2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

GENERATING CONTROLLABLE ULTRASOUND IMAGES OF THE FETAL HEAD

Lok Hin Lee and J. Alison Noble

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

ABSTRACT

Synthesis of anatomically realistic ultrasound images could be potentially valuable in sonographer training and to provide training images for algorithms, but is a challenging technical problem. Generating examples where different image attributes can be controlled may also be useful for tasks such as semi-supervised classification and regression to augment costly human annotation. In this paper, we propose using an information maximizing generative adversarial network with a least-squares loss function to generate new examples of fetal brain ultrasound images from clinically acquired healthy subject twenty-week anatomy scans. The unsupervised network succeeds in disentangling natural clinical variations in anatomical visibility and image acquisition parameters, which allows for user-control in image generation. To evaluate our method, we also introduce an additional synthetic fetal ultrasound specific image quality metric called the Fréchet SonoNet Distance (FSD) to quantitatively evaluate synthesis quality. To the best of our knowledge, this is the first work that generates ultrasound images with a generator network trained on clinical acquisitions where governing parameters can be controlled in a visually interpretable manner.

Index Terms— Generative adversarial networks, representation disentanglement, fetal ultrasound

1. INTRODUCTION

Current fetal sonographer training methods rely on expensive collations of clinically acquired data or fetal phantoms, which do not fully mimic the variability inherent in a clinical setting. Current state-of-the-art in ultrasound image analysis methods mostly rely on deep learning, which require large amounts of data. This limiting factor prompts the question of whether it is possible to simulate realistic ultrasound data, and if yes, under what conditions. Such an ability may not only be useful for deep learning research but in the future could support classroom-based training of sonographers in ultrasound image interpretation.

One method of ultrasound image synthesis involves simulating the propagation of the ultrasound wave through tissue in software with physics-based simulations. This involves numerical modelling of an intractable wave equation through inhomogeneous tissue media using large-scale ray tracing or finite elements calculations, which are computationally expensive [1]. These methods are also largely based upon a prior manual tissue segmentation or three-dimensional tissue model acquired from a separate imaging modality, which may not be available. In vivo tissue properties are also subject-specific making realistic simulation a challenge.

A second method involves simulating ultrasound images using artificial ultrasound phantoms. However, in this case the quality of the simulated ultrasound image is highly dependent upon the quality of material and construction. Furthermore, a physical ultrasound phantom is unable to accurately model the variations in anatomy that sonographers or deep learning models will encounter in clinical practice.

Recent approaches involving deep learning methods have experienced some success in automated ultrasound image generation. Generative adversarial network (GAN) based architectures have been used to simulate ultrasound images from synthetic data [2] and ultrasound phantoms [3]. However, to our knowledge, there are no publications that have trained GANs with clinically acquired unlabelled ultrasound images, which have more anatomical and image variation than simulated or phantom-based approaches. Furthermore, both prior works require a supervising signal either in terms of a prior tissue segmentation or spatial coordinates, which increases the cost of data collection compared to our method which is unsupervised.

In this work, we employ an information maximizing GAN [4] (InfoGAN) to generate fetal brain images from clinical fetal ultrasound scans. As a consequence of the structure of an InfoGAN, the model is able to untangle latent representations of the generated image. We show that this leads to control over the visibility of specific anatomical structures as seen in a fetal brain and imaging parameters representative of the variability in what a sonographer might see in a fetal ultrasound scan. We also quantitatively analyze the quality of the images by

This work is supported by the Croucher Foundation Croucher Scholarship for Doctoral Study.

978-1-5386-9330-8/20/$31.00 ©2020 IEEE

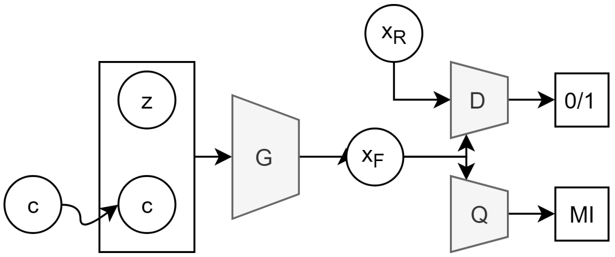

Fig. 1. A schematic view of an InfoGAN. z represents the latent incompressible noise variable that is input into the generator (”G”), whilst c represents the latent code that may be sampled from a given prior. The generator generates fake examples of data x_F, which are fed into the discriminator (”D”) along with real examples of data x_R. The discriminator then learns to discern between the generated fake examples of data and the real data. However, the latent code is also fed into a separate auxiliary network (”Q”) that estimates the posterior $P(c|x_F)$ so that the mutual information (”MI”) between the distribution and the latent code can be calculated and maximized in the loss function. In practice, Q and D share most layers and weights.

adopting a GAN quality metric that is used in the domain of natural images [5] and modifying it such that it is fetal ultrasound image specific.

2. METHODS

2.1. Data

Our dataset consists of fetal head ultrasound images acquired from routine clinical 20-week anatomy pregnancy scans. The whole dataset consists of 117,558 frames, manually extracted from 662 separate scans where the fetal skull was visible. Each frame was cropped so that extraneous patient data and imaging parameter text is removed and only the ultrasound image was visible. Each frame was then downsampled to 90x86 pixels. Random horizontal flipping was used during the training process to further augment the data.

2.2. Information Maximizing Generative Adversarial Networks

We model the distribution of the real data $P_{data}(x)$ using an information maximizing generative adversarial network (“InfoGAN”) (Figure 1) [4], which is able to learn untangled representations as a consequence of the network structure. With InfoGANs, a generator $G(z, c; \theta_G)$ attempts to create false examples of data x_F by taking as an input both a latent noise variable $z \sim P(z)$, as well as a latent code $c \sim P(c)$. As before, the discriminator network attempts to discriminate between x_F and x_R. However, to prevent the generator from ignoring the latent code, an additional cost term is introduced that rewards maximizing mutual information between c and the predicted distribution $P(c|x_F)$. A lower bound of the mutual information between c and x_F is calculated by estimating the posterior $P(c|x)$ using an auxiliary network Q.

Table 1. An overview of the generator and discriminator architectures investigated. CONV represents a convolutional layer, TCONV represents a transposed convolutional layer, BN represents Batch Normalization, FC represents a fully connected layer, and & represents the concatenation operation. (K3, S2, O256) implies a layer with a kernel size of 3x3, a stride of 2 and 256 channels.

Generator	Discriminator/Auxiliary
Input z	Input(96×80×1)
FC(O15360), BN, ReLU	CONV(K5,S2,O64), BN, LReLU
Reshape(5×6, O512) & Input c	CONV(K5,S2,O64), BN, LReLU
TCONV(K3,S2,O256), BN, ReLU	CONV(K5,S2,O128), BN, LReLU
TCONV(K3,S2,O128), BN, ReLU	CONV(K5,S2,O256), BN, LReLU
TCONV(K3,S2,O64), BN, ReLU	CONV(K5,S2,O512), BN, LReLU
TCONV(K3,S2,O1), Tanh	FC(O1), Loss

We argue that using an InfoGAN may lead to control over the visibility of specific anatomical structures in a fetal brain scan or control over imaging parameters representative of the variability seen in a routine fetal ultrasound scan. These effects can be modelled by the latent codes c and therefore disentangled from the noise vector z.

In order to successfully train the network, we use spectral normalization [6] and mini-batch discrimination [7]. This empirically increases training stability and reduces the propagation of temporary errors made by the discriminator back to the generator during the training process. We also investigate the use of a least squared loss (L-S) for the discriminator [8] and compare its results with those generated using a standard cross-entropy discriminator loss.

2.3. Network Architecture

An overview of the network architectures that are investigated is provided in Table 1. The generator takes in a noise vector z that is sampled from a Normal distribution $N(0, 1)$ and latent codes c. We vary the sampling of c from $Uniform(-1, 1)$ and $N(0, 1)$ as well as the number of latent codes n in different experiments c_n.

The noise vector z is projected to the initial feature size (5×6) and channel count (512). In order to emphasize the use of the latent codes to the generator, each sampled code is tiled to the initial feature size and concatenated to the initial channel count. This is then alternatively up-sampled and convolved to create a final generated image. We use the Adam optimizer and learning rates of 4e-4 and 1e-4 for D and G respectively.

The discriminator and auxiliary network share all of the initial convolution layers until the final fully connected layer. This is because both networks require a high level embedding of the image. This reduces the computational and memory requirements of the network. For each latent code variable c_n, the auxiliary network generates an estimate of the posterior distribution.

3. RESULTS

3.1. Fréchet SonoNet Distance

We modify a strategy that is used to quantify GAN performance in the natural image domain, the Fréchet Inception Distance (FID) [5] and adapt it to analyze ultrasound generator performance. FID measures the Fréchet distance between estimated multivariate normal distributions fitted to the activations of the coding layer of an ImageNet pre-trained Inception-v3 layer for real and generated examples respectively. A smaller FID value implies a closer distribution distance between real and generated images, and therefore more similar image features and higher image quality. However, FID analysis is limited by the fact that the Inception-v3 architecture is pre-trained on ImageNet. Learnt features that are extracted are therefore ImageNet specific, which might lead to bias and inaccuracy in the results for ultrasonography data in our case, which has different characteristics to natural images.

Therefore, in addition to the FID, we calculate the Fréchet SonoNet Distance (FSD) by adding a global average pooling layer to the final image feature extractor layer for a pre-trained SonoNet-64 [9] and measuring the resulting Fréchet distance between the multivariate normal distributions fitted onto the hidden representations for real and generated examples. SonoNet-64 is a ultrasound classification neural network that is trained on over 47k frames acquired from clinical fetal ultrasound scans. It is therefore fetal ultrasound specific and more appropriate for our quantitative analysis. We find FSD a good predictor of generator image quality during training and show the model with the minimum FSD in the different architectures to maximize generated image quality.

3.2. Quantitative Evaluation

We evaluate the performance of the InfoGANs by generating 5,000 images every 10k iterations and calculating the Fréchet distances between the real dataset and generated images (Fig. 2). Each GAN architecture was trained with $c_n, n \in \{3, 4...9\}$ for each latent code distribution. Each GAN was trained for a minimum of 700k iterations.

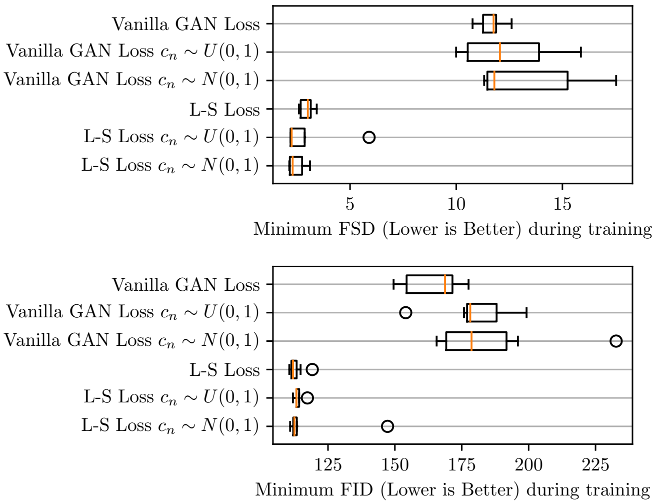

Fig. 2. The figure shows the minimum Fréchet distances each generative architecture experienced during training. For the architectures with latent codes as an input parameter, the number of latent codes were varied to investigate whether the number of latent codes affected FSD $c_n, n \in \{3, 4...9\}$. The networks with no latent codes were trained with 7 different random seeds to obtain a comparative result.

3.3. Qualitative Evaluation

We found that using a least square GAN loss enables the generator to capture more of the real data distribution, reflected in the lower distances. Furthermore, using a least squares loss stabilizes training and reduces the incidence of mode collapse, as can be seen by the reduced variation in both FID and FSD. However, there does not appear to be significant differences between using a latent code that was drawn from a uniform distribution or a normal distribution. Changing the number of latent codes n does not significantly vary final FSD, but the human explainability of additional codes decreased as the number of latent codes increased. This may be because the generator is forced to use increasingly minor variations in image generation to map to latent codes as variation is exhausted in order to maintain image realism. We also found that FSD and FID broadly correlated with each other, and that FSD is a useful measure of ultrasound image quality.

We vary the disentangled latent codes of the best-performing model on a FSD basis and generate images with disentangled, controllable representations seen in Figure 3. Qualitatively, this network appears to have captured variation of the size of the fetal skull with c_1, skull width with c_2, the visibility of the brain mid-line

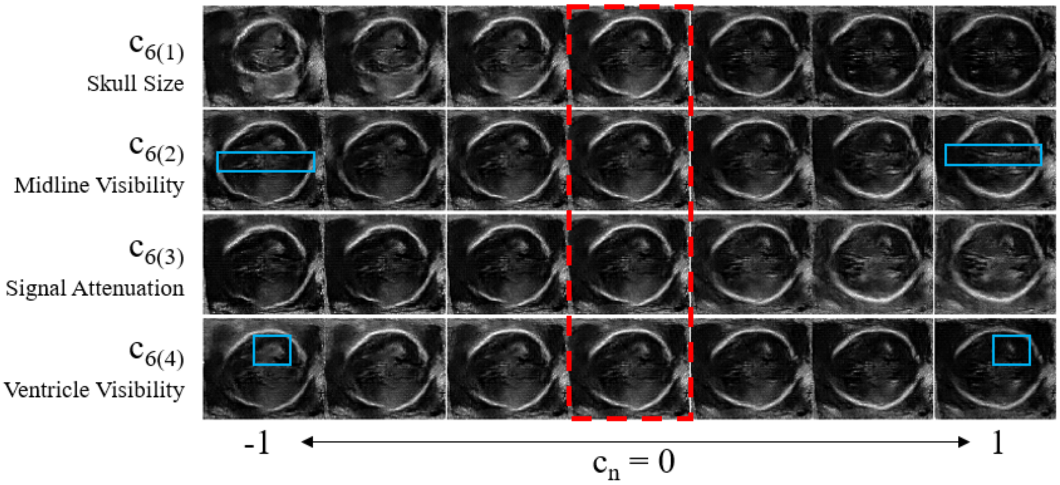

Fig. 3. A noise inputs z were used to generate a base image, highlighted in red. Images were generated from a Least Squares GAN network with six latent codes $c_6 \sim U(-1,1)$. Each row represents a latent code $c_{6(1-4)}$ that is uniformly changed from -1 to 1. Skull size, skull width, and the visibility of internal structures and image attenuation are captured within the latent codes and can be changed for a given base image. The anatomical changes of interest are highlighted in blue. The two latent codes left out did not correspond to interpretable changes in the generated image. Images are best viewed digitally.

with c_3, and ultrasound beam attenuation with c_4.

Of note is that these factors can be varied during generation for a given noise vector for different representations of the same base image. Arbitrary images of fetal ultrasound heads may therefore be generated by varying the noise vector z, and attributes such as anatomical presentation and image parameters can then be user-modified by changing the latent codes c_n.

In subsequent runs with the same network architecture, we found that similar human-interpretable results can be mapped to the latent codes. However, the mapping between the interpretation and the latent code change, as a consequence of the unsupervised training process.

4. CONCLUSION AND DISCUSSION

In this paper, we have proposed using an information maximizing generative adversarial network to generate synthetic fetal head ultrasound images. We show that this network architecture allows control of various image parameters and anatomical presentations for a given generator input, which is not possible with a traditional GAN. We additionally introduce an additional ultrasound-specific image generator quantitative metric named the Fréchet SonoNet Distance, and find that it correlates well with image quality. Successful controllable synthesis of high quality fetal ultrasound data has the potential to fast track training of sonographers where large, labelled datasets are not available.

5. REFERENCES

[1] J. Gu and Y. Jing, “Modeling of wave propagation for medical ultrasound: a review,” T-UFFC, 2015.

[2] Francis Tom and Debdoot Sheet, “Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning,” in ISBI, 2018.

[3] Yipeng Hu et al., “Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks,” in MICCAI RAMBO, 2017.

[4] Xi. Chen et al., “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” in NeuRIPS. 2016.

[5] Martin Heusel et al., “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” in NeuRIPS, 2017.

[6] Takeru Miyato et al., “Spectral normalization for generative adversarial networks,” in ICLR, 2018.

[7] Tim Salimans et al., “Improved techniques for training gans,” in NeuRIPS, 2016.

[8] Xudong Mao et al., “On the effectiveness of least squares generative adversarial networks,” TPAMI, 2018.

[9] Christian F Baumgartner and otheres, “Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound,” TMI, 2017.