Multi-Atlas Segmentation using Patch-Based Joint Label Fusion with Non-Negative Least Squares Regression

Mattias P. Heinrich, Matthias Wilms, and Heinz Handels

Institute of Medical Informatics, University of Lübeck, Germany
heinrich@imi.uni-luebeck.de, www.mpheinrich.de

Abstract. This work presents a patch-based multi-atlas segmentation approach based on non-negative least squares regression. Our approach finds a weighted linear combination of local image patches that best models the target patch, jointly for all considered atlases. The local coefficients are optimised with the constraint of being positive or zero and serve as weights, of the underlying segmentation patches, for a multi-atlas voting. The negative influence of erroneous local registration outcome is shown to be reduced by avoiding negative weights. For challenging abdominal MRI, the segmentation accuracy is significantly improved compared to standard joint least squares regression and independent similarity-based weighting. Our experiments show that restricting weights to be non-negative yields significantly better segmentation results than a sparsity promoting ℓ_1 penalty. We present an efficient numerical implementation that rapidly calculates correlation matrices for all overlapping image patches and atlases in few seconds.

Keywords: linear regression, generative model, cross-correlation

1 Introduction

Automatic segmentation of anatomical structures is an essential part of medical imaging applications, including but not limited to visualisation, computer-aided diagnosis and image-guided interventions. Atlas-based segmentation, in particular when employing multiple atlases and deformable registration, has demonstrated to achieve high accuracy and very good generalisation to different modalities, application domains and anatomic sights. Several algorithms have been proposed to rate different atlases [1] based on their suitability to segment a given target image in order to optimally fuse them. Patch-based approaches have the advantage against global methods of being able to estimate locally varying weights and can therefore better deal with local registration errors or imaging artefacts [2]. Recently, generative models that jointly estimate atlas weights have become popular [3, 4]. In Sec. 3, we present an elegant mathematical formulation that jointly estimates a linear combination of all atlas image patches that best approximates the target patch using **only positive or zero coefficients**. By employing non-negative least squares [5], we can avoid unrealistic

negative weights of atlases, which may cancel out correct label votes. Furthermore, we present a numerically efficient solution that enables rapid calculation of local correlation matrices and makes our fusion approach at least an order of magnitude faster than previous joint fusion methods [4]. Together with a fast deformable registration our method enables computation times of < 2 minutes per scan, which closes the performance gap to machine learning approaches. In Sec. 4, we demonstrate the suitability of our approach for the segmentation of organs in clinical abdominal magnetic resonance images (MRI) that pose different challenges than more commonly studied brain images. We show experimental comparisons with majority voting, locally weighted fusion (based on normalised correlation), sparse Lasso regression [6], least squares regression [3] and joint label fusion [4].

2 Related Work

Multi-atlas label fusion is an active field of research in medical imaging. Global performance of multiple raters or atlases can been estimated by the popular STAPLE (Simultaneous truth and performance level estimation) framework [1]. Recent work has shown that segmentation and registration errors are highly non-uniform across the image domain so that locally varying performance levels should be estimated [2] for every overlapping patch within the target scan. STAPLE estimates dependency across atlases based on their agreement of segmentation labels, however, atlas-based registration provides additional intensity information (of registered atlas images), which has been exploited e.g. in [7, 8]. In these works, label fusion is formulated as a generative model, which explains the formation of the target appearance through multiple atlases. Non-local patch based approaches that directly obtain weights from distance measures of image similarity [9, 10] are popular when only a rough (e.g. affine) alignment of atlases to the target scan is available. These methods, with the exception of [11], usually rate each atlas independently and may miss relations between them. Joint label fusion with least square regression was proposed in [3] and similarity-based distances between atlases and target scan and amongst atlases themselves in [4]. An additional sparsity constraint together with least squares regression was used in [6]. Yet, the task to select the right number of atlases is difficult in this context as discussed in [12]. Our work is closely related to the joint least square approaches of [3] and [6], but demonstrates the potential degradation of the label fusion through negative weights and proposes an elegant and effective solution by using a non-negativity constraint in the regression.

3 Methods

Our aim is to segment an unseen 3D intensity target scan $\mathbf{I}$ using n atlas scans $\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_n$ and their corresponding manual segmentations $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$,

which have already been brought into spatial alignment by deformable registration. To find the optimal label for a location x in the target scan, we consider its intensity patch $\mathbf{X}(x)$ (in $\mathbf{I}$) for a neighbourhood radius r defined by Ω_x with a size of $R = (2r+1)^3$. From the atlas database, we have n patches $\mathbf{Y}_1(x), \mathbf{Y}_2(x), \ldots, \mathbf{Y}_n(x)$. To avoid the negative influence of local contrast variations we first subtract the mean of all patches and divide them by their standard deviations, yielding normalised patches $\mathbf{X}'(x)$, $\mathbf{Y}'_1(x)$, etc:

$$\mathbf{X}'(x) = \frac{\mathbf{X}(x) - \overline{\mathbf{X}(x)}}{\sqrt{\frac{1}{R}\sum_{y\in\Omega_x}((\mathbf{X}(y) - \overline{\mathbf{X}(x)})^2}} \text{ with } \overline{\mathbf{X}(x)} = \frac{1}{R}\sum_{y\in\Omega_x}\mathbf{X}(y) \tag{1}$$

We now seek the coefficient vector $\boldsymbol{c} = c_1, c_2, \ldots, c_n$ that minimises the least squares error ρ^2 of a linear weighted combination of atlas patches:

$$\rho^2 = \sum_{y\in\Omega_x}(\mathbf{X}'(y) - \sum_{i}^{n} c_i \mathbf{Y}'_i(y))^2 \tag{2}$$

Let $\mathbf{A}$ be a $R \times n$ matrix, where the *i-th* column represents the atlas patch $\mathbf{Y}'_i(x)$ and $\boldsymbol{b}$ a vector of length R, which stands for the target patch $\mathbf{X}'(x)$. The straightforward solution of the optimal $\boldsymbol{c}$ in Eq. 2 without further constraints can be obtained by least squares regression. In our method, we propose to add the constraint that all coefficients have to be non-negative, hence solving a non-negative least squares problem [5]:

$$\underset{\boldsymbol{c}}{\operatorname{argmin}}||\mathbf{A}\boldsymbol{c} - \boldsymbol{b}||_2^2 \text{ subject to } c_i \geq 0 \;\forall\; i = \{1, 2, \ldots, n\} \tag{3}$$

The estimated $\boldsymbol{c}$ can be used to obtain an optimal reconstruction of the target patch from a weighted combination of all atlas patches for which $c_i > 0$. Fig. 1 shows an example, where the advantages of the non-negativity constraint can be clearly seen in comparison to the classical least squares solution. Patches with poor visual similarity (bottom row) are assigned zero weights in our approach. Using least squares, these weights become negative and lead to a cancellation of any votes for the kidney (shown with red labels) in the patch-based fusion.

Relations to sparsity promoting minimisation: An alternative solution to finding an optimal patch combination, [6], introduces an additional ℓ_1 penalty on the coefficients: $\lambda^*|\boldsymbol{c}|_1$, where $\lambda^* = \lambda \max(|\mathbf{A}\boldsymbol{b}|)$. Adding this term for Eq. 3 instead of the non-negativity constraint transforms the minimisation into a Lasso problem [13], which promotes a sparse selection of atlases. In [6] it was shown that a sparse selection of atlases may improve label fusion. It, however, adds an additional free parameter λ, which regulates the degree of sparsity. The minimisation of Lasso is computationally more expensive than NNLS and no gain is achieved for the image registration step (since all registered atlases are required regardless). Furthermore, it has been recently shown in [14] that a non-negativity constraint may achieve a superior sparse selection without the need for balancing the ℓ_1 penalty. We will further strengthen this theoretical assumption in our experiments.

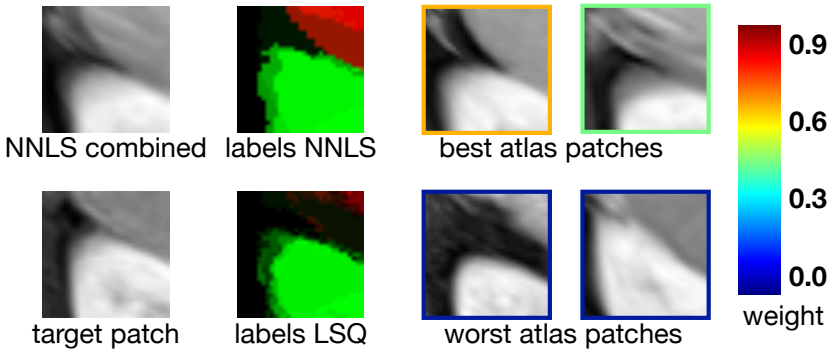

Fig. 1. Visual example of non-negative least square (NNLS) patch-based fusion for a challenging anatomical location close to spleen and kidney. After calculating the coefficients $\boldsymbol{c}$ the best and worst patches from the atlas database can be determined. The coloured borders indicate the NNLS-weights (both bottom patches have zero weight). When the NNLS weights are applied for patch-based label the probabilistic segmentation estimate (green indicates kidney, red spleen) is improved compared to least squares (LSQ) and the combined intensity patches reconstruct the target patch visually well. In the least squares approach the negative weight of e.g. the patch at the bottom right cancels out votes for the spleen (red colour) and leads to a deterioration of the fusion.

Efficient implementation for overlapping patches: We assume $R > n$, i.e. the resulting linear system is overdetermined and the same solution is obtained for $\mathbf{A}^T\mathbf{A}\boldsymbol{c} = \mathbf{A}^T\boldsymbol{b}$. Using matrix notation this yields the following equations (omitting x for brevity):

$$\begin{bmatrix} \sum(\mathbf{Y}'_1)^2 & \sum\mathbf{Y}'_1\mathbf{Y}'_2 & \dots & \sum\mathbf{Y}'_1\mathbf{Y}'_n \\ \sum\mathbf{Y}'_2\mathbf{Y}'_1 & \sum(\mathbf{Y}'_2)^2 & \dots & \sum\mathbf{Y}'_2\mathbf{Y}'_n \\ \vdots & \vdots & \ddots & \vdots \\ \sum\mathbf{Y}'_n\mathbf{Y}'_1 & \sum\mathbf{Y}'_n\mathbf{Y}'_2 & \dots & \sum(\mathbf{Y}'_n)^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \mathbf{Y}'_1\mathbf{X}' \\ \mathbf{Y}'_2\mathbf{X}' \\ \vdots \\ \mathbf{Y}'_n\mathbf{X}' \end{bmatrix} \tag{4}$$

This reduces the number of equations from R to n and it becomes clear that all entries are normalised cross-correlations (scaled by a factor R) between the locally considered patches. Since the elements in Eq. 4 have to be calculated for nearly all overlapping patches in an image, many calculations are repetitive. Following the work of [15] on local canonical correlation, which is based on the principle of guided filtering [16], we can replace the summations in Eq. 4 with box convolution filters. The filter operations have a computational complexity that is independent of the patch size R and thus enables fast computation for large neighbourhoods. For implementation details we refer to [15], but in contrast to that work, we consider correlation (not covariance) matrices and we have to calculate local standard deviations (numerator of Eq. 1). Note that this simplification cannot be used in *joint label fusion*, since the product of absolute differences cannot be expanded that way (c.f. Eq. 18 [4]). To stabilise the solution a small regularisation constant ϵ may be added to the diagonal of $\mathbf{A}^T\mathbf{A}$.

manual segmentation [17] majority voting NNLS fusion

Fig. 2. Visual comparison of segmentation outcome for two different multi-atlas label fusion techniques. The coronal slices show the following segmented abdominal organs: ■ liver, ■ spleen, ■ left, ■ right kidney, ■ left psoas major muscle (pmm) and ■ right pmm. Substantially improved organ delineation can be seen for the proposed non-negative least square regression compared to majority voting, especially for the right kidney and liver. Note that the bladder is not visible in this slice.

The choice of this parameter will be discussed later. After an atlas weighting vector $\boldsymbol{c}$ has been obtained for every voxel x, a structured label combination of the overlapping patches is considered. Each voxel votes for all voxels in its neighbourhood Ω_x based on its local label patch. This is equivalent to spatially filtering the weighting field of each label with a uniform kernel of size Ω_x and has already been used in [4] and [16].

4 Experiments

We evaluated our method for ten abdominal MRI scans from the VISCERAL3 training dataset [17] (see Fig. 2). The images were first resampled to uniform voxel-spacings and cropped to have same dimensions. Pair-wise deformable registrations were performed using the **deeds** framework [18], which is based on discrete optimisation together with self-similarity context [19] as matching term. For image resolutions of 2.5 mm (≈3 million voxels per volume) a single registration takes only ≈10 seconds. Five multi-scale levels with control-point spacings of $[8, 7, 6, 5, 4]$ voxels, $l_{\max} = [6, 5, 4, 3, 2]$ displacement steps, and quantisations of $q = [5, 4, 3, 2, 1]$ voxels were used, which yields a capture range of 75 mm.

We used a leave-one-out cross-validation for our experiments, i.e. for every target scan there are $n = 9$ atlases. The patch-radius was fixed to $r = 4$ voxels, as this yielded the best accuracy for all evaluated techniques on our data. In our NNLS implementation, we first calculate cross-correlations between all atlases and between all atlases and the target scan using roughly $\frac{(n+3)n}{2} = 65$ box filter operations. Each volumetric filter takes about 0.05 sec. (in total of 3 sec.). Fig. 3 demonstrates the advantages compared to a naïve implementation. Only voxels without unanimous decision of all atlases (and voxels within a margin of

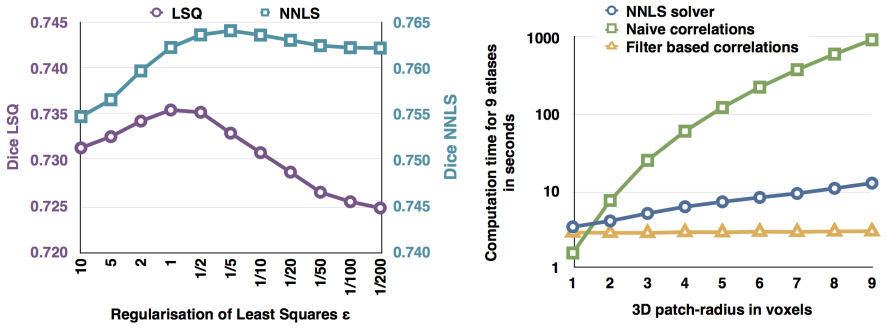

Fig. 3. Left: Influence of regularisation weight ϵ for least square regression with and without non-negative constraint on segmentation accuracy (note: the y-scales are different). Best results for NNLS are obtained for $0.5 \leq \epsilon \leq 0.1$ and $1 \leq \epsilon \leq 0.5$ for LSQ. Right: Computation times (on logarithmic scale) for calculation of all correlation matrices in Eq. 4 using naïve approach and our proposed filter-based (exact) solution, which is independent of the radius. The NNLS solution takes slightly longer for larger radii because more voxels are considered (larger spatial influence of each patch).

r around them) need to be considered for weighted NNLS fusion (so only for these, correlation matrices have to be stored). The NNLS optimisation was then performed for an average of 500×10^3 voxels per atlas in 6 sec. for all atlases. This computation time compares favourably (roughly 15 times faster) with an average of 120 sec. for *joint label fusion* using the original implementation of [4] on the same machine. Our reference implementation will be made available at `www.mpheinrich.de/software.html`.

The visual results in Fig. 2 show an improved segmentation compared to majority voting especially at boundaries between organs, e.g. liver and right kidney, which are challenging for the deformable registration. Experiments have been performed using the same registration outcome for several comparable state-of-the-art approaches. The resulting segmentations S have been numerically evaluated using the manual annotations M of seven organs (see Fig. 2 for a description) provided by [17] using the Dice overlap $D = 2|S \cap M|/(|S| + |M|)$. Our approach significantly improves with an average Dice of $D = 0.764$ over majority voting ($D = 0.713$), independent local NCC-similarity weighted fusion ($D = 0.729$), joint least square regression [3] ($D = 0.735$) and ℓ_1-Lasso regression ($D = 0.734$). The results obtained for the joint distance based fusion of [4] are only slightly inferior to ours with $D = 0.756$, but incur a much large computation time. Fig. 4 compares the distribution of Dice overlaps for all methods. Our approach implicitly yields roughly 40 % of zero coefficient. When combining the non-negativity and ℓ_1 constraints, a larger sparsity can be obtained (69 %, 53 % and 43 % for λ=0.5, 0.2 and 0.1 respectively). In accordance with [14] adding the ℓ_1 penalty deteriorates the coefficient estimation compared to only using NNLS. For $\lambda \leq 0.1$ both sparsity and accuracy of ℓ_1+NNLS come close to our results.

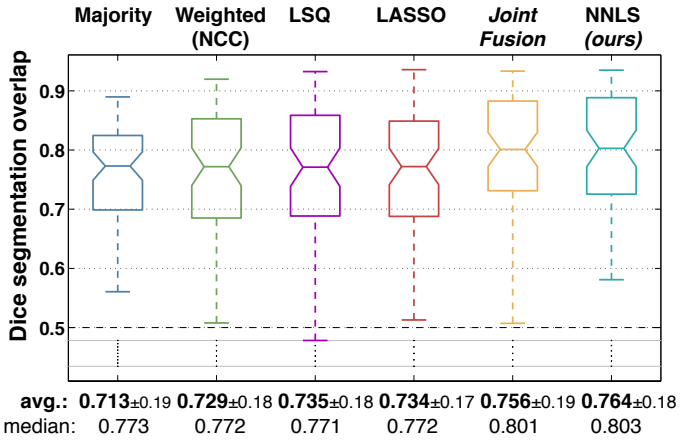

Fig. 4. Numerical results for segmentation accuracy in 10 MRI scans of [17] using multi-atlas fusion in a leave-one-out fashion. Average Dice scores over seven abdominal structures are compared for majority voting, locally weighted fusion with normalised cross-correlation (NCC), least squares regression (LSQ) analogous to [3], joint label fusion [3] and our proposed nonnegative least square (NNLS) regression. The improvements of NNLS compared to LSQ, LASSO and NCC are significant with $p = 0.049$, $p = 0.031$, and $p = 0.022$ respectively using a signed Wilcoxon rank-sum test.

5 Conclusion

We have presented a very accurate and fast multi-atlas label fusion technique that jointly models patch correlations using non-negative least squares regression. Experimental validation for abdominal scans shows significant improvements over straightforward patch-based regression or independent similarity-based weighting. On average our approach also outperforms *joint label fusion* [4], which is computationally much more involved. We have demonstrated that for this application non-negativity constraints are favourable over sparsity promoting ℓ_1 regularisation with clear advantages in computational demand and parameter choice. Future work could employ this fast and reliable fusion strategy together with approximate (pseudo-geodesic) pair-wise deformable registrations in order to further narrow the performance gap to machine learning approaches.

Acknowledgements. We are very grateful to H. Matuschek (University of Potsdam) for making available an efficient implementation for non-negative least square optimisation within the Eigen3 library (http://eigen.tuxfamily.org).

References

1. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans. Medical Imaging **23**(7) (2004) 903–921

2. Commowick, O., Akhondi-Asl, A., Warfield, S.K.: Estimating a reference standard segmentation with spatially varying performance parameters: Local MAP STAPLE. IEEE Trans. Medical Imaging **31**(8) (2012) 1593–1606
3. Wang, H., Suh, J.W., Das, S., Pluta, J., Altinay, M., Yushkevich, P.: Regression-based label fusion for multi-atlas segmentation. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 1113–1120
4. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P., et al.: Multi-atlas segmentation with joint label fusion. IEEE Pattern Analysis and Machine Intelligence **35**(3) (2013) 611–623
5. Lawson, C.L., Hanson, R.J.: Solving least squares problems. Volume 161. SIAM (1974)
6. Zhang, D., Guo, Q., Wu, G., Shen, D.: Sparse patch-based label fusion for multi-atlas segmentation. In: Multimodal Brain Image Analysis. Springer (2012) 94–102
7. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. Medical image analysis **18**(6) (2014) 881–890
8. Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Trans. Medical Imaging **29**(10) (2010) 1714–1729
9. Coupé, P., Manjón, 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) (2011) 940–954
10. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE Trans. Medical Imaging **30**(10) (2011) 1852–1862
11. Asman, A.J., Landman, B.A.: Non-local statistical label fusion for multi-atlas segmentation. Medical image analysis **17**(2) (2013) 194–208
12. Awate, S.P., Zhu, P., Whitaker, R.T.: How many templates does it take for a good segmentation?: Error analysis in multiatlas segmentation as a function of database size. In: Multimodal Brain Image Analysis. Springer (2012) 103–114
13. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288
14. Slawski, M., Hein, M.: Sparse recovery by thresholded non-negative least squares. In: Advances in Neural Information Processing Systems. (2011) 1926–1934
15. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Multispectral image registration based on local canonical correlation analysis. In: MICCAI 2014. Springer (2014) 202–209
16. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Analysis and Machine Intelligence **35**(6) (2013) 1397–1409
17. Hanbury, A., Müller, H., Langs, G., Weber, M.A., Menze, B.H., Salas Fernandez, T.: Bringing the algorithms to the data: Cloud-based benchmarking for medical image analysis. In: Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics. LNCS 7488, Springer (2012) 24–29
18. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans. Medical Imaging **32**(7) (2013) 1239–1248
19. Heinrich, M.P., Jenkinson, M., Papiez, B.W., Brady, M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N., eds.: MICCAI. LNCS. Springer (2013) 187–194