ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

Segmentation Performance Prediction of Medical Images using Reverse Classification Accuracy Method in the Absence of ground Truth

Varsha E. Jaware [1], Rajesh H. Kulkarni [2]

P.G. Student, Department of Computer Engineering, JSPM Narhe Technical Campus, Pune, Maharashtra, India[1]

Associate Professor, Department of Computer Engineering, JSPM Narhe Technical Campus, Pune, Maharashtra, India[2]

ABSTRACT: Image segmentation is a pre-requisite to medical image analysis. Medical image segmentation has long been recognised as a difficult problem. In Medical images, accurate segmentation is a key step in contouring during radiotherapy planning. Magnetic Resonance Imaging (MRI) are the most widely used radiographic techniques in diagnosis, clinical studies and treatment planning. In this paper, the technique incorporated is a Reverse Accuracy Classification (RCA) framework that predicts the performance of a segmentation method on new data. RCA takes the predicted segmentation from a new image to train a reverse classifier that is evaluated to a set of reference images with available Ground Truth. The ideal concept is that to rank the best segmentation results in the database without making the manual label. Then match the rank between the predictions of the truth saved in database. Rather than going in a traditional way of classification, a reverse way of testing is followed which lead to the prediction of segmentation quality in the absence of Ground Truth.

KEYWORDS: Machine Learning, image Segmentation, Ground Truth, classification, performance evaluation

I. INTRODUCTION

Image Segmentation is an important factor for analysis which extract clinical useful information from the segmented images. Image segmentation is the process of partitioning a digital image into multiple segments which is typically used to locate objects and boundaries (lines, curves, etc.) in images. The field of medical imaging has undergone revolutionary advances over of the past 2 decades. New medical imaging technologies have provided physician's powerful, non-invasive techniques to probe the structure, operate and pathology of the physique. Segmentation among the particles of the varied components is extremely vital to medical call. Despite of the significant advancement in this field, medical image segmentation remains an unresolved problem.

With increasing use of Magnetic Resonance (MR) imaging and Computed Topography (CT) scan for diagnosis, treatment planning and clinical studies, it has become almost compulsory to use computers to assist radiological experts in clinical diagnosis and treatment planning. Machine Learning classification has been the most important computational development in the last years to satisfy the primary need of clinicians for automatic early diagnosis. A two-class classification model for grouping and linear classifier to combine these features was proposed in [14]. To demonstrate the power of the classification model, a simple algorithm was used to randomly search for good segmentations. In cases of new data, it became very difficult to predict the real performance of a particular segmentation method when no reference image is available. RCA framework method was used to predict segmentation accuracy in such cases of absence of GT. Objective of this study is to find out how RCA model aims to answer the question on how to estimate the performance of models in cases where ground truth is not available [1].

Characterizing the performance of image segmentation approaches has been a persistent challenge.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

Performance analysis plays a vital role since segmentation algorithms often have limited accuracy and precision. Interactive drawing of the desired segmentation by human raters has always been an acceptable approach, and yet suffers from intra-rater and inter-rater variability. Automated algorithms have been sought in order to remove the variability introduced by raters, but such algorithms must be assessed to check their endurance whether they are suitable for the task or not. The performance of raters (human or algorithmic) generating segmentations of network or individual battery energy of a node [1]. Medical images has been difficult to quantify because of the difficulty of obtaining or estimating a known true segmentation for clinical data. An example of segmented image is shown in Fig.1

Fig.1 A Segmented image of Femur Bone

The trend towards large-scale studies including population imaging poses a new challenge in terms of quality control. This is a particular issue when automatic processing tool such as Image segmentation methods are employed to derive quantitative measures or bio markers for further analyses. However it is important to detect when an automatic method fails to avoid inclusion of wrong results which further lead to incorrect conclusions. In order to overcome this challenge, reverse classification accuracy approach is used for predicting segmentation quality. Traditionally segmentation performance was evaluated by selection of a random data set and obtaining a manual expert segmentation for it which was then compared to the automatic one. In order to solve segmentation problems, different methods were proposed based on statistical models [2], multi-atlas label propagation [3] and supervised classification [4].

On deploying a segmentation method there is hardly any idea to predict its real performance. Assessing the performance using traditional evaluation measures is a challenging problem when no GT is available for comparison. Perhaps, it is difficult to assess the accuracy level on clinical data [15] and thus find out the failure of automatic segmentation method. Particularly when segmentation is an intermediate step in large scale automated analysis where no visual quality control is present. This plays a vital role in large scale studies like the UK Biobank Imaging Study [16] where automated methods are applied on thousands of images, and segmentation is used for further statistical population analysis.

II. RELATED WORK

In many domains from graphics, remote sensing to marketing strategies, it has become difficult to retrieve an objective performance evaluation without GT. The lack-of-label problem has been tackled by exploiting transfer learning in [5] using a reverse validation procedure when the number of labeled data is limited in the general machine learning domain. The basic idea of reverse validation [5] is based on reverse testing [6] new proposed classifier is initially trained on prediction of the test data and later evaluated on Training data. This thought of reverse testing is closely related to the approach of RCA in [1].

The segmentation performance is evaluated by several tasks like separating the perceptual salient structures [7], automatically generating semantic GT [8], [9] or by observing at contextual properties [10].A method for

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

analyzing objective metrics like precision and recall with no GT is proposed but it failed with data sets that had partial GT as probabilistic model approach is used.

System proposed by [11] used region-correlation matrix to quantify the performance of various segmentation algorithms. The drawbacks found with this system was that there was no independence for evaluation of the segmentation performance considering a particular method for an image.

Meta-evaluation framework, which is a standalone method where image features are used to provide ranking of different methods [17] used in a machine learning setting that provides a ranking of different methods. Standalone methods have the advantage that they do not require a manually segmented reference image for comparison and can therefore be used for real-time evaluation. But it faced a problem with the estimation of segmentation performance of an individual image. An effort to reckon objective metrics like precision and recall with missing GT is proposed by [18]. The drawback with it was that it could be used for data sets with partial GT since under the same assumptions it applies a probabilistic model.

Unsupervised methods in [19], [20] aim to obtain the segmentation accuracy directly from the images and label maps like geometrical and information-theoretic features. It enables the quantification of the quality of an image segmentation result. These evaluation criteria reckon some statistics for each region or class in a segmentation result. Such an evaluation criterion can be useful for different applications like the comparison of segmentation results, the automatic choice of the best fitted parameters of a segmentation method for a given image, or the definition of new segmentation methods by optimization. But the application in medical setting is unclear as unsupervised methods are applied to scenarios where the main motto of segmentation is to yield visually consistent results that are meaningful to a human observer.

Among all ensembles approaches Random Forest (RF) [11] produced the best accuracies in many scientific fields. Random forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification and the tree "votes "for that class. The forest chooses the classification having the most votes (over all the trees in the forest). After time δt it calls the optimization function to select the path and send RREP. Optimization function uses the individual node's battery energy; if node is having low energy level then optimization function will not use that node.

III. SYSTEM OVERVIEW

A system framework aimed to assess the performance of the segmented images in a reverse manner has helped to solve cases where there is no reference available to compare with. RCA in [1] is applied to estimate the segmentation quality by using one of the specific segmentation method like Multi Atlas, Single Atlas or Random Forest. RCA model was applied with Single Atlas, Constrained CNN and Atlas Forest. But among those algorithm Constrained CNN failed to give a high accuracy as compared to Single Atlas and Atlas Forest. This model worked well on large body organs like brain, stomach, liver, pelvis and but estimated less accuracy for smaller organs like spleen, adrenal glands and clavicles. System has two phases, mainly segmentation phase and then comparing it with the reference database for predicting the accuracy in a reverse manner. The supposition mentions that the RCA classifier which is trained using predicted segmentation that acts as pseudo GT, worked well if the segmentation quality was high for a new image. Similarly, if the segmentation quality revealed to be poor then it would poorly perform on the reference images. Considering the scenario that in cases of absence of GT images, it becomes difficult to analyze the segmentation performance. It is also necessary to check when the automatic segmentation method fails RCA has the advantage of allowing to predict the accuracy for each individual case and also allows drawing conclusions for the overall performance of a particular segmentation method.

RCA Model:

The RCA framework is based on the thought of training reverse classifiers on individual images utilizing their predicted segmentation as pseudo GT. In this system three different methods for RCA classifier along with combination of different segmentation methods. Fig.2 shows the detailed architecture of the RCA framework.

Consider the segmentation result of image I, called Si. A segmentation Si is obtained, by training a chosen algorithm with reference database (for which GT is available). Unfortunately for this image no GT is available, thus there is no real accuracy. Hence RCA classifier is used for that. RCA classifier is trained using the result of

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

segmentation Si. Then test back to the reference database to obtain the segmentation. From those segmentation in the reference database, since GT is present calculate the accuracies. The final prediction accuracy of Si by RCA, is the maximum accuracy of the segmentation present among the reference database.

Segmentation phase: In this step, segmentation of MRI images of organs like brain, liver, lung, etc. is done using Single Atlas, Multi Atlas and Convolutional Neural Network (CNN).

RCA Model: The input received from segmentation phase is then applied to RCA model for classification and predicting its performance by comparing it with the reference database. **Reference Database:** For storing images two partitions of database named as Training and Testing database. Absence of GT is only for Testing database (which we wish to predict using the RCA). To get the segmentation on testing database, reference (Training) database to train a classifier. This training database has GT, which is used to train and segment testing database and also to predict the segmentation accuracy using improved RCA model. Fig. 2 describes the architecture of the RCA model.

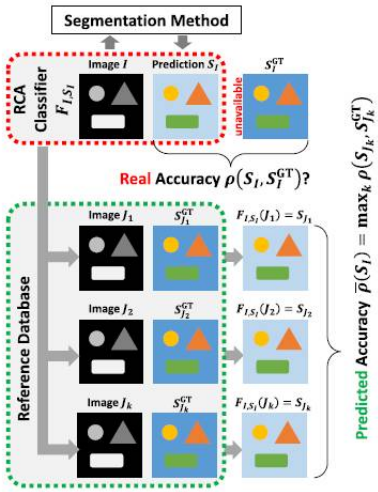

Fig.2 Architecture of the System

IV. PROPOSED ALGORITHM

A. Classifier Algorithms:

Random Forest, Multi Atlas and CNN method are used for Segmentation Phase. RCA model includes Single Atlas, Atlas Forest and Constrained CNN methods for classification.

a) Atlas Forests:

RCA classifier is based on the recently introduced concept of Atlas Forests (AFs) which demonstrates the feasibility of using Random Forests (RFs) [11] to encode individual atlases, i.e., images with corresponding segmentations. For general image segmentation tasks Random Forests have become popular and perform a vital role as they naturally handle multi-class problems and are computationally efficient. They do not (necessarily) require preregistration of the images neither at training nor testing time, since they operate as voxel-wise classifiers. Registering location probability maps to each atlas and new image, is not a general requirement for using AFs to encode atlases. In fact, the way AFs are employed within RCA framework does not require any image registration. The forest-based RCA classifiers were trained all with the same set of parameters of maximum depth 30 and 50 trees. Atlas Forests give high accuracy in this work for RCA model.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

b) Deep Learning:

CNNs is also used as one of the RCA classifiers. A Deep Medic is utilized for a 3D CNN architecture for automatic segmentation [21]. The architecture is computationally efficient as it can handle large image context by using a dual pathway for multi-scale processing. CNNs are able to learn highly complex and discriminative data associations between input data and target output. The architecture of the network is defined by the number of layers and the number of activation functions in each layer. In CNNs, each activation function corresponds to a learned convolutional filter, and each filter produces a feature map (FM) by convolving the outputs of the previous layer. Through the sequential application of many convolutions, highly complex features are learned that are then used to produce voxel-wise predictions at the final, fully-connected layer. CNN shows less accuracy as compared to the other Atlas forests and Random Forests.

c) Atlas-based label propagation:

The third approach considered is atlas-based label propagation. Label propagation using multiple atlases are used to on many segmentation problems [3]. A common procedure in multi-atlas methods is to use non-rigid registration to align the atlases with the image to be segmented and then perform label fusion strategies to obtain predictions for each image point. Multi-atlas methods are based on registration but are not strictly voxel-wise classifiers as they operate on the whole image during registration, the final stage of label fusion can be considered as a voxel wise classification step. An approach that has been originally developed in the context of segmentation of cardiac MRI [22] is used. For the purpose of RCA, however, there is only a single atlas and thus no label fusion is required.

d) Random Forest:

Random Forests grows many classification trees. Each tree is grown as follows:

I) If the number of cases in the training set is N, sample N cases at random but with replacement, from the original data. This sample will be the training set for growing the tree.

II) If there are M input variables, a number m is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.

III) Each tree is grown to the largest extent possible. There is no pruning.

B. Mathematical Model:

Equation of a proxy measure for predicting the segmentation accuracy is as follows, where ρ is any evaluation metric, such as DSC, ASD or HD

*SI denotes the predicted segmentation that here acts as pseudo GT

*I be an image that has been segmented by any segmentation method

*Segmentation Function FI, SI (J) = SJ

*Image J which produces a segmentation SJ

$$\rho.\ (SI\ ) = \max 1 \leq k \leq m\ \rho(FI, SI\ (Jk\ ),\ SGT\ Jk\ )$$

V. SIMULATION RESULTS

A. *Performance Metrics for Prediction Accuracy:*

The Dices similarity coefficient is the most widely used measure for evaluating segmentation performance , and the main results in this study focused on how well the DSC, predicted good results using RCA framework. In order to quantify prediction accuracy, three different measures, namely the correlation between predicted and real DSC, the mean absolute error (MAE), and a classification accuracy were incorporated in this model. Arguably, the most important measure for direct evaluation of how well RCA works is the MAE directly reveals how close the predicted DSC is to the real one. Correlation is of great importance, as it conveyed a relation between predicted and real scores. Segmentation Failure Detection in clinical measures it is of great importance to be able to detect when an automated method fails. Low real DSC scores if obtained are correctly predicted and failed segmentations are

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

identified from the score. It is important to exclude failed segmentations from the subsequent analysis as it leads to wrong results or conclusions.

B. Summary:

An appealing property of the proposed framework is that unlike the supervised methods no training data is required that captures examples of good and bad segmentations. Instead, RCA simply relies on the availability of a reference database with available GT segmentations. The drawback, however, is that assumption of a linear relationship between predicted and real scores which should be close to an identity mapping, something only found in the case of using Single-Atlas label propagation. In the case of off-diagonal correlation, as for example found for Atlas Forests, an extension to RCA could be considered where the predictions are calibrated. This, however, requires training data from which a regression function could be learned. CNN show a very less accuracy level for classification as compared to Atlas forest and Single Atlas.

Table No. 1 Detecting Segmentation Failure Using Random Forest

RCA Classifier	Correlation	MAE	Accuracy
Random Forest (proposed)	89%	9%	95%
Atlas Forest (existing)	85%	9.6%	88.4%
Single Atlas (existing)	87%	9.7%	92.8%

VI. CONCLUSION AND FUTURE SCOPE

A survey of the earlier related work has been done successfully which involves a unique method for classifying images in a reverse way. The initial studies showed that RCA model worked well with Atlas Forest and Single Atlas but not with CNN. RCA has advantages in the absence of Ground Truth for classifying correct segmented results and find the best match for the new image as it is ideal for integration into automatic processing pipelines in clinical routine. In large scale studies it is important to detect failed segmentations as it is not feasible to employ manual quality control with visual inspection. RCA helps as an effective tool to automatically extract the subset of high quality segmentations and also detect failure cases. In future RCA can improve on distance based measures and also find correct accuracy for small organs like spleen, clavicles etc.

REFERENCES

1. Vanya V. Valindria, Ioannis Lavdas, Wenjia Bai, Konstantinos Kamnitsas, Eric O. Aboagye,Andrea G. Rockall, Daniel Rueckert, and Ben Glocker,*"Reverse Classificaton Accuracy: Predicting Segmentation Performance in the absence of Ground Truth,"* IEEE Transactions on Medical Imaging, Vol. 36, No. 8, 2017.
2. T. Heimann and H.-P. Meinzer, *"Statistical shape models for 3D medical image segmentation: A review"* Med. Image Anal., vol. 13, no. 4, pp. 543-563, 2009.
3. J. E. Iglesias and M. R. Sabuncu, *"Multi-atlas segmentation of biomedical images: A survey"* Med. Image Anal., vol. 24, no. 1, pp. 205-219,2015.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

4. E. Geremia et al., *"Classification forests for semantic segmentation of brain lesions in multi-channel MRI"* in Decision Forests for Computer Vision and Medical Image Analysis. London, U.K.: Springer, pp. 245260, 2013.
5. E. Zhong, W. Fan, Q. Yang, O. Verscheure, and J. Ren, *"Cross validation framework to choose amongst models and datasets for transfer learning"* in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, pp. 547-562, 2010.
6. W. Fan and I. Davidson, *"Reverse testing: An efficient framework to select amongst classifiers under sample selection bias,"* in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 147-156, 2006.
7. F. Ge, S. Wang, and T. Liu, "New benchmark for image segmentation evaluation," J. Electron. Image. vol. 16, no. 3, pp. 033011, 2007.
8. L. Goldmann et al., *"Towards fully automatic image segmentation evaluation,"* in Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst., pp. 566-577, 2008.
9. H. Li, J. Cai, T. N. A. Nguyen, and J. Zheng, *"A benchmark for semantic image segmentation"*,in Proc. IEEE Int. Conf. Multimedia Expo, pp. 16,2013.
10. K. Sikka and T. M. Deserno*"Comparison of algorithms for ultrasound image segmentation without ground truth*", Proc. SPIE, vol. 7627, pp. 76271C176271C9, 2010.
11. Leo Breimen, *"Random Forest"*, 2001.
12. K.Yogeswara Rao, M.James Stephen, D.Siva Phanindra,*"Classification Based Image Segmentation Approach"*, IJCST, Vol 3, Issue 1, 2012.
13. Xiaofeng Ren and Jitendra Malik,*"Learning A Classification Model for Segmentation"*.
14. Dr Stefan Teipel, Michel J Grothe, , Henryk Barthel,*"Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection "*,Lanclet Neurology,2015.
15. B. van Ginnekan, T. Heimann, M.Styner,*"3D Segmentation in the clinic: A Grand Challenge "*, in proc. MICCAI Workshop 3D Segmentation Clinic, pp. 7-15, 2007.
16. C. Sudlow et al.,*"UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age "*, PLoS Med., vol. 12, no. 3, p.e1001779, 2015.
17. H. Zhang, S. Cholleti, S.A. Goldman, J.E.Fritts,*"Meta-evaluation of Image Segmentation using Machine Learning "*,in Proc. IEEEE Comput. Soc. Conf. Compute. Vis. Pattern REcognit. (CVPR). vol. 1, pp. 1138-1145,2006.
18. B.Lamiroy and T. Sun,*"Computing Precision and Recall with Missing or Uncertain Ground Truth"*, in Graphics Recognition. New Trends and Challenges. Berlin, Germany: Springer- Verlag, pp. 149-162,2013
19. S. Chabrier, B. Emile, C. Rosenberger, H. Laurent,*"Unsupervised Performance Evaluation of Image Segmentataion"*, EURASIP J. Appl.Signal Process., vol. 2006, pp. 217-229, 2006.
20. H. Zhang, J. E. Fritts, and S. A. Goldman, *"Image Segmentation Evaluation: A survey of Unsupervised Methods"*,Computer Vision and Image Understanding , vol. 110, no. 2, pp. 260-280, 2008.
21. K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, B. Glocker, *"Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation,"* Medical Image Analysis, 2016.
22. W. Bai, W. Shi, D. P. ORegan, T. Tong, H. Wang, S. Jamil- Copley, N. S. Peters, and D. Rueckert,*"A Probabilistic Patch-Based Label Fusion Model for Multi-Atlas Segmentation with Registration Refinement: Application to Cardiac MR Images"*, IEEE Transactions on Medical Imaging, vol. 32,no. 7, pp. 1302-1315, 2013.