CNSEG-GAN: A LIGHTWEIGHT GENERATIVE ADVERSARIAL NETWORK FOR SEGMENTATION OF CRL AND NT FROM FIRST-TRIMESTER FETAL ULTRASOUND

Md. Mostafa Kamal Sarker[1] *Robail Yasrab*[1] *Mohammad Alsharid*[1,3]
Aris T. Papageorghiou[2] *J. Alison Noble*[1]

[1]Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[2]Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK
[3]Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, UAE

ABSTRACT

This paper presents a novel, low-compute and efficient generative adversarial network (GAN) design for automatic segmentation called CNSeg-GAN, which combines 1-D kernel factorized networks, spatial and channel attention, and multi-scale aggregation mechanisms in a conditional GAN (cGAN) fashion. The proposed CNSeg-GAN architecture is trained and tested on a first-trimester ultrasound (US) scan video dataset for automatic detection and segmentation of anatomical structures in the midsagittal plane to enable Crown Rump Length (CRL) and Nuchal Translucency (NT) measurement. Experimental results shows that the proposed CNSeg-GAN is x15 faster than U-Net and yields mIoU of 78.20% on the CRL and 89.03% on the NT dataset, respectively with only 2.19 millions in parameters. The accuracy of this lightweight design makes it well-suited for real-time deployment in future work.

***Index Terms*—** First trimester, ultrasound, video segmentation, midsagittal plane, generative adversarial network.

1. INTRODUCTION

Fetal ultrasound (US) imaging is a crucial part of pregnancy care, allowing healthcare providers to monitor fetal growth and health. The first-trimester fetal US scan (also known as the dating or nuchal scan) is carried out between 11^{+0} to 13^{+6} $weeks^{days}$ of gestation to assure pregnancy viability, accurately date the pregnancy and to assess the risk of chromosomal anomalies [1]. During the first-trimester scan, sonographers acquire various imaging planes, also known as standard planes (SP), to visualize required anatomical structures. These SP include the midsagittal plane for measuring Crown-Rump Length (CRL) and Nuchal Translucency (NT).

Previous deep learning-based US segmentation models have reported good accuracy [2]. However, previously reported models have a large number of parameters which limits applicability for real-time deployment on the growing number of ultrasound-based devices that have limited computational and memory resources. Fetal standard plane segmentation is challenging for first-trimester scans due to relatively small size of fetus and its substructures. The use of automated segmentation as an overlaying-based assistive workflow tool could help increase the accuracy (and speed) of anatomical assessment and measurement, as well as support newly-qualified operators in these tasks.

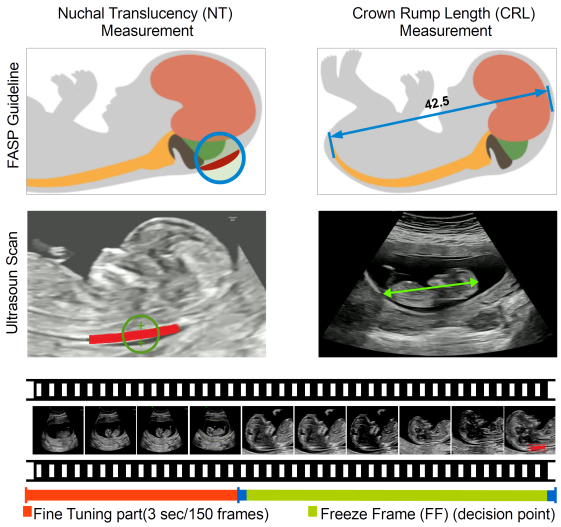

Fig. 1: An overview of the fetal anatomy (CRL and NT) for the first-trimester US scans. Video frames are acquired from frozen video segments (green), and fine-tune segments (red). For more details, readers are referred to [3].

Contribution: A lightweight cGAN model (2.19 million parameters) called CNSeg-GAN is proposed to segment anatomical structures in the mid-sagittal plane (MSP) from an US scan video. It segments the key CRL and NT structures using a multiscale aggregation strategy with spatial and channel attention modules to capture the correlation between the spatial and channel features from coarse-to-fine pixel levels and to discriminate between fetal structures and the speckled

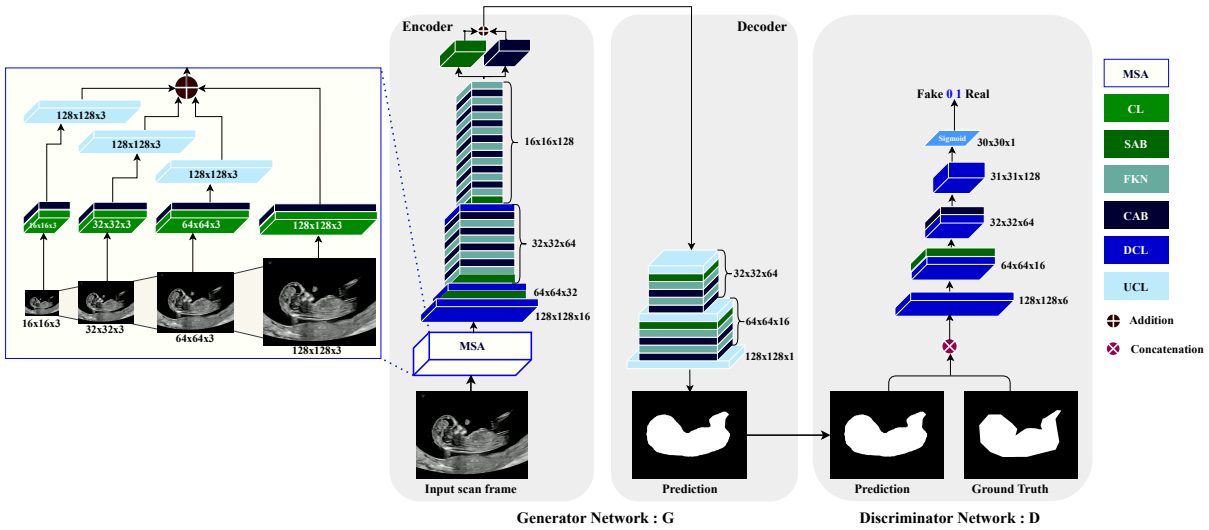

Fig. 2: An overview of the proposed CNSeg-GAN architecture for automated CRL and NT segmentation in first-trimester US scans.

or blurry background. The CNSeg-GAN has been designed to be suitable for real-time overlaying and guidance. The model processing speed is 110 frames-per-second on a single Nvidia Quadro RTX 5000 (16GB) GPU.

Related Work. Recent reports on CNN-based automated first-trimester fetal biometry include measurement of the CRL [4], brain [5] and NT [5]. Cengiz et al. [4] designed a U-Net based CNN for CRL measurement assuming first-trimester standard plane images as input. Similarly, Bano et al. [6] proposed to estimate fetal biometry from a limited set of data using a single frame-based semantic segmentation method. Prior segmentation methods [2, 7] did not utilize GAN-based real-time methods the CRL and NT segmentation. Here, we argue that to be useful in clinical practice, a fully-automated lightweight neural network must work in real-time. The proposed model provides pixel-wise semantic segmentation of the midsagittal standard plane from free-hand US video scans.

2. METHODS

2.1. Data Acquisition and Pre-processing

Routine clinical first-trimester fetal US scans were available from a large-scale study PULSE. According to NHS Fetal Anomaly Screening Programme (FASP) guidelines [8] CRL and NT measurements should be made in the first trimester US scan. US video is acquired through screen-grab signals at 30 frames per second of a GE Voluson E8 version BT18 (GE Healthcare, Zipf, Austria) US machine. On average, a complete first-trimester US scan takes 13.73 ± 4.18 minutes, with an average of $24,720 \pm 7,534$ frames per scan video. The fetal structure (CRL) and NT mask were segmented using a training set of manually annotated video clips. The data distribution used in this work is summarized in Table 1.

Table 1: Details of datasets and tasks used in this study.

Anatomy	Datasets	Video Segments	Frames
CRL	Training	100	12534 (77.9%)
	Test	28	3559 (22.1%)
NT	Training	110	10174 (79.3%)
	Test	36	2647 (20.7%)

2.2. Model Architecture

The key modules of our proposed model architecture are designed in accordance with the cGAN baseline of pix2pix [9] methods. It consists of two main networks: the generator G and the discriminator D. Fig. 2 depicts the overall CNSeg-GAN architecture further explained in detail next. **Generator Network (G):** As shown in Fig. 2 G consists of two modules, namely an encoder and a decoder. ***Encoder*** module of CNSeg-GAN comprises of one multiscale aggregation block (MSA), three downsampling convolutional layers (DCL), four spatial attention blocks (SAB), twelve 1-D kernel factorized networks (FKN), and twelve channel attention blocks (CAB). Here, the spatial attention block (SAB) feature map F_{SAB} is defined as,

$$F_{SAB} = F^A \oplus (F_r^D \otimes exp(F_{r,t}^B \otimes F_r^C))_r, \qquad (1)$$

where feature maps are denoted $F^A = (ReLU(BN(CL)))$, BN denotes batch normalization, $F_{r,t}^B$, F_r^C, and F_r^D denotes new feature maps generated from F^A by applying reshape (r) and transpose (t) to the height (H) and width (W) only (channel (C) is fixed) with exp for the softmax function. Moreover, the spatial attention block (CAB) feature map F_{CAB} is defined as,

$$F_{CAB} = F^A \oplus (F_r^A \otimes exp(F_{r,t}^A \otimes F_r^A))_r, \qquad (2)$$

where F_r^A is an altered version of F^A that has undergone reshaping in the channel dimension, $F_{r,t}^A$ is an altered version of F^A that has undergone reshaping and transposition in the channel (C) dimension. In addition, the feature map of kernel factorized networks (FKN) F_{FKN} is defined as,

$$F_{FKN} = F^v(F^h(F^v(F^h(F^k)))) \oplus F^k, \tag{3}$$

where F^k is the input feature, $F^v = ReLU(BN(F^k_{1\times3}))$ representing an FKN with a vertical kernel of 1×3, and $F^h = ReLU((F^k_{3\times1})$ with a 3×1 horizontal kernel. The CNSeg-GAN ***Decoder*** module contains three UCL, two SAB, four FKN, and four CAB consecutively and is shown in Fig. 2. These blocks create the segmentation masks with a size of 128×128 from the encoder module prediction by applying a threshold of 0.5.

Discriminator Network (D). Fig. 2 presents the D as four layers, including four DCLs, one SAB, and one CAB. The first 3 DCLs consist of a CL with kernel size 4×4, stride 2, and padding 1. A SAB and CAB are used after the second and third DCL blocks, respectively. Finally, a sigmoid activation function is included in the final layer of D.

2.3. Model Training and Implementation

Model Training. We train the G and D networks in CNSeg-GAN via adversarial back-propagation. Assume that i is an input US image and o is the corresponding segmentation mask ground truth. Let d be a random variable included as a dropout in the decoder's layers to prevent the model from overfitting and to expand the learning process. Further, let $G(i,d)$and $D(i,G(i,d))$ describe the outputs of the generator and discriminator, respectively. The loss function of the generator ℓ_{Ge} is defined as,

$$\begin{aligned}\ell_{Ge}(G,D) = \mathbb{E}_{i,o,d}(-\log(D(i,G(i,d)))) \\ + \gamma\mathbb{E}_{i,o,d}(\ell_{L_1}(o,G(i,d)))\end{aligned}. \tag{4}$$

Here γ is an empirical weighting factor. The loss function of the generator ℓ_{Di} is defined as,

$$\begin{aligned}\ell_{Di}(G,D) = \mathbb{E}_{i,o,d}(-\log(D(i,o))) \\ + \mathbb{E}_{i,o,d}(-\log(1-D(i,G(i,d))))\end{aligned}. \tag{5}$$

Here, $-\log(1-D(x,G(x,z))$ and $-\log(D(x,y))$ represents the predicted segmentation and the ground-truth mask, respectively.

Implementation Details. The CNSeg-GAN architecture was implemented by PyTorch v0.4.1. US video frames were scaled to 128 × 128 pixels. A learning rate of 0.0002, batch size of 2, Adam optimization with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ were used. Data augmentation was not used for the proposed model, all CNSeg-GAN G and D layers were trained from scratch. To train the comparative CNN models, data augmentation with a pre-trained model (ResNet101) was used.

3. RESULTS AND DISCUSSION

Table 2.3 reports the performance evaluation for the CNSeg-GAN model and five comparative CNN segmentation models using accuracy (AC), Dice score (DS) and mean intersection over union (mIoU) metrics.

Quantitative Evaluation of Trained Models: We trained and tested benchmark CNN-based segmentation models (U-Net [10], PAN [11], MA-Net [12], DeepLabV3+ [13], and NHG [2]) which were selected due to their high benchmark segmentation performance on public medical imaging and computer vision datasets. The experimental results in Table 2.3 show that the CNSeg-GAN model outperforms all other tested models in terms of AC, DS, and mIoU on both study datasets. CNSeg-GAN achieves AC, DS, and mIoU scores of 95.22%, 90.85%, and 78.20% on the CRL dataset and 99.84%, 93.80%, and 89.03% on the NT dataset. CNSeg-GAN yields 2.49%, 2.64% higher DS, and 1.28%, 18.47% higher mIoU on the CRL and NT datasets respectively compared with the U-Net. The superior performance is particularly notable for the NT dataset where it is more challenging to segment the small structure within a complex background. Fig. 2 presents an example box plot of mIoU for all CRL test samples. The coloured boxes show the range of scores for different models. Performing the Wilcoxon signed-rank test between the CNSeg-GAN model and second-best model (U-Net) on the CRL and NT datasets was statistically significant (p-value < 0.001).

Qualitative Analysis: Fig. 4 presents a qualitative evaluation of the CNSeg-GAN model with the second-best model (U-Net). We show sample frame segmentations of the CRL and NT and their ground truth. These US frames contain speckle and incomplete anatomical boundaries. In the first row, the proposed approach can be seen to accurately segment the boundaries in the presence of hypoechoic tissue regions. However, U-Net does not accurately delineate the entire fetal region where the pixel intensities change for the fetal structures. The attention mechanism of the CNSeg-GAN model aids detection of hypoechoic tissue pixels even when speckle

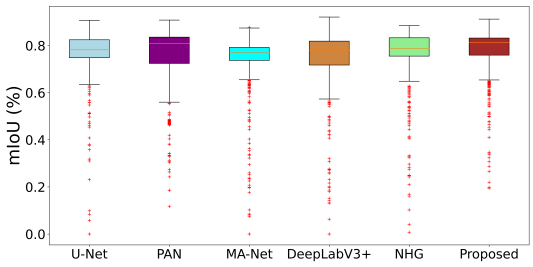

Fig. 3: Box plot of mIoU for different benchmark CNN models and ours on the CRL test dataset.

Table 2: Quantitative analysis of trained models on CRL and NT test datasets.

Methods	Parameters	CRL			NT		
	Millions	AC(%)	DS(%)	mIoU(%)	AC(%)	DS(%)	mIoU(%)
U-Net [10]	51.51	95.12 ± 0.06	88.36 ± 0.08	76.92 ± 0.09	**99.91** ± 0.01	91.16 ± 0.11	70.56 ± 0.13
PAN [11]	43.25	94.95 ± 0.03	87.76 ± 0.09	76.31 ± 0.12	99.93 ± 0.01	92.28 ± 0.11	71.93 ± 0.12
MA-Net [12]	166.43	94.55 ± 0.04	86.68 ± 0.11	74.75 ± 0.12	99.90 ± 0.01	89.29 ± 0.13	68.30 ± 0.15
DeepLabV3+ [13]	45.66	95.54 ± 0.02	86.59 ± 0.12	74.74 ± 0.13	99.89 ± 0.01	89.21 ± 0.14	68.42 ± 0.15
NHG [2]	11.46	92.32 ± 0.03	88.24 ± 0.09	74.42 ± 0.04	92.49 ± 0.05	82.14 ± 0.01	67.37 ± 0.01
CNSeg-GAN (proposed)	**2.19**	**95.22** ± 0.03	**90.85** ± 0.07	**78.20** ± 0.09	99.84 ± 0.01	**93.80** ± 0.07	**89.03** ± 0.10

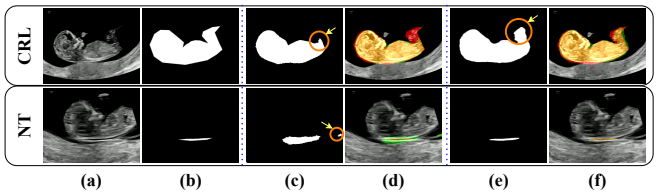

Fig. 4: Examples of proposed segmentation results on compared to U-Net model on CRL and NT datasets. (a) input video frame, (b) ground truth mask, (c) U-Net prediction, (d) U-Net model prediction overlaid on the US frame, (e) proposed model prediction (f) proposed model prediction overlaid on the US frame. Note that the colors of the overlay visualization results are as follows: TP (orange), FP (green), FN (red) and TN (background).

is present. The second-row example illustrates another case where the CNSeg-GAN model works well. In this case, a multi-scale input and the attention blocks place more emphasis on neighboring hyperechoic and hypoechoic pixels, resulting in fewer false positives than for the U-Net.

4. CONCLUSION

We have presented a lightweight cGAN-based architecture that takes a free-hand US scan video frame as an input and outputs key structure segmentations (NT, CRL) for first-trimester fetal ultrasound images. The proposed low-compute CNSeg-GAN model relies on factorized kernels with spatial and channel attention blocks to perform segmentations. The proposed model has only 2.19 million parameters and outperformed other benchmark architectures in terms of DS and mIoU with better parameter efficiency on the reported test data. Our lightweight design allows for the real-time accurate segmentations of CRL and NT structures. In the future, we aim to extend the current architecture towards different fetal anatomies for segmentation and guidance tasks.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was approved by the UK Research Ethics Committee (Reference 18/WS/0051) and the ERC ethics committee.

6. ACKNOWLEDGEMENTS

This work was also supported in part by the InnoHK-funded Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE) Project 2.1 (Cardiovascular risks in early life and fetal echocardiography).

7. REFERENCES

[1] LJ Salomon et al., "Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan," *UOG*, vol. 37, no. 1, pp. 116–126, 2011.

[2] R Yasrab et al., "End-to-end first trimester fetal ultrasound video automated crl and nt segmentation," in *IEEE ISBI*. IEEE, 2022, pp. 1–5.

[3] L Drukker et al., "Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

[4] C Sevim et al., "Automatic fetal gestational age estimation from first trimester scans," in *ASMUS*. Springer, 2021, pp. 220–227.

[5] TM Christeena et al., "Deep learning measurement model to segment the nuchal translucency region for the early identification of down syndrome," *Measurement Science Review*, vol. 22, no. 4, pp. 187–192, 2022.

[6] B Sophia et al., "Autofb: Automating fetal biometry estimation from standard ultrasound planes," in *MICCAI*. Springer, 2021, pp. 228–238.

[7] Z Sobhaninia et al., "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *2019 41st annual intl. conference of the IEEE EMBC*. IEEE, 2019, pp. 6545–6548.

[8] "Fetal anomaly screen programme handbook," Report, NHS Screening Programmes, London, UK, 2015.

[9] P Isola et al., "Image-to-image translation with conditional adversarial networks," in *IEEE CVPR*, 2017, pp. 1125–1134.

[10] O Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*. Springer, 2015, pp. 234–241.

[11] L Hanchao et al., "Pyramid attention network for semantic segmentation," *arXiv:1805.10180*, 2018.

[12] F Tongle et al., "Ma-net: A multi-scale attention network for liver and tumor segmentation," *IEEE Access*, vol. 8, pp. 179656–179665, 2020.

[13] LC Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," *CoRR*, vol. abs/1802.02611, 2018.