# Hashing-Based Atlas Ranking and Selection for Multiple-Atlas Segmentation

Amin Katouzian[1(✉)], Hongzhi Wang[1], Sailesh Conjeti[2], Hui Tang[1], Ehsan Dehghan[1], Alexandros Karargyris[1], Anup Pillai[1], Kenneth Clarkson[1], and Nassir Navab[2]

[1] IBM Almaden Research Center, San Jose, CA, USA
akatouz@us.ibm.com
[2] Computer Aided Medical Procedures, Technische Universität München, Munich, Germany

**Abstract.** In this paper, we present a learning based, registration free, atlas ranking technique for selecting outperforming atlases prior to image registration and multi-atlas segmentation (MAS). To this end, we introduce *ensemble hashing*, where each data (image volume) is represented with *ensemble* of hash codes and a learnt distance metric is used to obviate the need for pairwise registration between atlases and target image. We then pose the ranking process as an assignment problem and solve it through two different combinatorial optimization (CO) techniques. We use 43 unregistered cardiac CT Angiography (CTA) scans and perform thorough validations to show the effectiveness and superiority of the presented technique against existing atlas ranking and selection methods.

## 1 Introduction

In atlas-based segmentation, the goal is to leverage labels in a single fixed (template) atlas for segmenting a target image. The assumption that the spatial appearance of anatomical structures remains almost the same across and within databases is not always held, which results in systematic registration error prior to label propagation. Alternatively, in MAS [2,15], multiple atlases are deployed, encompassing larger span of anatomical variabilities, for compensating large registration errors that may be produced by any single atlas and increasing performance. Thus, the challenge is to optimally select a number of outperforming atlases without compromising segmentation accuracy and computational speed.

For this reason, different atlas selection methods have been proposed based on (1) image similarity between atlases and target image, defined over original space or manifolds [1,11,17], (2) segmentation precision [7], and (3) features representations for supervised learning [12]. Although the aim is to reduce the number of required pairwise registrations between query image and less applicable (dissimilar) atlases, all above mentioned selection techniques themselves rely
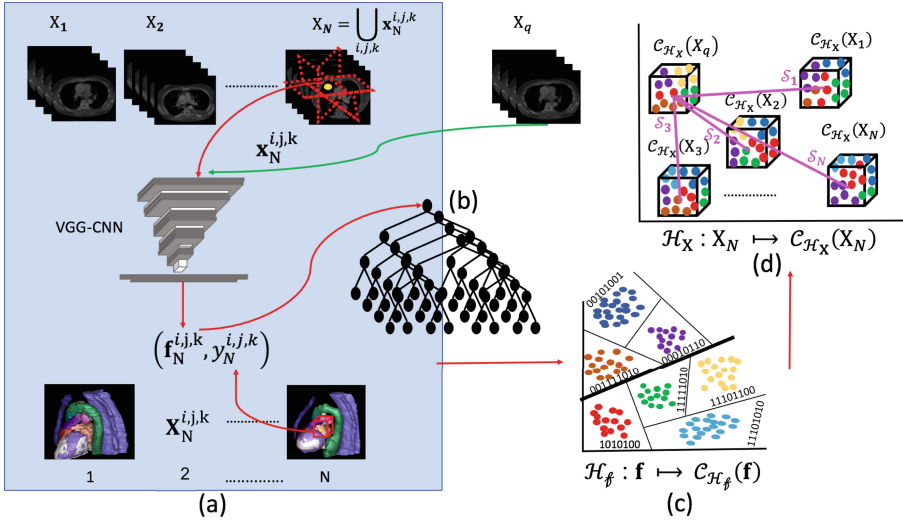
**Fig. 1.** Overall schematic of proposed method. **Training:** The VGG-convolutional neural network (CNN) [14] is used for feature extraction from 3D atlases (a) and training of mHF $H_f$ (b). The feature space is parsed and hashed $C_{H_f}(f)$ by preserving similarity among each and every organ (c). **Retrieval:** The VGG-CNN features are extracted from query ($\mathbf{X}_q$) and fed to the learnt mHF to generate *ensemble* hash codes $C_{H_X}(\mathbf{X}_q)$. The CO is then used as similarity $\mathcal{S}$ measure for retrieving $\mathcal{N}$ closest matches (d).

on non-rigid registration as a preprocessing step. In fact, at the first glance, registration seems to be inevitable since the selection strategy is often established on the premise of capturing similarity between pairwise images.

This motivated us to investigate potential extension of hashing forests [3] as an alternative solution where similarity can be measured within a registration free regime. The rationale behind the use of hashing forests is further substantiated by [8,10] where the former uses forests for the task of approximate nearest neighbor retrieval and the latter introduced a novel training scheme with guided bagging both applied for segmentation of CT images. Although they can be utilized for the purpose of label propagation as part of MAS, however, their direct generalizations to atlas selection are doubtful due to lack of ranking metric and strategy. In essence, we propose similar idea of preserving similarity in local neighborhoods but what makes our method suitable for atlas ranking is inclusion of hashing in neighborhood approximation, which serves as a basis for defining a definitive metric for ranking through CO techniques [13].

Our work is fundamentally different from [3] from two main perspectives. First, unlike [3], where each data point is represented by a single class sample or hash code (i.e. organ type, Fig. 1(c)) in hashing space, we parse the hashing space with ensemble of codes derived from features representing every organ within the volumetric CT images, Fig. 1(d). Secondly, due to ensemble representations, the retrieval/ranking task becomes a matching or assignment problem in Hamming

space that we solve through CO techniques. We use KuhnMunkres (also known as Hungarian) algorithm [9] as well as linear programming (LP) to rank and select the closest atlases to the query. We perform similar validation scheme to [11] and use normalized mutual information (NMI) as similarity measure for atlas selection. Finally, we quantify the overall performance through MAS algorithm proposed by [15].

Our contributions include: **(1)** extending [3] for volumetric data hashing and introducing the concept of *ensemble hashing* for the first time, **(2)** employing hashing for atlas selection and ranking as part of MAS, **(3)** eliminating pairwise registration, which makes atlas selection extremely fast, and **(4)** deploying CO as a solution to atlas ranking problem that will also be shown to be a viable solution for similarity matching in the context of *ensemble hashing*.

## 2    Methodology

### 2.1    Volumetric Ensemble Hashing Through mHF

Figure 1 shows the schematic of proposed method and we refer readers to [3] for detailed description about mHF. For a given subset of training atlases $\mathcal{X}_N = \{\mathbf{x}_v : \mathbb{R}^3 \to \mathbb{R}\}_{v=1}^N$ and corresponding labels $\mathcal{Y}_N = \{\mathbf{y}_v : \mathbb{R}^3 \to \mathbb{N}\}_{v=1}^N$ we perform random sampling with minimum rate of $f_{s_{min}}$ to extract features from data represented in three orthogonal planes centered at $(i, j, k)$ spatial coordinate using VGG network $\mathcal{F}_N^{(i,j,k)} = \{\mathbf{f}_v^{(i,j,k)} \in \mathbb{R}^d\}_{v=1}^N$. For simplicity, $\mathcal{X}_N$, $\mathbf{x}_v$, $\mathbf{f}_v^{(i,j,k)}$ are used interchangeably with $\mathcal{X}$, $\mathbf{x}$, $\mathbf{f}$, respectively throughout paper. An $n$ bit hash function $h_f$ is defined that maps $\mathbb{R}^d$ to $n$-dimensional binary Hamming space $h_f : \mathbb{R}^d \to \{1, 0\}^n$ with $\mathcal{C}_{h_f}(\mathbf{f}) = h_f(\mathbf{f})$ codeword. The hashing forest comprises of $K$ independently hashing trees $\mathcal{H}_f = \{h_f^1, \cdots, h_f^K\}$ that encodes each sample point $\mathbf{f}$ from $\mathbb{R}^d$ to $nK$-dimensional Hamming space $\{1, 0\}^{nK}$ such that $\mathcal{H}_f : \mathbf{f} \to \mathcal{C}_{H_f}(\mathbf{f}) = \mathcal{C}_{h_f^1}(\mathbf{f}), \cdots, \mathcal{C}_{h_f^K}(\mathbf{f})$. Given $\mathcal{H}_f$, we encode all organs in training atlases $\mathcal{X} \in \mathbb{R}^{d \times N \times o}$ as $\mathcal{H}_f : \mathcal{F} \to \mathcal{C}_{H_f}(\mathcal{F}) = \mathcal{C}_{h_f^1}(\mathcal{F}), \cdots, \mathcal{C}_{h_f^K}(\mathcal{F}) \in \{1, 0\}^{nK \times N \times o}$, where $o$ is the total numbers of organs.

The mHF parses and hashes the latent feature space while preserving similarity among organs, Fig. 1(c). For *ensemble* representation of each volume, we incorporate the coordinates of sampling points $(i, j, k)$ into hashing scheme as follows:

$$\mathcal{H}_\mathcal{X} : \underbrace{\bigcup_{(i,j,k)} \mathbf{f}_N^{(i,j,k)}}_{\mathcal{X}_N} \to \bigcup_{(i,j,k)} C_{\mathcal{H}_f}\left(\mathbf{f}_N^{(i,j,k)}\right) = \underbrace{\bigcup_{(i,j,k)} \left\{ C_{h_f^1}\left(\mathbf{f}_N^{(i,j,k)}\right), \cdots, C_{h_f^k}\left(\mathbf{f}_N^{(i,j,k)}\right) \right\}}_{C_{H_\mathcal{X}}(\mathcal{X}_N)}$$

(1)

where $C_{H_\mathcal{X}}(\mathcal{X}_N) \in \{1, 0\}^{(nk \times N \times n_s)}$ and $n_s$ is the number of sampling points.

### 2.2    Retrieval Through Combinatorial Optimization (CO)

In classical hashing based retrieval methods, given a query $\mathbf{x}_q$, the inter-sample similarity $S$ could be computed as pairwise Hamming distance $D_H$ between
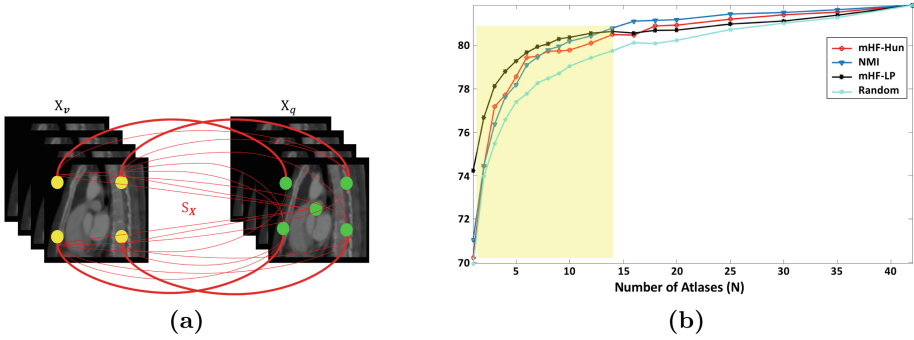
**Fig. 2.** Distance matching illustration between $\mathcal{X}_v$ and $\mathcal{X}_q$ volumes represented by 4 and 5 sampling points, respectively. The top matches are depicted by thicker edges (a). The MAS performance using different atlas ranking techniques. The average Dice value is reported over all organs when $N = [1{:}9,\ 10{:}2{:}20,\ 25,\ 30,\ 35,\ 42]$ (b). The highlighted area covers number of atlases where mHF-LP outperforms or its performance almost becomes equal to NMI (N = 14).

the hash codes of query $C_\mathcal{H}(\mathbf{x}_q)$ and each sample in training database $C_\mathcal{H}(\mathbf{x}_v)$ through logical $\mathtt{xor}$ as $S_{\mathbf{x}_{qv}} = D_H(\mathbf{x}_q, \mathbf{x}_v) = C_\mathcal{H}(\mathbf{x}_q) \oplus C_\mathcal{H}(\mathbf{x}_v) \in \mathbb{N}$. The samples whose hash codes fall within the Hamming ball of radius $r$ centered at $C_\mathcal{H}(\mathbf{x}_q)$ *i.e.* $D_H(\mathbf{x}_q, \mathbf{x}_v) \leq r$ are considered as nearest neighbors. However, for our problem, the pairwise Hamming distance between hash codes of two volumes is $S_{\mathcal{X}_{qv}} = D_H(\mathcal{X}_q, \mathcal{X}_v) = C_\mathcal{H}(\mathcal{X}_q) \oplus C_\mathcal{H}(\mathcal{X}_v) \in \mathbb{R}^2$ and finding the nearest neighbors seems intractable.

Both volumes $\mathcal{X}_v$ and $\mathcal{X}_q$ are represented by $n_s^v \times l$ and $n_s^q \times l$ features in latent space, resulting in $n_s^v$ and $n_s^q$ hash codes, where $n_s^v$ and $n_s^q$ are number of sampling points, Fig. 2(a) ($n_s^v = 4$, $n_s^q = 5$). The Similarity $S_\mathcal{X}$ can be posed as multipartite Hamming distance matching problem by resolving correspondence between pairwise Hamming tuples of length 2. To this end, we construct set of nodes in each volume, where each node comprises of position of sampling point $(i, j, k)$ and corresponding hash code. The problem now is to find a set of edges that minimizes the matching cost (total weights), which can be tackled by CO methods like assignment problem [4,5] in 2-D Hamming space.

## 2.3   Similarity Estimation Through Assignment Problem with Dimensionality Reduction

Motivated by [5], we assign costs $c_{qv}$ to pairwise Hamming distances $S_{\mathcal{X}_{qv}}$, which represents the likelihood of matching sampling data in two volumes. The overall cost shall be minimized with respect to $c_{qv}$ as follows:

$$\min \sum_q \sum_v c_{qv} S_{\mathbf{x}_{qv}} \ \text{s.t.} \ \begin{cases} \sum_q c_{qv} = 1 \forall q \\ \sum_v c_{qv} = 1 \forall v \\ c_{qv} \in [0, 1] \end{cases} \tag{2}$$

**Table 1.** The MAS results (Dice: mean $+/-$ Std) for all organs using mHF-Hun, mHF-LP, and NMI atlas selection techniques when $N = [1{:}10, 12, 14]$ atlases are used.

| Organ / Method | ST | A-Ao | D-Ao | R-PA | Ve | R-At | L-At | RV |
|---|---|---|---|---|---|---|---|---|
| mHF-Hun | 77.91±5.61 | 77.72±8.92 | 85.51±7.57 | 85.44±4.16 | 88.95±3.75 | 86.22±5.39 | 90.69±6.81 | 86.13±7.80 |
| mHF-LP | **78.95±3.28** | **77.48±1.79** | 85.64±1.69 | 85.29±1.80 | **89.64±0.86** | **86.05±1.67** | 90.52±1.23 | **85.90±1.64** |
| NMI | 77.94±4.08 | 76.85±2.89 | **86.25±1.57** | **85.79±1.73** | 88.41±2.24 | 85.73±2.10 | **90.60±1.39** | 85.64±2.18 |

| Organ / Method | LV | Myo | L-PA | T-PA | Ao-R | Ao-A | S-VC | I-VC |
|---|---|---|---|---|---|---|---|---|
| mHF-Hun | 78.94±6.11 | 84.06±8.27 | 79.53±10.30 | 82.23±6.17 | 65.82±14.38 | 61.57±9.66 | 73.82±7.10 | 60.95±12.06 |
| mHF-LP | 78.31±1.67 | 83.90±1.40 | **80.49±2.46** | **82.54±1.64** | **66.14±2.27** | **67.32±2.70** | **75.25±2.76** | 61.67±1.57 |
| NMI | **78.53±1.94** | **84.07±1.81** | 79.13±2.88 | 82.49±2.09 | 63.58±5.39 | 65.47±7.09 | 74.58±4.22 | **62.30±2.23** |

Given $n_s^v$ and $n_s^q$ sampling points, $(n_s^v \times n_s^q)!$ solutions exist, which makes computation of all cost coefficients infeasible as each volume is often represented by 3000 sampling points, on average. To overcome this limitation, we will solve the following LP problem that shares the same optimal solution [5]:

$$\widetilde{S}_{\mathcal{X}_{qv}} = \min \sum_q \sum_v c_{qv} \hat{S}_{\mathbf{x}_{qv}} \text{ where } \hat{S}_{\mathbf{x}_{qv}} = \begin{cases} S_{\mathbf{x}_{qv}} & \text{if } S_{\mathbf{x}_{qv}} \leq \eta \\ \infty & \text{if } S_{\mathbf{x}_{qv}} > \eta \end{cases} \tag{3}$$

where $\eta = \frac{1}{n_s^q} \sum_q S_{\mathbf{x}_{qv}}$. As a complementary analysis, we will solve the same problem using Hungarian algorithm and refer readers to [9] (Table 1).

## 3 Experiments and Results

We compare our method against an atlas selection technique (baseline) similar to [1]. The performance of each algorithm is evaluated by MAS with joint label fusion employing [15]. Like [12], in our quantification, we use the Dice Similarity Coefficient (DSC) between manual ground truths and automated segmentation results. We also justify the need for our proposed ranking technique in MAS against deep learning segmentation methods and deploy multi-label V-net [6].

### 3.1 Datasets

We collected 43 cardiac CTA volumes with labels for 16 anatomical structures including sternum(ST), ascending(A-Ao)/descending aorta(D-Ao), left(L-PA)/right (R-PA)/trunk pulmonary artery(T-PA), aorta/arch(Ao-A)/root(Ao-R), vertebrae(Ve), left(L-At)/right atrium(R-At), left(LV)/right ventricle(RV), lLV myocardium(Myo), and superior(S-VC)/inferior vena cava(I-VC). Each image has isotropic in-plane resolution of $1\,\mathrm{mm}^2$. The slice thickness varies from $0.8\,\mathrm{mm}$ to $2\,\mathrm{mm}$. All images are intensity equalized to eliminate intensity variations among patients and then resampled to voxel size of $1.5\,\mathrm{mm}$ in all dimensions.
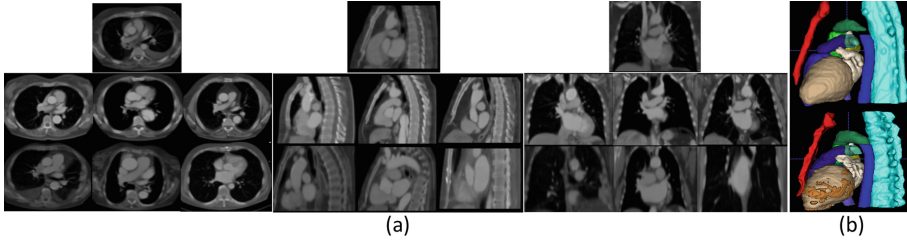
**Fig. 3.** The mHF-LP qualitative results. Query volume (top row) in axial, coronal, and sagittal views (from left to right) and corresponding retrieve volumes in corresponding views. The middle and bottom rows show the most similar and dissimilar cases (from left to right), respectively (a). The 3D visualization of segmentation results for the query volume (top) and corresponding manual ground truth (bottom) (b).

## 3.2    Validation Against Baseline

In this section, we validate the performance of the proposed hashing based atlas ranking selection strategies (mHF-LP, described in Sect. 2.3, and mHF-Hungarian (mHF-Hun)) on segmentation results against [1] where NMI is used as similarity metric for atlas selection and registration. We use $f_{s_{min}} = 50$ and resize each sampling voxel to $224 \times 224$ as an input to VGG network. We then extract $d = 4096$ features from FC7 layer for training the mHF. The forest comprises of 16 trees with depth of 4 to learn and generate 64 bits hashing codes.

We indicate that the NMI atlas selection method requires deformable registration whereas neither mHF-LP nor mHF-Hun does. Once atlases are ranked and selected using any approach, a global deformable registration is performed as part of MAS. We perform 43-fold leave-one-out cross validation and train models using all 16 labels (organs: $\iota = 1, \cdots, 16$). Figure 2(b) shows the average Dice values over all organs when $N$ nearest atlases are selected using NMI, mHF-LP, and mHF-Hun algorithms. To speed up experiments, we introduced intervals while increasing $N$ as the top similar cases weigh more in segmentation performance. As seen, the mHF-LP outperforms up to $N = 14$ where its performance equals the NMI (Dice = 80.66%). This is an optimal number of selected atlases where only 1.22% of performance is compromised in contrast to the case that we use all atlases ($N = 42$, Dice = 81.88%).

We performed an additional experiment by selecting atlases randomly and repeated the experiment five times. The averaged results are shown in Fig. 2 (cyan graph). This substantiates that the performance of our proposed techniques is solely depending on retrieval efficacy and not registration as part of MAS. Looking at qualitative results demonstrated in Fig. 3, we can justify the mHF-LP superior performance when a few atlases are selected ($N \leq 9$). As we can see, the top ranked atlases are the most similar ones to query and therefore they contribute to segmentation significantly. As we further retrieve and add more dissimilar atlases, the mHF-LP performance almost reaches a plateau.
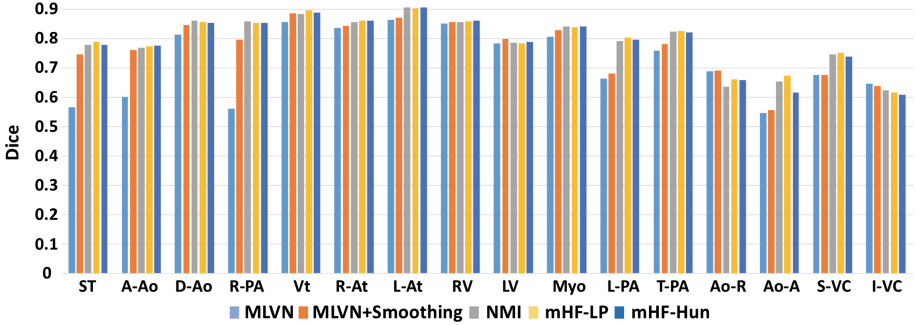
**Fig. 4.** The Dice segmentation results of all organs using MLVN with and without smoothing in contrast to MAS algorithm with proposed atlas selection techniques and NMI (baseline).

### 3.3 Validation Against Multi-label V-Net (MLVN)

We perform a comparative experiment to study the need for such atlas selection techniques as part of MAS process in contrast to CNN segmentation methods. We performed volumetric data augmentation through random non-rigid deformable registration during training. The ratio of augmented and non-augmented data was kept 1:1. Due to high memory requirements in 3D convolution, the input image is down-sampled to $2\,\mathrm{mm} \times 2\,\mathrm{mm} \times 3.5\,\mathrm{mm}$, and a sub-image of size $128 \times 192 \times 64$ is cropped from the center of the original image and fed to the network. For segmentation, the output of the MLVN [6] is up-sampled to the original resolution padded into the original spatial space. We preserve the same network architecture as [6] and implemented the model in CAFFE.

Figure 4 demonstrates the results of MLVN segmentation with and without smoothing on 4-fold cross validation along with mHF-LP, mHF-Hun, and NMI. As expected, we obtained better results with smoothing. We performed the t-test and found significant difference between generated results by MLVN and the rest ($p > 0.05$). The performance of MLVN is fairly comparable with the rest excluding the results for L-PA and Ao-A. Both are relatively small organs and may not be presented in all volumes, therefore, we could justify that the network has not seen enough examples during training despite augmentation.

### 3.4 Discussion Around Performance and Computational Speed

The MLVN is very fast at testing/deployment stage, particularly when performed on GPU. It takes less than $10\,\mathrm{s}$ to segment a 3D volume on one TITAN X GPUs with $12\,\mathrm{GB}$ of memory. However, The averaged segmentation performance ($\bar{\mathrm{Dice}}_{\mathrm{MLVN+smoothing}} = 0.7663$) was found to be smaller than the rest ($\bar{\mathrm{Dice}}_{\mathrm{mHF\text{-}LP}} = 0.7969$, $\bar{\mathrm{Dice}}_{\mathrm{mHF\text{-}Hun}} = 0.7909$, $\bar{\mathrm{Dice}}_{\mathrm{NMI}} = 0.7921$). The MAS generates more accurate results but at the cost of high computational burden due to the requirement for pairwise registrations and voxel-wise label fusion, of

which the latter is not the focus of this work. We refer readers to [16] where the trade-off between computational cost and performance derived by registration has been thoroughly investigated.

We parallelized the pairwise registrations between atlases and deployed the MAS on Intel(R) Xeon(R) CPU E5-2620 v2 with frequency of 2.10 GHz. In the NMI based atlas selection technique, each deformable registration took about 55 s on 3 mm$^3$ resolution. In contrast, solving the LP problem took only 10 s. By looking at Fig. 2, using $N = 14$, where we achieve reasonbaly good segmentation performance (Dice = 80.66%), we can save up to $14 \times 45 = 630$ s. The computational speed advantage of proposed method can be more appreciated in presence of large atlases especially when at least equal performance is achievable. Moreover, in the presence of limited data, we can achieve reasonably good segmentation performance using MAS algorithm with ranking, which seems very challenging for CNNs as they are greatly depending on availability of large amount of training data.

## 4   Conclusions

In this paper, for the first time, we proposed a hashing based atlas ranking and selection algorithm without the need for pairwise registration that is often required as a preprocessing step in existing MAS methods. We introduced the concept of *ensemble hashing* by extending mHF [3] for volumetric hashing and posed retrieval as an assignment problem that we solved through LP and Hungarian algorithm in Hamming space. The segmentation results were benchmarked against the NMI based atlas selection technique (baseline) and MLVN. We demonstrated that our retrieval solution in combination with MAS boosts up computational speed significantly without compromising the overall performance. Although the combination is still slower than CNN based segmentation at deployment stage it generates better results especially in presence of limited data. As future work, we will investigate the extension of the proposed technique for organ- or disease-specific MAS by confining the retrieval on local regions (organ level) rather global (whole volume level).

## References

1. Aljabar, P., et al.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage **46**(3), 726–738 (2009)
2. Artaechevarria, X., et al.: Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Med. Imag. **28**(8), 1266–1277 (2009)
3. Conjeti, S., et al.: Metric hashing forests. Med. Image Anal. **34**, 13–29 (2016)
4. Jain, A.K., et al.: Matching and reconstruction of brachytherapy seeds using the Hungarian algorithm (MARSHAL). Med. Phys. **32**(11), 3475–3492 (2005)
5. Lee, J., et al.: Reduced dimensionality matching for prostate brachytherapy seed reconstruction. IEEE Tran. Med. Imaging **30**(1), 38–51 (2011)

6. Tang, H., et al.: Segmentation of anatomical structures in cardiac CTA using multi-label V-Net. In: Proceedings of the SPIE Medical Imaging (2018)
7. Jia, H., et al.: ABSORB: Atlas building by self-organized registration and bundling. NeuroImage **51**(3), 1057–1070 (2010)
8. Konukoglu, E., et al.: Neighbourhood approximation using randomized forests. Med. Image Anal. **17**(7), 790–804 (2013)
9. Kuhn, H.W., et al.: The Hungarian method for the assignment problem. Naval Res. Logist. Q. **2**, 83–97 (1955)
10. Lombaert, H., Zikic, D., Criminisi, A., Ayache, N.: Laplacian forests: semantic image segmentation by guided bagging. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 496–504. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10470-6_62
11. Lotjonen, J.M.P., et al.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage **99**, 2352–2365 (2010)
12. Sanroma, G.: Learning to rank atlases for multiple-atlas segmentation. IEEE Tran. Med. Imaging **33**(10), 1939–1953 (2014)
13. Schrijver, A., et al.: A Course in Combinational Optimization. TU Delft, Delft (2006)
14. Simonyan, K., et al.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2015)
15. Wang, H., et al.: Multi-atlas segmentation with joint label fusion. IEEE Tran. PAMI **35**(3), 611–623 (2013)
16. Wang, H., et al.: Fast anatomy segmentation by combining low resolution multi-atlas lebel fusion with high resolution corrective learning: an experimental study. In: Proceedings of the ISBI, pp. 223–226 (2017)
17. Wolz, R., et al.: LEAP: learning embeddings for atlas propagation. NeuroImage **49**(4), 1316–1325 (2010)