RESEARCH STATEMENT

Duygu Sarikaya (duygusar@buffalo.edu)

My research interests span computer vision methods for medical imaging problems. While my recent research uses advances of deep learning with a multi-task and multimodal approach in robot-assisted surgery (RAS) video understanding, my previous research has mainly revolved around exploring fusion models in automatic segmentation of brain magnetic resonance imaging (MRI) relating the tumor and anatomical structures of the brain.

1 Research Background

Automatic understanding of medical images has been an active research area, and universally-accepted, standard solutions are high in demand. The recent advances of computer vision and machine learning algorithms coupled with the use of digital imaging modalities facilitating efficient storage and access, has led to the improvements in automatic understanding of medical imaging data [1]. Scalable and efficient computer vision algorithms are used in medicine for a broad range of open research problems such as diagnostics, image-guided therapy, automation, and augmented reality for surgical planning and navigation. Intelligent designs and algorithms answer to the needs of medicine by addressing the time, cost and expertise concerns while hoping to create objective, universally-accepted, standardized, and validated metrics for medical imaging [2].

Many recent studies have shown that exploiting the relationship across different tasks, jointly reasoning multi-tasks [3, 4, 5], and taking advantage of a combination of shared and task-specific representations [6] perform astonishingly better than their single-task counterparts. Data fusion and use of multiple medical imaging modalities has gained significance in clinical assessment; diagnostics and treatment planning, following the lower costs and increased quality of medical imaging. Medical image fusion include a broad range of techniques from image registration and fusion, to fusion at decision level. The use of multi-modal medical images or fusion of other sources of information offer a greater diversity of the features that can exploit the underlying features and a prior knowledge that might have gone unnoticed otherwise [7].

My research facilitates information fusion whether it is multiple modalities combining shared and task-specific representations, joint relationships across different tasks, or using smart fusion of a priori knowledge inferred from multiple atlases. My recent research focuses on the specific problems of robot-assisted surgery (RAS) video understanding, more specifically, automatically detecting and localizing the tools, and action recognition of surgical gestures and tasks in RAS videos. My previous research addresses the problem of automatic segmentation of brain magnetic resonance imaging (MRI) relating the tumor and anatomical structures of the brain. My research shows that using an appropriate fusion approach increases the accuracy of these tasks and performs better than approaches without fusion. Therefore, fusion models show great promise in the medical domain.

2 Recent Research

Robot-assisted surgery (RAS) is the latest form of development in today's minimally invasive surgical technology. The robotic tools help the surgeons complete complex motion tasks during procedures with ease. Despite its advances in minimally invasive surgery, the steep learning curve of the robot-assisted surgery devices remains a disadvantage [8]. The need for universally-accepted and validated metrics and quantitative skill assessment via automation is addressed in the community [2]. Early identification of technical competence in surgical skills is expected to help tailor training to personalized needs of surgeons in training [2, 9]. The insights drawn may be applied in effective skill acquisition, objective skill assessment, real-time feedback, and human-robot collaborative surgeries.

I evaluate the problem with a computer vision approach and mainly focus on video understanding of surgical videos. The main questions I am trying to solve with my research are: How do we understand the surgical setting and how do we model the surgeon's actions? In order to model the actions of the surgeons, I look into the sub-problems of detecting and localizing the tools in the video and recognizing what actions are taking place by using temporal information as well as image cues.

ATLAS Dione Dataset for robot-assisted surgery (RAS) video understanding

Along with my research, I have prepared ATLAS Dione dataset and made it publicly available to encourage further research on RAS video understanding [10]. ATLAS Dione provides video data (86 full subject study videos, 910 action clips with a total of 5 hours) of ten surgeons from Roswell Park Cancer Institute (RPCI) (Buffalo, NY) performing six different surgical tasks on the daVinci Surgical System (dVSS) with annotations of robotic tools per frame (for a subset of 99 action clips), actions taking place and their timestamps. It also provides information on the surgeon expertise levels based on the Dreyfus model [11]. It addresses the need in the community as public datasets for robot-assisted surgery (RAS) video understanding are limited. ATLAS Dione includes artefacts, camera movement and zoom, and a wide range of free movement. You can access this dataset at `www.roswellpark.edu/education/atlas-program/dione-dataset`

Detection and Localization of Robotic Tools in Robot-Assisted Surgery (RAS) Videos Using Deep Neural Networks for Region Proposal and Detection

With this research, I propose a solution to the tool detection and localization open problem in robot-assisted surgery (RAS) video understanding, using a strictly computer vision approach and the recent advances of deep learning. I propose an architecture using multimodal convolutional neural networks for fast detection and localization of tools in RAS videos [10]. To my knowledge, this approach is the first to incorporate deep neural networks for tool detection and localization in RAS videos. The architecture applies a Region Proposal Network (RPN), and a multimodal two stream convolutional network for object detection, to jointly predict objectness and localization on a fusion of image and temporal motion cues. My approach differs from most state-of-the-art solutions that require either kinematic data captured with additional tools or manual initialization of tools for tracking in surgical videos. The system I propose is strictly computer vision based and fully automated. The results with an Average Precision (AP) of 91% and a mean computation time of 0.1 seconds per test frame detection indicate that the architecture is superior to conventionally used methods for medical imaging while also emphasizing the benefits of using region proposal networks (RPN) for precision and efficiency for localization of the tools.

Long-term Recurrent Convolutional Networks for Robot Assisted Surgery (RAS) Video Understanding

I model the activities taking place during surgical tasks and address the problem of activity recognition in robot-assisted surgery videos. I propose a multimodal novel architecture based on long-term recurrent neural networks (LSTM). My architecture learns temporal dynamics along with shared and task-specific representations. Early experiments show that my joint model reaches an Average Precision (AP) with a 7% improvement compared to state-of-the-art methods. More information about this research will be provided upon publication.

Segmentation of the Anatomical Structures of Magnetic Resonance Human Brain Images

I propose an algorithm that automatically segments the anatomical structures of magnetic resonance human brain images (MRI) [12, 13]. By propagating a priori knowledge we gather from multi-atlases, we are able to quantitatively investigate the spatial and morphometric correspondences of brain structures. The common approaches for label fusion of multi-atlases mostly rely only on voting, however I use an intensity similarity based weighted voting. This information fusion step helps me to use the intensity similarity information of corresponding pixels as well as the spatial correspondence on the graph that is built with the voxels of the brain. Finally I cluster the graph using multiway cut for 3D segmentation of the subject image.

Automatic Brain Tumor Segmentation with Markov Random Field (MRF) on Supervoxels

In this research, I address the problem of segmentation of tumor in brain magnetic resonance images (MRI) [14, 15]. I define 3D joint histograms that are representative of each subject MRI in three different modalities. Then, I perform oversegmentation on each subject MRI and define a Markov random field (MRF) on the supervoxels using the probability distributions over the voxels as the unary term and the edge cues of shared boundaries of two supervoxels to define the binary term. The likelihood model on the intensities is based on histogram matching and fusion of this a priori knowledge in a specific way relating the tumor and brain structures. This likelihood model is the key to my research contribution. Inference is ultimately performed with graph cuts.

3 Future Research

My future research plans revolve mainly around developing novel computer vision and machine learning techniques with a medical image computing focus. I anticipate that the advances in deep learning, fusion models, multi-modality, multi-task learning and metric learning will help with medical image computing problems greatly, and are worth contemplating upon. I believe applications of this kind of research will be diverse in near future. Both my research findings so far and the recent advances in the community show that these topics have true potential and shows great promise in medical image computing domain. My agenda is not limited to developing more accurate and customized solutions to medical image computing problems, but encompasses a concern for developing novel computer vision approaches that are generalizable across different computer vision and machine learning domains as well.

References

[1] C. H. Chen, “Computer Vision in Medical Imaging”, *World Scientific Publishing Co., Inc.*, River Edge, NJ, USA, 2013.

[2] S. Kumar, N. Ahmidi, G. Hager, P. Singhal, J. J. Corso and V. Krovi, “Surgical performance assessment”, in *ASME Dynamics Systems and Control Magazine*, vol. 3, no. 3, pp. 7-10, 2015.

[3] I. Kokkinos, “UberNet: Training a 'Universal' Convolutional Neural Network for Low-, Mid-, and High-Level Vision using Diverse Datasets and Limited Memory”, *Computing Research Repository (CoRR)*, abs. 1609.02132, 2016.

[4] M. Long and J. Wang, “Learning Multiple Tasks with Deep Relationship Networks”, *Computing Research Repository (CoRR)*, abs. 1506.02117, 2015.

[5] J. Yao, S. Fidler, and R. Urtasun, “Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation”, *Proc. Computer Vision and Pattern Recognition (CVPR)*, June.2012.

[6] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, “Cross-stitch Networks for Multi-Task Learning”, in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] J. A. Pappachen and B. V. Dasarathy, “Medical image fusion: a survey of the state of the art”, *Inf. Fusion*, vol.19, pp. 4-19, September. 2014.

[8] M. Meadows. “Robots lend a helping hand to surgeons”, in *U.S. Food and Drug Administration (FDA) Consumer Magazine*, vol. 36, no. 3, pp. 10-15, May/Jun. 2002.

[9] K. A. Guru *et al.*, “Cognitive skills assessment during robot-assisted surgery: separating wheat from chaff”, in *British Journal of Urology (BJU) International*, vol. 115, no.1, pp.166-174, Jan. 2014.

[10] D. Sarikaya, J. J. Corso, K. A. Guru, “Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection”, *IEEE Transactions on Medical Imaging*, February. 2017.

[11] S. E. Dreyfus and H.L. Dreyfus, “A five-stage model of the mental activities involved in directed skill acquisition”, Operations Research Center, Cambridge, MA, USA, Tech. Rep., 1980.

[12] D. Sarikaya, L. Zhao, and J. J. Corso, “Multi-atlas brain MRI segmentation with multiway cut”, in *Proc. MICCAI Workshops: The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS13), 2013*

[13] A.M. Mendrik *et al.*, “MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans”, *Computational Intelligence and Neuroscience, special issue on Simulation and Validation in Brain Image Analysis*, Article ID 813696, 2015.

[14] L. Zhao, D. Sarikaya, and J. J. Corso, “Automatic brain tumor segmentation with MRF on supervoxels”, in *Proc. MICCAI Workshops: The Multimodal Brain Tumor Segmentation Challenge (BRATS), 2013*

[15] B. H. Menze *et al.*, “The multimodal brain tumor image segmentation benchmark (BRATS)”, *IEEE Transactions on Medical Imaging*, October, 2015.