Ultrasound Guided Pedicle Screw Entry Point Identification for Spinal Fusion Surgery

Xiao Qi[1], Michael Vives[3], and Ilker Hacihaliloglu [2,4]

[1] Department of Electrical and Computer Engineering, Rutgers University, USA
[2] Department of Biomedical Engineering, Rutgers University, USA
[3] Department of Orthopaedics, Rutgers New Jersey Medical School, USA
[4] Department of Radiology, Rutgers Robert Wood Johnson Medical School, USA
`xiao.qi@rutgers.edu`   `ilker.hac@soe.rutgers.edu`

Abstract

Accurate identification of the location the vertebra and corresponding pedicle is critical during pedicle screw insertion for percutaneous spinal fusion surgery. Currently, two dimensional (2D) fluoroscopy based navigation systems have extensive usage in spinal fusion surgery. Relying on 2D projection images for screw guidance results in high misplacement rates. Furthermore, fluoroscopy-based guidance exposes the surgical staff and patient to harmful ionizing radiation. Real-time non-radiation-based ultrasound (US) is a potential alternative to intra-operative fluoroscopy. However, accurate interpretation of noisy US data and manual operation of the transducer during data collection remains a challenge. In this work we investigate the potential of using multi-modal deep convolutional neural network (CNN) architectures for fully automatic identification of vertebra level and pedicle from US data. Our proposed network achieves 93.54% vertebra identification accuracy on in vivo US data collected from 27 subjects.

1 Introduction

Due to the proximity of many important neural and vascular structures, accurate screw placement is important during spinal fusion surgery. Misplacement will impact fusion rate and even cause adjacent level disease (Manbachi, et al., 2014). Percutaneous pedicle screw insertion (PPSI) procedures, based on Computer assisted orthopedic surgery (CAOS) systems, have gained widespread acceptance due to advantages, such as reduced length of stay, reduced blood loss and improved accuracy(Mobbs, et al., 2011). One of the main disadvantage of CAOS is the exposure to radiation since the main intra-operative imaging modality is 2D/3D fluoroscopy.

Due to its real-time, 2D/3D, and non-radiation-based imaging capabilities, ultrasound (US) has been investigated as an alternative imaging modality in various CAOS procedures (Hacihaliloglu, 2017a). Despite promising results, US has not become the standard-of-care in CAOS. One of the biggest

---

[1]   Corresponding author

challenges of incorporating US into CAOS procedures is related to the manual operation of the US transducer during data collection. In order to overcome this problem, last year our group has presented an approach, based on convolutional neural networks (CNNs), for automatic identification of the pedicle scan plane from sagittal in vivo US data (Qi, et al., 2018). However, automatic identification of vertebra level is also important since the optimal entry point for pedicle screw is located at intersection of transverse and sagittal scan plane (Oh, et al., 2013).

With these in mind, we propose a deep learning framework for both sagittal and transverse scan plane classification for automatic and accurate identification of pedicle screw entry point from real-time US data. The proposed system, through utilizing a combination multiple features and a novel multimodal CNNs architecture, achieves accurate and robust results overcoming the disadvantage of US images not corresponding to any anatomical structure, and facilitates surgeon to select the correct scan plane during the procedure. We validate our method on scans collected from 27 subjects.

2   Materials and Methods

2.1   Data Collection

After obtaining approval from the Rutgers University institutional review board (IRB), US Lumbar spine images were captured from 27 subjects with written consent. For training, 313 sagittal scan US images, using longitudinal paramedian sagittal scanning approach, and 433 transverse US images, using transverse scanning approach, were collected from 18 subjects. Sagittal scans included 130 images of articular process (AP) view, 91 images of transverse process (TP) view, 92 images of laminar view. Transverse scans included 195 images of pedicle view, and 238 images of vertebral bone view. In order to reduce over-fitting and to make pre-trained CNNs more robust for identification task, data augmentation was performed on the training dataset. Thus, the size of training dataset is increased by 10 times (7,460 total scans). Nine additional new subjects were scanned in order to collect test dataset. From these nine subjects we collected 66 AP scan plane view images, 66 laminar view images, 39 TP view images, 126 pedicle view images, and 122 vertebral bone view images.

The AP view scans, during longitudinal paramedian sagittal scanning approach, covers parts of the pars interarticularis and the inferior point of the superior articular facet. The transverse interlaminar view scans cover the top area of pedicle, referred as pedicle view scans. Therefore, AP and pedicle view scans can be used as reference for placement of pedicle screw from sagittal view and transverse view separately1 (Oh, et al., 2013). Our methods, explained in the below sections, are geared towards the classification of these two scans.

2.2   Image Enhancement

As a first step in our proposed work, we enhance the collected US data, denoted as *US(x,y)*, by extracting image local phase information using band-pass quadrature filters (Hacihaliloglu, 2017). Real-time feedback of acoustic bone shadow information can guide the orthopedic surgeon to a standardized diagnostic viewing plane with minimal artifacts and can provide additional information for scan plane localization. Accordingly, the *LP(x,y)* image is used for the enhancement of bone shadow regions (Hacihaliloglu, 2017b). Figure 1 displays the extracted *LP(x,y)* and multi feature image (MF(x,y)) obtained by combining US(x,y), LP(x,y) and BSE(x,y) images as three channel input.
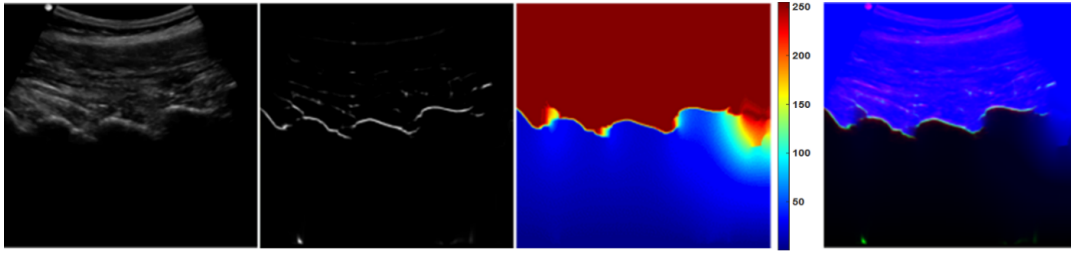
Figure 1: From left to right: B Mode US image (*US(x,y)*), local phase image (*LP(x,y)*), bone shadow image (*BSE(x,y)*), and multi feature image (*MF(x,y)*) obtained by combining *US(x,y), LP(x,y)* and *BSE(x,y)* images as three channel input.

Multi-modal deep learning algorithms have shown that cross modality feature learning results in the improvement of the learning task (Ngiam, et al., 2011). Consequently, we design a new multi-modal CNN architecture for processing the enhanced images, *LP(x,y)* and *BSE(x,y)*, together with the B-mode US data, *US(x,y)*. In order to show the improvements achieved, by using multi-modality deep learning, we investigate mono-modality CNN architectures as well.

2.3 Scan Plane Identification Using Transfer Learning and Mono-modality CNNs

Transfer learning makes use of pre-trained CNNs to solve identification tasks in a new domain with a small dataset (Oquab, et al., 2014). Therefore, during this work we apply transfer learning from pre-trained CNNs to the task of paramedian sagittal and transverse spinal US image scan plane identification. We use AlexNet (Krizhevsky, et al., 2012) and VGG-11 (Russakovsky, et al., 2014) pre-trained networks. During transfer learning the last layers (freeze first two convolution layers) of these networks were re-trained using our collected US data. In order to investigate the improvements on classification, we also trained each network by using *LP(x,y)*, *BSE(x,y)*, and *MF(x,y)* images. Finally, we investigate the performance of the networks when all the layers are re-trained using our collected data.

2.4 Scan Plane Identification Using Multi-modal CNNs

Our proposed multi-modal CNNs architecture consists of two same convolutional network streams for processing *US(x,y)* images and the corresponding enhanced images (*LP(x,y), BSE(x,y)*), or multi-feature image (*MF(x,y)*) respectively (Fig.2-a). Once the two individual stream networks are trained, the image features from these two trained individual architectures are concatenated together by a late fusion approach. During this work we investigated the following combinations as multi-modal data: (1) *US(x,y)+LP(x,y)*, (2) *US(x,y)+BSE(x,y)*, (3) *US(x,y)+MF(x,y)*. Figure 2-a shows an example where two modified Alexnet architectures are joint together with a late fusion operation (FusCNN1). We have also constructed a second multi-modal CNN by using a network structure similar to VGG-11 (FusCNN2).

2.5 Training

The collected dataset is set up for 5-fold cross-validation by leave-one-out for each class of images. Furthermore, all the collect test datasets were also used for testing. We also made sure that during the 5-fold cross-validation of the dataset the training and testing data did not include the same patient scans.

3 Results

For all the investigated networks the highest identification accuracy was obtained when using the *MF(x,y)* images which was obtained by combining *US(x,y)*, *LP(x,y)* and *BSE(x,y)* images as three channel input. Overall our proposed FusCNN1 architecture using *US(x,y)* and *MF(x,y)* as multi-modal data achieved the highest accuracy (Fig 2 b-c). The identification accuracy of VGG-11 for each type of image is significantly higher than the identification accuracy of Alexnet for same image type (paired t-test $p<0.005$) (Fig.2-b). FusCNN1 architecture, based on VGG-11, outperforms FusCNN2 which is based on Alexnet (Fig.2-c) for multi-modal fusion using *US(x,y)* and *MF(x,y)* images (paired t-test <0.05).

4 Discussions and Future Work

We presented a deep learning-based computational method for accurate and robust identification of pedicle plane from 2D US data. We have shown that by using multi-modality features classification accuracy improves significantly. In future, we will also investigate incorporation of data from more patients to increase the size of training data set, since it could yield higher classification accuracy.
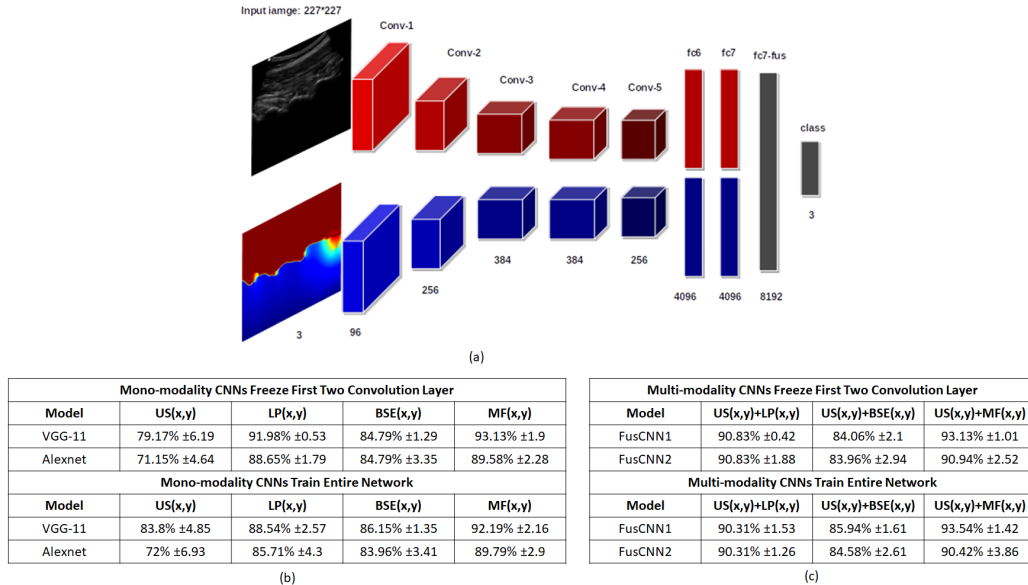

(a)

Mono-modality CNNs Freeze First Two Convolution Layer					Multi-modality CNNs Freeze First Two Convolution Layer			
Model	US(x,y)	LP(x,y)	BSE(x,y)	MF(x,y)	Model	US(x,y)+LP(x,y)	US(x,y)+BSE(x,y)	US(x,y)+MF(x,y)
VGG-11	79.17% ±6.19	91.98% ±0.53	84.79% ±1.29	93.13% ±1.9	FusCNN1	90.83% ±0.42	84.06% ±2.1	93.13% ±1.01
Alexnet	71.15% ±4.64	88.65% ±1.79	84.79% ±3.35	89.58% ±2.28	FusCNN2	90.83% ±1.88	83.96% ±2.94	90.94% ±2.52
Mono-modality CNNs Train Entire Network					Multi-modality CNNs Train Entire Network			
Model	US(x,y)	LP(x,y)	BSE(x,y)	MF(x,y)	Model	US(x,y)+LP(x,y)	US(x,y)+BSE(x,y)	US(x,y)+MF(x,y)
VGG-11	83.8% ±4.85	88.54% ±2.57	86.15% ±1.35	92.19% ±2.16	FusCNN1	90.31% ±1.53	85.94% ±1.61	93.54% ±1.42
Alexnet	72% ±6.93	85.71% ±4.3	83.96% ±3.41	89.79% ±2.9	FusCNN2	90.31% ±1.26	84.58% ±2.61	90.42% ±3.86

(b)   (c)

Figure 2: (a) Proposed multi-modal FusCNN1 architecture. For the architecture depicted in the above diagram the two multi-modal inputs are B mode US image, *US(x,y)*, and corresponding enhanced bone shadow image, *BSE(x,y)*. Each stream (blue, red) consists of five convolutional layers and two fully connected layers. Both streams coverage in one fully connected layer and a softmax classifier (gray). (b) Quantitative results obtained from the investigated mono-modality CNNs. (c) Quantitative results obtained from the proposed multi-modality CNNs.

Acknowledgement

References

Hacihaliloglu, I., 2017a. Ultrasound imaging and segmentation of bone surfaces: A review.. *Technology,* 5(02), pp. 74-80.

Hacihaliloglu, I., 2017b. Enhancement of bone shadow region using local phase-based ultrasound transmission maps. *International journal of computer assisted radiology and surgery,* June, 12((6)), pp. pp. 951-60.

Hacihaliloglu, I., Rasoulian, A., Abolmaesumi, P. & Rohling, R. N., 2014. Local Phase Tensor Features for 3D Ultrasound to Statistical Shape+Pose Spine Model Registration. *IEEE Transaction on Medical Imaging (TMI) ,* 33(11), pp. pp. 2167 - 2179.

Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems,* pp. pp. 1097-1105.

Manbachi, A., Cobbold, R. S. & Ginsberg, H. J., 2014. Guided pedicle screw insertion: techniques and training. *The Spine Journal,* 14(1), pp. pp. 165-179.

Mobbs, R. J., Sivabalan, P. & Li, J., 2011. Technique, challenges and indications for percutaneous pedicle screw fixation. *Journal of Clinical Neuroscience,* 18(6), pp. pp.741-749.

Ngiam, J. et al., 2011. *Multimodal deep learning.* pp. 689-696, InProceedings of the 28th international conference on machine learning (ICML-11).

Oh, C. H. et al., 2013. Technical Report of Free Hand Pedicle Screw Placement using the Entry Points with Junction of Proximal Edge of Transverse Process and Lamina in Lumbar Spine: Analysis of 2601 Consecutive Screws. *Korean J Spine,* 10(1), pp. pp. 7-13.

Oquab, M., Bottou, L., Laptev, I. & Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks.. *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 1717-1724.

Qi, X. et al., 2018. Automatic Scan Plane Identification from 2D Ultrasound for Pedicle Screw Guidance. *EPiC Series in Health Sciences,* Volume 18, pp. 1-7.

Russakovsky, O. et al., 2014. *ImageNet Large Scale Visual Recognition Challenge.* s.l., s.n., pp. 1717-1724 .