2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

AUTOMATIC MULTIMODAL REGISTRATION VIA INTRAPROCEDURAL CONE-BEAM CT SEGMENTATION USING MRI DISTANCE MAPS

Zachary Augenfeld[1] *MingDe Lin* [2,3] *Julius Chapiro, MD* [2] *James Duncan*[1,2]

[1] Department of Biomedical Engineering, Yale University, New Haven, CT, USA
[2] Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA
[3] Visage Imaging, Inc., San Diego, CA, USA

ABSTRACT

Accurate multimodal registration is integral when fusing spatial information from two or more medical images. Specifically, image-guided procedures involve acquiring images under less than ideal conditions, so the ability to map segmented regions of interest from a diagnostic image into the intraprocedural imaging domain becomes especially important. However, standard methods of multimodal registration may not be feasible, depending on intraprocedural image quality. In order to deal with such cases, in this paper we propose a series of two convolutional neural networks performing segmentation of the target image, the predicted outputs of which are then utilized as inputs to a feature-matching registration framework. Each network is trained in a supervised fashion, with robust point matching (RPM) being performed interstitially to generate signed distance maps to be included in training the second network. By supplementing the target image segmentation with dense spatial information derived from the source image, the accuracy of both segmentation and registration is improved. This provides the first fully automated framework for registering diagnostic MRI to intraprocedural cone-beam CT (CBCT) images used to guide transcatheter arterial chemoembolization (TACE), a standard interventional therapy used to treat primary liver cancer.

Index Terms— Multimodal registration, Cone-beam CT, Image-guided intervention, Deep learning, Segmentation, Liver cancer

1. INTRODUCTION

Liver cancer is the fourth most common cause of cancer death worldwide, with an estimated 782,000 deaths in 2018[1]. Its incidence is the fastest rising of any cancer, with an increase of 75% since 1990[2]. Although hepatocellular carcinoma (HCC), the most common type of primary liver cancer, can be treated with surgical resection and liver transplantation in some cases, local ablation therapies are much more common. These therapies are minimally invasive procedures in which tumors are targeted locally with embolization agents and/or chemotherapy. Specifcally, in transcatheter arterial chemoembolization (TACE), lipiodol is delivered to tumors via a catheter traveling from the femoral artery to the hepatic artery, guided by cone-beam CT (CBCT) imaging acquired intraprocedurally. Multi-parameter MRI, including gadolinium-enhanced T1 MRI (Gd-MRI), is acquired for diagnosis and planning purposes, such that segmentations are performed manually for tumors as well as the whole liver.

CBCT differs from conventional CT by utilizing a cone-beam X-ray source, 2D flat panel detectors, and the use of Feldkamp-Davis-Kress backprojection for reconstruction[3]. This allows CBCT images to be acquired relatively rapidly, permitting it to be a practical choice for guiding interventional therapies such as TACE. However, in addition to a heightened sensitivity to scatter, yielding more noise and artifacts, CBCT suffers from a contrast-to-noise ratio that is 1.5-2 times lower than that of conventional CT[4]. TACE CBCT image contrast is further reduced by imaging of the catheter, which gets mapped to extremely high voxel intensities, allowing less dynamic range for abdominal tissue. This degraded contrast complicates resolving tumor location, thereby necessitating the mapping of diagnostic tumor segmentations into the intraprocedral imaging domain, a task currently estimated by eye. Our method automatically registers the Gd-MRI to the CBCT image, potentially improving precision in targeting tumors.

The most commonly used method of multimodal registration is maximization of mutual information (MMI)[5], which relies on sufficient dynamic range to create corresponding partitions of each image by binning its relative intensities. But, as discussed above, these CBCT images suffer from a greatly reduced contrast, making MMI infeasible. Usually, when intensity-based registration methods aren't viable, feature-based registration approaches such as robust point matching (RPM)[6] may be used. However, corresponding features in each image are required, and the intraprocedural nature of TACE CBCT acquisition precludes the ability to manually label any features or landmarks in real time. There is recent work in training convolutional neural networks (CNNs) to learning registration parameters directly[7, 8], but most of these approaches require training data consisting of affinely pre-registered corresponding image pairs as input and known nonrigid transformation parameters as the target output. The lack of trustworthy a priori MR-CBCT registrations to train with makes this approach impractical. Instead, we propose a supervised deep learning framework which instead performs CBCT liver segmentation as a front end to RPM in order to register the corresponding livers.

Lastly, signed distance functions (SDFs) have been used extensively in image segmentation. Historically associated with level set approaches[9, 10], SDFs have more recently been leveraged in deep learning approaches as well, such as CNN segmentation regularization[11]. We propose a novel approach of assisting the training of a second deep neural network performing the same segmentation task with signed distance maps derived from MRI liver masks registered to CBCT using the initial network's predictions, improving both the segmentation and registration performance.

2. METHODS

Each section below describes the constituent parts of our overall proposed framework in detail. The entire pipeline is shown in Fig. 1.

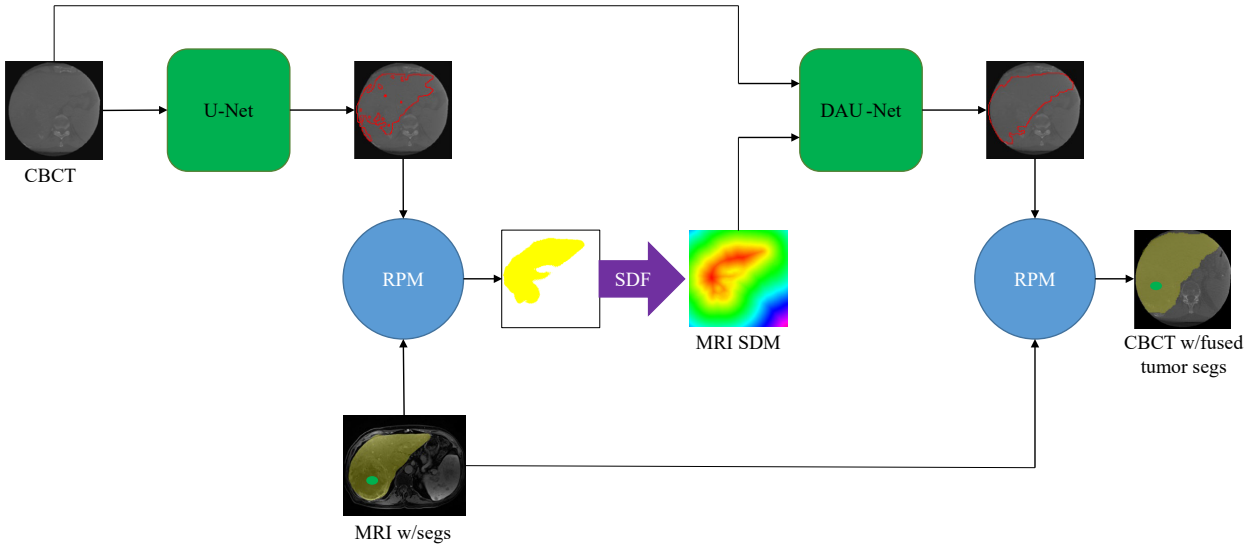


Fig. 1. Overall architecture and configuration of proposed framework.

2.1. Cone-Beam CT Liver Segmentation

A U-Net[12] was chosen as the CNN architecture, due to its current ubiquity in medical image segmentation. We adopt one of the architectures analyzed in [13], choosing a four-level U-Net designed to expect input images with 2mm isotropic voxel spacing. Since implementing a fully 3D U-Net would limit the amount of training data, all convolutional layers perform 2D convolution. However, in order to provide context in the transverse dimension while still benefiting from the relative computational simplicity of 2D processing, a 2.5D approach was implemented, in which slabs of 7 adjacent slices are processed at a time as channels of a single image. All kernel sizes and numbers of filters per layer remain the same as in [13]. Meanwhile, the size of the output predictions remain a single slice at a time. To reduce complexity, upsampling layers were used on coarse feature maps as opposed to learnable deconvolutional layers. Batch normalization[14] and RELU activations were used in each convolutional layer except the last, which used a sigmoid activation standard in binary classification tasks. It was also determined that instead of using standard binary crossentropy as a loss function, implementing a form of generalized Dice loss (GDL)[15, 16] led to more stable training. Instead of only accepting binary values, GDL extends the cardinalities and intersections used in Dice to account for output prediction scores being in the range $[0, 1]$. We found that for our purposes, the more robust implementation was that in [16], as defined below:

$$GDL = 1 - \frac{2\sum_i \hat{y}_i y_i}{\sum_i \hat{y}_i^2 + \sum_i y_i^2}$$

where $y \in \{0, 1\}$ is the true label and $\hat{y} \in [0, 1]$ is the output score of the network.

The distance-assisted U-Net (DAU-Net) is a second U-Net having the same architecture and implementation as the first, but instead of receiving a 7-slice slab as input, it receives 14 slices: the 7 slices of the CBCT image, and the corresponding slices of the signed distance map obtained from the registered MRI segmentation, both of which are described below.

2.2. Robust Point Matching and Signed Distance Maps

Robust point matching[6] (RPM) is a features-based registration framework which utilizes deterministic annealing to prevent local minima when optimizing. Its soft assignment implementation determines one-to-one correspondences between point sets without penalty terms and allows for excluding outlying points. The most prominent attribute of RPM, for the purpose of intraprocedural image registration, is its speed. At high spatial resolution, 3D CBCT liver masks may contain tens of thousands of points, so the ability to adjust parameters to quickly assign correspondences for a subset of the points, while simultaneously disregarding false positive, is quite useful. Since segmentation is performed at 2mm voxel spacing, segmentation outputs are upsampled to the CBCT's original resolution (0.655mm isotropic) before thresholding to create a binary mask. Using RPM, the manually segmented MRI liver mask is registered to this full-resolution predicted CBCT liver mask.

Even though the combination of CBCT liver segmentation and RPM registration is sufficient for an automated multimodal registration framework, we propose a novel method of jointly improving upon the initial segmentation and registration by incorporating signed distance maps (SDMs). The initially registered MRI liver masks are transformed to a signed distance map by the following formula:

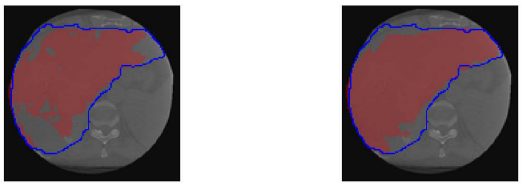
$$SDM(x) = \begin{cases} d(x, \partial L), & x \in \text{exterior of } L \\ 0 & x \text{ on } \partial L \\ -d(x, \partial L) & x \in \text{interior of } L \end{cases}$$

where L is the liver volume, ∂L is its surface, d refers to Euclidean distance, and x is the 3D coordinate. It is important to note that despite liver segmentation being performed in 2D, predicting one slice at a time, both registrations and SDMs are performed in 3D. So, including SDM slabs in training the second CNN is akin to encoding dense three-dimensional spatial information from which the distance-assisted U-Net (DAU-Net) may learn more complex and/or contextual features. For instance, even though the initial U-Net is agnostic to where within the 3D image its current slab was sampled

Table 1. Segmentation Results

Model	DSC	AUC
U-Net	0.863 ± 0.042	0.970 ± 0.013
DAU-Net	$0.878^{\star} \pm 0.024$	0.979 ± 0.008

*statistically significant difference from U-Net ($p < 0.05$)



U-Net Segmentation DAU-Net Segmentation

Fig. 2. Segmentation Performance. Blue outline: ground truth (manually segmented) liver. Red overlay: CNN segmentation output.

from, the DAU-Net is trained with information that may be correlated with slice index, potentially allowing it to incorporate indirect knowledge of the z-coordinate when updating its feature maps.

3. EXPERIMENTS AND RESULTS

3.1. Data and Implementation

The images acquired from 16 HCC patients having undergone TACE were included in our experiment. Each patient's image set consisted of diagnostic Gd-MRI and intraprocedural CBCT, acquired 1-2 months apart, from each of which whole livers were manually segmented. Due to patients being treated at different sites, MR acquisitions were performed on different scanners, but all CBCT images were acquired using a Philips C-arm system with equal image size and resolution. We implemented 8-fold cross validation, with each fold containing 12 image sets for training, 2 image sets for validation, and 2 test image sets. All networks were implemented using Keras and trained using NVIDIA GTX 100 Ti GPUs. Each CBCT image was resampled to 2mm isotropic voxel spacing, yielding an image size of $132 \times 132 \times 98$ before padding necessary for U-Net inputs. To train the initial U-Net, we randomly sampled slabs of 7 adjacent transverse slices with a mini-batch size of 9. Registrations were performed using the RPM algorithm included in BioImage Suite[17], using upsampled U-Net predictions and original MRI liver segmentations as input. The second U-Net was trained similarly to the first, but the CBCT training image slabs were concatenated with corresponding slabs of signed distance maps generated from the registered MRI liver masks.

3.2. Segmentation Results

To evaluate segmentation performance, we computed Dice similarity coefficients (DSC) between predicted liver segmentations and manual segmentations, as well as the area under the curve (AUC) of the receiver operating characteristic curve. DSC is defined as:

$$DSC(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

The outputs of both the initial U-Net and the DAU-Net were evaluated with these metrics, the means and standard deviations of which are reported in Table 1. These results show that our proposed MRI distance-assisted U-Net outperforms the baseline U-Net when performing CBCT whole liver segmentation. A sample test slice displaying this improvement can be seen qualitatively in Fig. 2.

3.3. Registration Results

To evaluate registration performance, we examined how the given MR mask and image map to the CBCT mask and image, respectively. Registrations were performed on test data as described in section 3.1 and DSC was computed between manual CBCT liver segmentations and registered MRI liver segmentations. However, in order to confirm that the RPM algorithm did not excessively match liver surfaces at the cost of structural alignment, mutual information (MI) was computed between the CBCT image and registered MRI as well. MI is defined as:

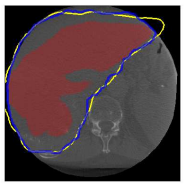
$$MI(X,Y) = \sum_x \sum_y p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)}$$

Table 2. Registration Results

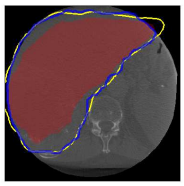
Seg. Method	Liver DSC	MI
U-Net	$0.809^{\dagger} \pm 0.041$	0.0163 ± 0.0088
DAU-Net	$0.833^{\star\dagger} \pm 0.029$	$\mathbf{0.0184^{\star} \pm 0.0107}$
Manual	0.914 ± 0.024	0.0187 ± 0.0113

*statistically significant difference from U-Net ($p < 0.05$)
†statistically significant difference from Manual ($p < 0.05$)

Lastly, these metrics were computed for the RPM registration obtained when matching manually segmented livers from each modality. The mean liver DSC (Eq. 1) and MI metrics for registrations using each CBCT segmentation derived from the initial U-Net, the proposed DAU-Net, and the manual segmentation are provided in Table 2. Although neither model's predictions align whole liver volumes as accurately as when manual segmentations are utilized, as indicated by the liver DSC values, both U-Nets yield segmentations that lead to registrations whose MI values do not significantly differ from that derived from the manual segmentations. More specifically, $p = 0.231$ and $p = 0.840$ for the U-Net and DAU-Net, respectively, when compared to using manual segmentations. At the same time, the DAU-Net provides a statistically significant improvement of both DSC and MI with respect to the initial U-Net, with $p = 0.0095$. It is worth noting that the registration using manual CBCT segmentations is not possible to normally obtain during a TACE procedure, as mentioned in Section 1, but acts here more as an upper bound. Qualitatively, Fig. 3 shows how the MR liver registration improves on an example test slice. Overall this indicates that not only does assisting a U-Net with distance maps improve both segmentation and registration accuracy, but it also yields mulitmodal registration results that are comparable to the best one can do given manual segmentations in both modalities in an automated fashion.



Using U-Net Output



Using DAU-Net Output

Fig. 3. Registration Performance. Blue outline: Manually segmented CBCT Liver. Yellow outline: MR liver registered to manually segmented CBCT liver. Red overlay: MR liver registered to CNN segmentation output.

4. CONCLUSION

In this paper, we present a fully automated end-to-end framework in which MRIs used for diagnosing and assessing HCC are registered to intraprocedural CBCT images used for guiding TACE. By providing source image distance maps to a deep segmentation network, we can improve not only target image segmentation but the multimodal registration accuracy itself. By providing dense 3D spatial location information, even when training in 2D (or 2.5D), we provide the U-Net with the potential to learn more complex or more contextual features. In addition, this is (to our knowledge) the first fully automated way of fusing tumor location derived from diagnostic MRI into the intraprocedural CBCT imaging domain. Since this multimodal registration problem is motivated by the limitations of image-guided interventional therapies such as TACE, future work will include assessing tumor registration accuracy, such as a tumor Dice similarity coefficient. Also, since the RPM registration used is dependent on the initial liver segmentation, it may be worth looking into improving the lead-off U-Net, such as by providing it with shape priors. Lastly, the method of improving a segmentation model by serially connecting another but providing distance maps in addition to images during training begs the question of whether repeating the process refines the registration accuracy. So, future work will also include implementing a cascade of U-Nets, each trained with more accurately registered distance maps, and evaluating whether each can improve the results of its preceding U-Net in an iterative manner.

5. REFERENCES

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] Global Burden of Disease Liver Cancer Collaboration, "The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015," *JAMA Oncology*, vol. 3, no. 12, pp. 1683–1691, 12 2017.

[3] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[4] Vania Tacher, Alessandro Radaelli, MingDe Lin, and Jean-François Geschwind, "How I do it: Cone-beam CT during transarterial chemoembolization for liver cancer," *Radiology*, vol. 274, no. 2, pp. 320–334, 2015.

[5] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proceedings of the IEEE*, vol. 91, no. 10, pp. 1699–1722, Oct 2003.

[6] Steven Gold, Anand Rangarajan, Chien-Ping Lu, Suguna Pappu, and Eric Mjolsness, "New algorithms for 2D and 3D point matching: pose estimation and correspondence," *Pattern Recognition*, vol. 31, no. 8, pp. 1019 – 1031, 1998.

[7] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Fast predictive multimodal image registration," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 858–862.

[8] Zhenlin Xu and Marc Niethammer, "DeepAtlas: Joint semi-supervised learning of image registration and segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. 2019, pp. 420–429, Springer International Publishing.

[9] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, Feb 2001.

[10] L. Zhang, X. Wu, and Z. Sheng, "A fast image segmentation approach based on level set method," in *2006 8th international Conference on Signal Processing*, Nov 2006, vol. 2.

[11] Shusil Dangi, Cristian A. Linte, and Ziv Yaniv, "A distance map regularized CNN for cardiac cine MR image segmentation," *Medical Physics*, vol. 46, no. 12, pp. 5637–5651, 2019.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2015*. 2015, pp. 234–241, Springer International Publishing.

[13] Hans Meine, Grzegorz Chlebus, Mohsen Ghafoorian, Itaru Endo, and Andrea Schenk, "Comparison of U-net-based convolutional neural networks for liver segmentation in CT," *arXiv e-prints*, Oct 2018.

[14] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv e-prints*, Feb 2015.

[15] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2017, pp. 240–248, Springer International Publishing.

[16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[17] Xenophon Papademetris, Marcel Jackowski, Nallakkandi Rajeevan, R Todd Constable, and LH Staib, "Bioimage suite: An integrated medical image analysis suite," *The Insight Journal*, vol. 1, no. 3, pp. 13, 2005.