DISPARITY MAP PREDICTION FROM STEREO LAPAROSCOPIC IMAGES USING A PARALLEL DEEP CONVOLUTIONAL NEURAL NETWORK

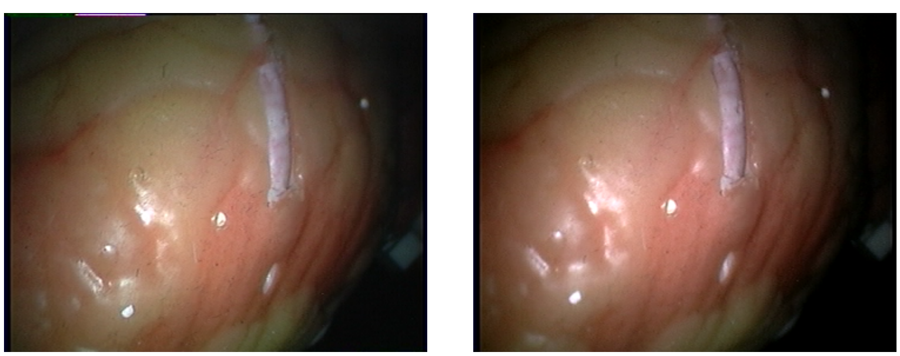
Bálint Antal
Faculty of Informatics
University of Debrecen
Debrecen, POB 400, 4002, Hungary
antal.balint@inf.unideb.hu

ABSTRACT

One of the main computational challenges in supporting minimally invasive surgery techniques is the efficient 3d reconstruction of stereo endoscopic or laparosocopic images. In this paper, a Convolutional Neural Network based approach is presented, which does not require any prior knowledge on the image acquisition technique. We have evaluated the approach on a publicly available dataset and compared to a previous deep neural network approach. The evaluation showed that the approach outperformed the previous method.

1 INTRODUCTION

Minimally invasive surgery (MIS) became a wide-spread technique to have surgical access to the abdomen of patients without casing major damages in the skin or tissues. Since MIS supporting techniques like laparascopy or endoscopy provide a restricted access to the surgeon, computer-aided visualization systems are developed. One of the major research areas in the 3d reconstruction of stereo endoscope images. See Figure 1 for an example stereo cardiac laparoscopy image pair.



(a) Left image (b) Right image

Figure 1: An example stereo laparoscopy image pair Pratt et al. (2010) Stoyanov et al. (2010).

In this paper we present an approach to predict the disparity maps of a stereo image pair by creating a very deep parallel Convolutional Neural Network (CNN). The CNN maps input RGB image patches to disparity map patches, avoiding the procedure of establishing a correspondence between the two sides which would require prior knowledge about the data.

The paper is organized as follows: Section 2 describes the proposed method and the methodology we used. We show our preliminary results in Section 3 and we draw conclusions in Section 4.

2 Disparity Map Reconstruction using Deep CNNs

In this section we describe the approach to predict disparity maps from the stereo images. Each stereo image pair is split into 24×24 overlapping parts and we can create a mapping with a very deep parallel CNN to a its respective disparity map patch.

2.1 Methodology

We have used the laparoscopic cardiac dataset provided by Pratt et al. (2010) Stoyanov et al. (2010). The dataset consists a pair of videos consisting of 2427 frames, with a spatial resolution of 360×288. The ground truth is contains a disparity map for each image.

We have used Keras Chollet (2015), Theano Bergstra et al. (2010) and cuDNN Chetlur et al. (2014) for deep learning. We have run the experiments on COTS PC with a NVIDIA Titan X installed.

2.2 The proposed parallel CNN architecture

The components of the proposed parallel CNN architecture can be categorized into two main building blocks: a CNN block (Table 1), which consists of several convolutional layers followed by max pooling and batch normalization and a Fully Connected block (Table 2). Table 3 shows the whole architecture. There are 3 CNN blocks on each side at the beginning. After the convolutional layers, the two side are merged and a succession of FC blocks maps the stereo images to a disparity map. To avoid overfitting, we used Tikhonov normalization on the weights and initialized them with the Glorot uniform formula Glorot & Bengio (2010). We have used Adamax **?** for training and allowed a maximum of 1000 epochs with early stopping. We have used a training set of 20 images and tested our approach on the remaining 2407 frames.

Table 1: The architecture of a CNN block.

Layer	**Shape**	**Activation**
Convolutional	$128 \times 5 \times 5$	ReLU Nair & Hinton (2010)
Convolutional	$64 \times 3 \times 3$	ReLU
Convolutional	$64 \times 3 \times 3$	ReLU
Convolutional	$32 \times 3 \times 3$	ReLU
Convolutional	$32 \times 3 \times 3$	ReLU
Max Pooling	2×2	
Batch Normalization		

Table 2: The architecture of a Fully connected (FC) block with p parameters.

Layer	**Shape**	**Activation**
Dense	p	ReLU
Dense	p	ReLU
Dropout (0.5)		

3 Results

To evaluate the accuracy of the proposed approach, we have calculated the root mean squared error (RMSE) of the predicted disparity maps to the ground truth. The proposed approach had a RMSE of 4.844. In comparison, another DNN-based approach Antal (2016) achieved 5.537 on the same images.

4 Conclusion

In this paper we have presented a very deep parallel convolutional neural network to predict disparity maps from stereo image pairs. We have used a stereo laparoscopic image data set and the

Table 3: The architecture of a CNN block.

Left	**Right**
CNN Block L1	CNN Block R1
CNN Block L2	CNN Block R2
CNN Block L3	CNN Block R3
Merge	
Flatten	
FC Block 1 (4096)	
FC Block 1 (2048)	
FC Block 1 (1024)	
Dense (24×24)	
Total parameters:	**38599425**

evaluation showed that it performed better then a previously publish deep neural network-based approach. Since the proposed approach does not require prior knowledge on the image acquisition, it is potentially more generalizable across devices. In the future, we will investigate this hypothesis.

ACKNOWLEDGMENTS

This work was supported in part by the project VKSZ 14-1-2015-0072, SCOPIA: Development of diagnostic tools based on endoscope technology supported by the European Union, co-financed by the European Social Fund. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

REFERENCES

Bálint Antal. Automatic 3d point set reconstruction from stereo laparoscopic images using deep neural networks. In *Proceedings of the 6th International Joint Conference on Pervasive and Embedded Computing and Communication Systems (PECCS 2016)*, pp. 116–121, 2016. ISBN 978-989-758-195-3. doi: 10.5220/0006008001160121.

James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: A cpu and gpu math compiler in python. In *Proc. 9th Python in Science Conf*, pp. 1–7, 2010.

Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cudnn: Efficient primitives for deep learning. *arXiv preprint arXiv:1410.0759*, 2014.

Franois Chollet. Keras. `https://github.com/fchollet/keras`, 2015.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *International conference on artificial intelligence and statistics*, pp. 249–256, 2010.

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814, 2010.

Philip Pratt, Danail Stoyanov, Marco Visentini-Scarzanella, and Guang-Zhong Yang. Dynamic guidance for robotic surgery using image-constrained biomechanical models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 77–85. Springer, 2010.

Danail Stoyanov, Marco Visentini Scarzanella, Philip Pratt, and Guang-Zhong Yang. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 275–282. Springer, 2010.