Standard Plane Localisation in 3D Fetal Ultrasound Using Network with Geometric and Image Loss

Yuanwei Li[1], **Juan J. Cerrolaza**[1], **Matthew Sinclair**[1], **Benjamin Hou**[1], **Amir Alansary**[1], **Bishesh Khanal**[2], **Jacqueline Matthew**[2], **Bernhard Kainz**[1], **Daniel Rueckert**[1]
[1]Imperial College London, [2]Kings College London
yuanwei.li09@imperial.ac.uk

Abstract

Standard scan plane detection in 3D fetal brain ultrasound (US) is a crucial step in the assessment of fetal brain development. We propose an automatic method for the detection of standard planes in 3D volumes by utilising a convolutional neural network (CNN) to learn the relationship between a 2D plane image and the transformation parameters required to move that plane towards the corresponding standard plane. In addition, we explore the effect of using two different training loss functions which exploit the geometric information and the image data of the extracted plane respectively. When evaluated on 72 subjects, our method achieves a plane detection error of 3.45mm and 12.4°.

1 Introduction

3D US imaging of the fetal brain enables clinicians to assess fetal brain development and detect growth abnormalities. However, this requires the accurate extraction of 2D standard scan planes such as the transventricular (TV) and transcerebellar (TC) plane that contain key anatomical structures [6]. This task is challenging, operator-dependent and time-consuming even for experienced sonographers. Hence, there is a strong need to develop automatic methods for 2D standard plane extraction from 3D volumes to improve clinical workflow efficiency.

Recently, several works have applied deep learning techniques to standard plane detection in fetal US by treating it as an image classification problem [1, 2]. But these methods identify standard planes from 2D US videos and are computationally infeasible for plane localisation in 3D volumes since there are infinitely many ways to sample an arbitrary 2D plane in 3D space. To this end, we approach the plane detection problem similar to the work of [5, 3] by using a CNN to predict the transformation parameters that define the plane position and orientation. The CNN learns a mapping between a 2D plane and the transformation required to move that plane towards the standard plane within a 3D volume. Our approach is iterative which uses multiple passes of the CNN to predict a more accurate plane location at each iteration during inference. In addition, we investigate the use of two training loss functions: (1) A geometric loss that minimises the mean-square-error (MSE) of the geometric transformations defining the planes [5] and (2) an image loss that minimises the MSE of the image data extracted from the planes by using a spatial transformer network (STN) [4].

2 Method

Planes and Transformations: Any plane in 3D space can be defined by a rigid transformation with respect to a reference plane. In Fig. 1a, we define an identity plane (black) with origin at the volume centre. T_{in} and T_{out} are defined using the identity plane as the reference and they move the identity plane to the blue plane and red plane respectively. Our CNN predicts ΔT which is defined using the

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

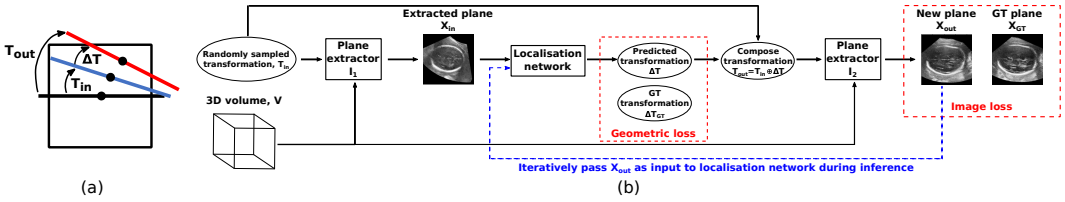

Figure 1: (a) Composition of transformations. Black: Identity plane. Blue: Input plane. Red: Output plane. (b) Overall framework of our proposed method for standard plane localisation.

blue plane as reference. We can compose transformations using $T_{out} = T_{in} \oplus \Delta T$ where $\oplus$ is the composition operator.

Localisation Network: The localisation network is a CNN that regresses the values of a 3D transformation ΔT given a 2D input image X_{in}. The network learns a mapping between the image extracted at the current plane position and the transformation required to move the current plane to a new position that is closer to the ground truth (GT) plane. Since we are predicting plane movement, the transformation has to be rigid, comprising translation $\boldsymbol{t}$ and rotation represented by quaternions $\boldsymbol{q}$: $\Delta T = (\boldsymbol{t}, \boldsymbol{q})$.

Our localisation network comprises 5 convolution layers, each followed by max-pooling. After the last pooling layer, the network splits into two branches, each comprising 2 fully-connected (FC) layers followed by a regression output layer. One branch regresses 3 parameters for translation ($\boldsymbol{t}$) while the other branch regresses 4 parameters for quaternions ($\boldsymbol{q}$). All convolution layers use 3x3 kernel with stride=1 and all pooling layers use 2x2 kernel with stride=2. ReLU activation function is applied after all convolution and FC layers. Drop-out is also added after each FC layer.

Plane Extractor: The plane extractor is a module in our network pipeline that extracts any arbitrary 2D plane image X from a 3D volume V. We denote $X = I(V, T, s)$ where $I(\cdot)$ is the plane extraction function, T is a transformation applied to the identity plane and s is the length of a square plane. The plane extractor does the following: (1) Initialise an identity plane of size $s \times s$ by creating a meshgrid of points that slice through the volume centre. (2) Apply T to the meshgrid of points to update their positions. (3) Sample the updated meshgrid from V using trilinear interpolation to obtain X. The plane extractor is based on STN [4] which is differentiable and can be trained with backpropagation. This allows the module to be incorporated into our network for end-to-end training.

Network Training: Network training is summarised in Fig. 1b. A training sample is represented by $(X_{in}, \Delta T_{GT}, X_{GT})$ where X_{in} is a plane image randomly sampled from V, ΔT_{GT} is the transformation that will move the randomly sampled plane to the GT plane location, and X_{GT} is the GT plane image. First, we randomly sample a transformation T_{in} from which we extract the corresponding plane image $X_{in} = I_1(V, T_{in}, s)$. X_{in} is then passed as input to the localisation network which predicts a transformation ΔT. A geometric loss is then formulated as the MSE between the GT and predicted transformation parameters: $L_{geom} = \|\Delta T_{GT} - \Delta T\|_2^2 = \|\boldsymbol{t}_{GT} - \boldsymbol{t}\|_2^2 + \left\|\boldsymbol{q}_{GT} - \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|}\right\|_2^2$.

The predicted transformation ΔT is relative to the X_{in} plane. We compose T_{in} and ΔT to obtain T_{out} which defines the new plane location relative to the identity plane. We then extract the image at the updated plane location $X_{out} = I_2(V, T_{out}, s)$. An image loss can subsequently be computed as the MSE between the GT and predicted plane images: $L_{img} = \|X_{GT} - X_{out}\|_2^2$. The combined loss function is given by: $L = \alpha\|\boldsymbol{t}_{GT} - \boldsymbol{t}\|_2^2 + \beta\left\|\boldsymbol{q}_{GT} - \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|}\right\|_2^2 + \gamma\|X_{GT} - X_{out}\|_2^2$ where α, β and γ are the loss weights.

Network Inference: During inference, our method adopts an iterative approach to find the standard plane. First, a plane is randomly initialised X_0 and passed as input to the localisation network which predicts a transformation ΔT that moves the current plane at T_0 to a new position T_1. The image X_1 extracted from the new plane location is then passed to the localisation network (blue arrow in Fig. 1b). This process is repeated for N iterations to give the final result as X_N. For each volume, we initialise 5 random planes and take their average as the final result.

Table 1: Evaluation of the proposed method for standard plane detection with different training losses. Results presented as (Mean ± Standard Deviation).

	δx (mm)	$\delta\theta$ (°)	PSNR	SSIM
M1: L_{geom}	6.51±4.86	14.1±8.2	15.6±2.0	0.393±0.082
M2: L_{img}	10.23±16.08	16.6±8.2	15.1±2.3	0.372±0.090
M3: $L_{geom} + L_{img}$	5.85±3.95	12.9±7.0	15.7±2.1	0.400±0.101
M3+: $L_{geom} + L_{img}$	**3.45±1.73**	**12.4±12.8**	**16.5±1.9**	**0.418±0.080**

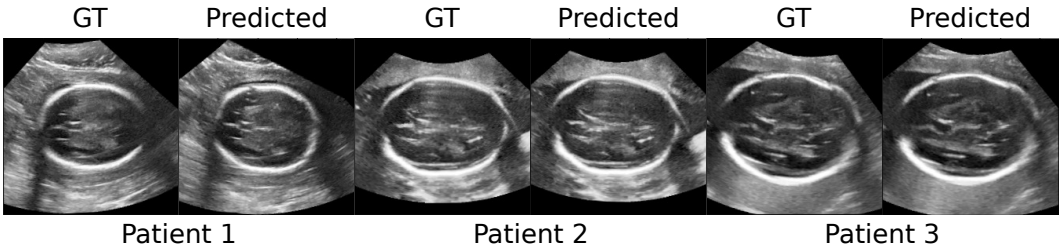

Figure 2: Visualisation of GT planes and planes predicted by M3+.

3 Experiments and Results

Data and Experiments: The proposed method is evaluated on 3D US volumes of fetal brain from 72 subjects for the detection of TV standard plane. A training/testing split of 70%/30% is used. All volumes are processed to be isotropic with mean dimensions of 324×207×279 voxels. The method is implemented using Tensorflow running on one NVIDIA Titan Xp GPU. We set s=225, N=10 and loss weights α=γ=1e4, β=1. Each of these losses can be removed by setting its weight to zero. Batch size is set to 8. Weights are initialised randomly from a distribution with zero mean and 0.1 standard deviation. Optimisation is carried out for 200,000 iterations using the Adam algorithm with learning rate=0.001, β_1=0.9 and β_2=0.999. The predicted plane is evaluated against the GT using distance between the plane centres (δx) and rotation angle between the planes ($\delta\theta$). Image similarity of the planes is also measured using peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

Results: Table 1 compares the plane detection results of using different training losses for the localisation network. Performance is improved when using both the geometric and image losses which complement each other by providing geometric and image information respectively (M3). We further improve our results by using three orthogonal plane images as network inputs as this provides more information about the 3D volume (M3+). M3 and M3+ take 0.53s and 1.28s to predict one plane per volume respectively. Fig. 2 shows a visual comparison between the GT planes and the planes predicted by M3+.

Conclusion: We presented a new method for standard plane detection in 3D fetal US by using a CNN to regress transformations iteratively. We use a combined training loss that accounts for both geometric and image information to improve detection accuracy. As future work, we are exploring other training loss functions and are extending the method to multiple planes detection.

References

[1] Baumgartner, C.F., Kamnitsas, K., Matthew, J., et al.: Sononet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE TMI 36(11), 2204–2215 (Nov 2017)

[2] Chen, H., Dou, Q., Ni, D., et al.: Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: MICCAI. pp. 507–514. Springer (2015)

[3] Hou, B., Alansary, A., McDonagh, S., et al.: Predicting slice-to-volume transformation in presence of arbitrary subject motion. In: MICCAI. pp. 296–304. Springer (2017)

[4] Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: NIPS. pp. 2017–2025

[5] Kendall, A., Grimes, M., Cipolla, R.: Posenet: A convolutional network for real-time 6-dof camera relocalization. In: ICCV. pp. 2938–2946. IEEE (2015)

[6] NHS: Fetal anomaly screening programme: programme handbook June 2015. Public Health England (2015)