enables segmentation without modifications of the algorithm. We do not report results for the multi-atlas segmentation in this experiment because the registration between the cropped images and the training images failed. Since the initial alignment does not lead to a rough overlap of the target regions, it is a very challenging registration problem. While it may be possible to develop initialization techniques that improve the alignment, we consider it a major advantage of the keypoint transfer that no modification is required to handle limited field-of-view scans.

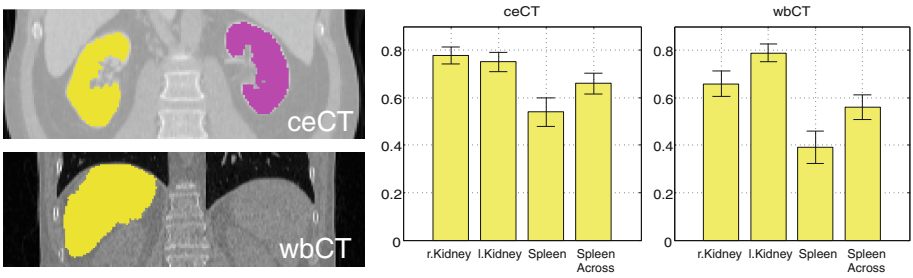


Fig. 7. Coronal views of scans with limited field-of-view showing the kidneys or the spleen, illustrated for ceCT and wbCT, respectively. Bars indicate the mean Dice and error bars correspond to standard error. 'Spleen Across' corresponds to using lung and liver keypoints to transfer spleen segmentations.

4 Conclusion

We introduced an image segmentation method based on keypoints that transfers label maps of entire organs. Relying on sparse correspondences between keypoints in the test and training images increases the efficiency of the method. Keypoint matches are further robust to variations in the field-of-view of the images, which enables segmentation of partial scans. Our algorithms for the keypoint voting and the segmentation transfer were derived from generative models, where latent random variables were marginalized out. The accuracy of our segmentation compares favorably to multi-atlas segmentation, while requiring about three orders of magnitude less computation time.

Acknowledgements. This work was supported in part by the Humboldt foundation, the National Alliance for Medical Image Computing (U54-EB005149), the NeuroImaging Analysis Center (P41-EB015902), the National Center for Image Guided Therapy (P41-EB015898), and the Wistron Corporation.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
2. Ballard, D.: Generalizing the hough transform to detect arbitrary shapes. Pattern Recogn. **13**(2), 111–122 (1981)
3. Coup, P., Manjn, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**(2), 940–954 (2011)
4. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Anal. **17**(8), 1293–1303 (2013)
5. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
6. Goksel, O., Gass, T., Szekely, G.: Segmentation and landmark localization based on multiple atlases. In: Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings, pp. 37–43, Beijing, China (2014)
7. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage **33**(1), 115–126 (2006)
8. Iglesias, J.E., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A.: Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 25–36. Springer, Heidelberg (2011)
9. Jiménezdel Toro, O., Müller, H.: Hierarchical multi-structure segmentation guided by anatomical correlations. In: Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings, pp. 32–36, Beijing, China (2014)
10. Langs, G., Hanbury, A., Menze, B., Müller, H.: VISCERAL: towards large data in medical imaging — challenges and directions. In: Greenspan, H., Müller, H., Syeda-Mahmood, T. (eds.) MCBR-CDS 2012. LNCS, vol. 7723, pp. 92–98. Springer, Heidelberg (2013)
11. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid Multi-organ Segmentation Using Context Integration and Discriminative Models. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 450–462. Springer, Heidelberg (2013)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
13. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of CT images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
14. Potesil, V., Kadir, T., Brady, S.: Learning new parts for landmark localization in whole-body CT scans. IEEE Trans. Med. Imaging **33**(4), 836–848 (2014)
15. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C., et al.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage **21**(4), 1428–1442 (2004)
16. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging **29**, 1714–1729 (2010)

17. Toews, M., Wells III, W., Collins, D.L., Arbel, T.: Feature-based morphometry: discovering group-related anatomical patterns. NeuroImage **49**(3), 2318–2327 (2010)
18. Toews, M., Wells III, W.M.: Efficient and robust model-to-image alignment using 3D scale-invariant features. Med. Image Anal. **17**(3), 271–282 (2013)
19. Jiménez del Toro, O., et al.: VISCERAL - VISual Concept Extraction challenge in RAdioLogy. In: Goksel, O. (ed.) Proceedings of the VISCERAL Challenge at ISBI, No. 1194 in CEUR Workshop Proceedings, pp. 6–15 (2014)
20. Zheng, Y., Georgescu, B., Comaniciu, D.: Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 411–422. Springer, Heidelberg (2009)