

Learning-Based Approach for Real-Time Versatile Ultrasound Computed Tomography

1st Seokhwan, Oh
Electrical Engineering
Korea Advanced Institute of
Science and Technology
Daejeon, Korea
joseph9337@kaist.ac.kr

2nd Gibbeum, Lee
Electrical Engineering
Korea Advanced Institute of
Science and Technology
Daejeon, Korea
pirensisco@kaist.ac.kr

3rd Myeong-Gee, Kim
Electrical Engineering
Korea Advanced Institute of
Science and Technology
Daejeon, Korea
suhakprince@kaist.ac.kr

Hyeon-Min, Bae
Electrical Engineering
Korea Advanced Institute of
Science and Technology
Daejeon, Korea
hmbae@kaist.ac.kr

Abstract— *Ultrasound computed tomography, implemented on the basis of circular transducer array in general, is known to provide quantitative characteristics of imaging objects. However, existing UCT solutions are not popular in clinical applications due to their usage restrictions and long reconstruction time. In this paper, we propose a versatile UCT system employing two facing transducer arrays to gather and analyze bidirectional reflected and traversed waves. A Neural Network (NN) approach is incorporated to reduce the acquisition time for real-time image reconstruction. To gradually supplement the details of lesions, a refinement network structure is proposed and the quality of reconstructed image exceeds the quality of conventional systems when judged by diverse test metrics*

Keywords— *ultrasound computed tomography, UCT, Deep learning, ultrasound imaging system, biomedical imaging, ultrasound diffraction tomography*

I. INTRODUCTION

Ultrasound is a non-invasive, safe, and widely available medical imaging modality. The majority of prior-arts in ultrasound imaging have focused on B-mode ultrasonography which utilizes reflection and scattering of the ultrasound waves. The operator dependency and the fact that the reconstructed image offers only qualitative information are two major limitations of conventional b-mode ultrasonography. Ultrasound Computed Tomography (USCT) that provides quantitative characteristics of a tissue such as the sound speed and attenuation overcomes the fundamental drawbacks of B-mode ultrasonography and has shown great potential in the area of cancer localization, where malignant lesions have abnormal acoustic properties [1].

Despite its safety and specificity, USCT suffers from several major weakness: limited versatility and computational complexity. Firstly, conventional USCT has a ring-shaped transducer alignment to acquire diffracted ultrasound wave in all directions. However, such transducer alignments exerts limitation on the usage of USCT; USCT can only be used on protruding parts of the body such as breast. Secondly, because the image reconstruction of USCT relies on the acoustic wave equation, numerically-intensive iterative computation is necessary [2]-[4], which results in significant processing time.

In addition, the convergence of the iterative methods is not guaranteed for imaging targets with high-contrast [3].

In order to overcome aforementioned issues, we propose a neural network (NN)-assisted USCT system based on two facing linear transducer arrays. The proposed USCT method employing learning-based deep neural networks (DNNs) enables a sensory-to-quantitative-image synthesis with reduced processing time and enhanced accuracy [5]. The proposed DNN featuring a coarse-to-fine network structure recovers high-frequency details and textures such as brain creases and vessels irrecoverable in conventional DNNs. Additionally, a generator-discriminator structure is applied to the proposed DNN to synthesize more photo-realistic image that cannot be represented by PSNR [6]-[8].

II. MEASUREMENT MODELS AND DATA ACQUISITION

The physics of propagating acoustic wave is governed by the Helmholtz's equation. The lossless propagation of ultrasound in biological tissue can be described as [9]

$$(\nabla^2 + k_0^2)p(\vec{r}) = O(\vec{r})p(\vec{r}) \\ = (k_0^2 - k(\vec{r})^2)p(\vec{r}), \quad (2.1)$$

where \vec{r} is the position vector, $k = w/c$ is the wavenumber at frequency w and sound speed c , and $O(\vec{r})$ is the object profile that we aim to reconstruct. The inhomogeneous wave equation is formulated with the Green's function $G(k_0, \vec{r})$, and the steady-state pressure $p^{steady}(\vec{r})$ generated in the region Ω with the incident pressure $p^{inc}(\vec{r})$ is [4]

$$p^{steady}(\vec{r}) = p^{sc}(\vec{r}) + p^{inc}(\vec{r}) \\ = \int_{\Omega} O(r)G(k_0, \vec{r} - r) p^{steady}(r)dr + p^{inc}(\vec{r}). \quad (2.2)$$

In two-dimensional (2-D) area, the Green's function from (2.2) can be given by

$$G(k_0, \vec{r}) = -\frac{j}{4}H_0^{(2)}(k_0|\vec{r}|), \quad (2.3)$$

where $H_0^{(2)}$ denotes the Hankel function.

A. Measurement System

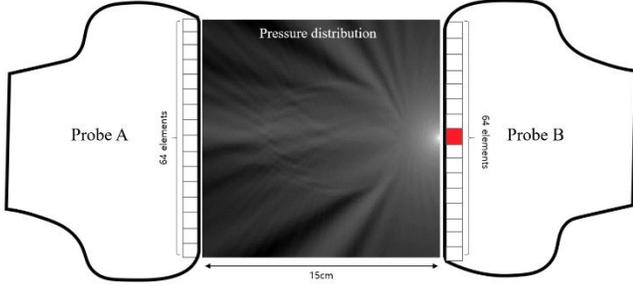


Figure 1. Top-view of the UCT system and actual steady-state pressure distribution of MRI-derived phantom.

Figure 1 illustrates the UCT system configuration applied in this work. Two facing linear transducer arrays, each with 128 piezoelectric elements, are employed to measure steady-state ultrasound pressure. Each transducer alternately emits a 0.3MHz continuous sinusoidal wave. The scattered pressure p_{trans}^{sc} measured by each transducer element can be described as

$$p_{trans}^{sc}(r^{tr}) = \int_{\Omega} O(r)G(k_0, r^{tr} - r) p^{steady}(r)dr, \quad (2.4)$$

where r^{tr} is the position vector of the transducer. The distribution of steady-state pressure can be obtained by solving (2.2) using the generalized minimal residual method [10].

B. Dataset

The training dataset for the proposed NN-based ultrasound system is created by gathering the magnitude and the phase of the scattered pressure of object phantoms measured at each transducer element. 2-D sliced MRI-derived numerical brain data obtained from Multimodal Brain Tumor Segmentation Challenge [11] are used as the object phantoms. The sound velocity of the brain phantoms was mapped between 1484 m/s and 1573 m/s depending on the parts of the brain. The 2-D pressure map for each phantom was generated by solving (2.2) and (2.4). The size of the pressure matrix is 128x128 where each column and row represents the transmitting element and receiving element, respectively. In order to utilize magnitude and phase information of the scattered pressure, the real and imaginary parts of the pressure matrix are separately handled in the NN. The pressure matrices obtained from 2400 2-D phantoms originated from 80 brains (60 for the train, 20 for the test) are used for the network training.

III. RECONSTRUCTION METHOD

In this section, the architecture of the proposed ultrasound imaging networks is presented. For the quantitative comparison of the proposed data-driven approach, a novel model-driven reconstruction algorithm is also proposed.

A. Iterative analytic reconstruction

The iterative analytic reconstruction method consists of two steps:

- 1) Reconstruct $O(r)$ by solving equation (2.2),
- 2) Improve prediction of the estimated steady-state pressure \hat{p}^{steady} with calculated $O(r)$.

The procedure starts by initializing \hat{p}^{steady} as a known free-space pressure distribution p^{inc} . Through iteration, \hat{p}^{steady} approximates the actual pressure distribution p^{steady} , and the accuracy of reconstructed object profile $O(r)$ improves. TVAL3 [12], a form of total variation minimization algorithm [13], is used for the reconstruction of $O(r)$, and the GMRES solver is adopted for the prediction of \hat{p}^{steady} .

B. Learning-Based reconstruction

It is well known that a network with the objectives of minimizing L1 and L2 losses often results in blurred images [14]. For detailed reconstruction of high-frequency image components, we propose an ultrasound imaging refinement network, referred to as a UINet, for the image synthesis. The UINet consists of 1) ultrasound encoding network that extracts useful features from real and imaginary pressure distribution, and 2) image decoder network that reconstructs the object profile from the encoded ultrasound data through a cascade refinement structure.

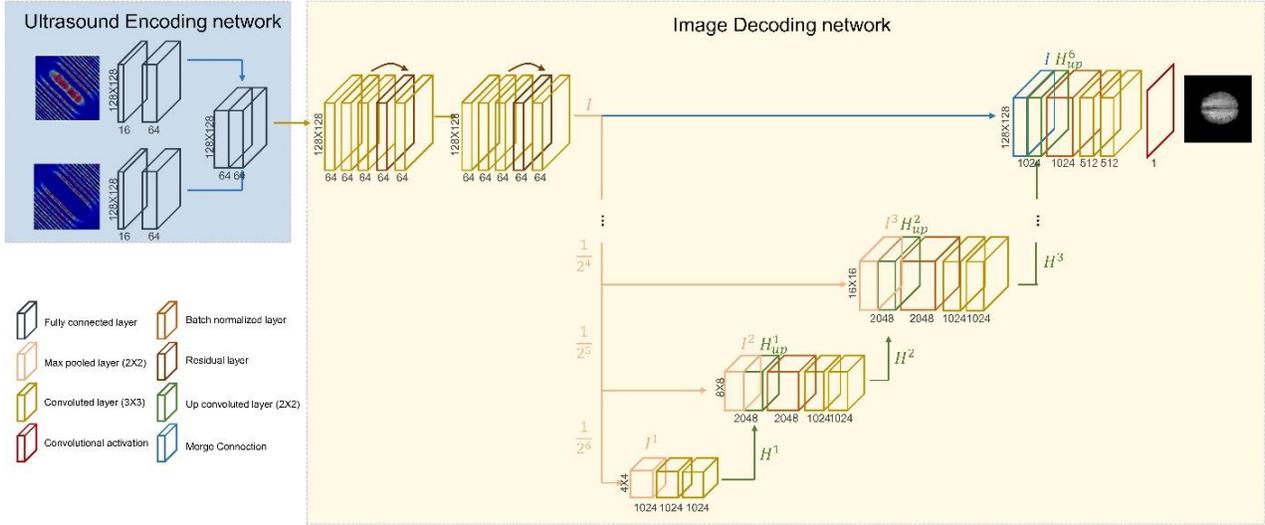
In addition, a generative adversarial network that is known to generate plausible-looking natural images using generator-discriminator structure [6, 15, 16] is also applied to UINet for further enhancement. The aforementioned Ultrasound imaging refinement generative adversarial network (UIGAN) is implemented by adding a discriminator to the UINet.

C. Network Architecture

The ultrasound encoder network extracts meaningful features from the measured pressure. Each encoder receives 128x128 real and imaginary ultrasound pressure matrix. The encoder network has 16 channels of fully connected layer with the ReLU activation functions followed by the same layer with 64 channels. The outputs of each encoder are concatenated and forwarded to the image decoder.

The image decoder network synthesizes the object profile from the encoded ultrasound features. The UINet aims at reconstructing the high-frequency details of object phantoms that can be lost in conventional encoder-decoder network architectures. Inspired from the Cascade Refinement Network [17], the proposed image decoder recovers the image details via progressive detail refinement process. The network starts with a series of 3x3 convolutional blocks and residual blocks [18] as shown in Fig. 2(a). Then, the filtered feature (I) is resized to a feature with lower resolution through a series of max-pooling. In our implementation, a 4x4 low-resolution features (I^1) are generated through 6 layers of max pooling. Then, I^1 is applied to the first refinement module (R^1) that is composed of convolutional layers, for the generation of refined features ($H^1 = R^1(I^1)$). The refined features are then convolutional up-sampled to the higher resolution (H^1_{up}) and concatenated with I^2 , a 5 times max-pooled output of the I . The refinement module receives the concatenated input and synthesizes a higher resolution output ($H^2 = R^2(I^2, H^1_{up})$). This procedure is repeated until the desired resolution is achieved and the signal propagation paths can deliver low level features I to high level refined features with limited loss on information. The number of

(a) UINet Architecture



(b) Discriminator Architecture

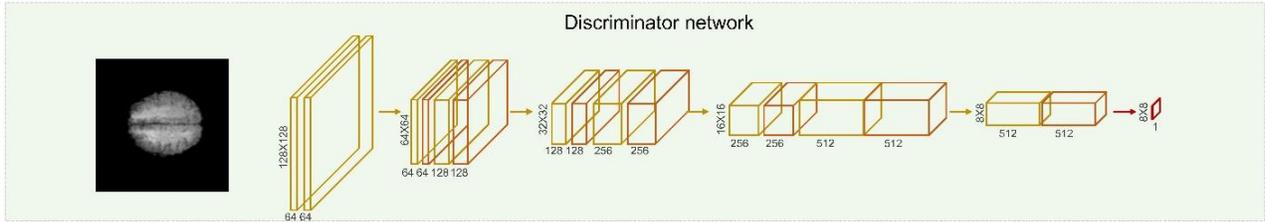


Figure 2 Architecture of the ultrasound encoder, image decoder, and discriminator network

refinement modules is set to 7 and 128x128 output features (H^7) are generated. For the final layer, 1x1 convolution is used to synthesize the object profile from H^7 .

The discriminator structure is applied to UIGAN for the generation of realistic natural images. While maintaining the generator’s loss function for the accuracy of object profile, the discriminator adds a perceptual loss to generate more plausible images. The discriminator is composed of convolutional layers and batch normalization layers [19]. For high-frequency image reconstruction, the Markovian Discriminator [16] architecture is employed. The markovian discriminator classifies local image patches as real or fake, opposed to verifying the entire image.

D. Training Details

The objective function of the UINet is

$$G^* = \arg \min_G E_{x,y} [\|y - G(x)\|^2], \quad (3.1)$$

wherein G^* tries to minimize the mean-squared difference between the ground truth (y) and the generated output $G(x)$ from the corresponding input (x). Unlike standard GAN, the proposed UIGAN maintains L2-distance of the generator to strictly regulate the output close to the ground-truth. In addition, the UIGAN includes a discriminator loss to the L2-distance to create undistinguishable outputs by the discriminator network D . Thus, the objective function of the UIGAN is

$$G^* = \arg \min_G \max_D E_{x,y} (\|y - G(x)\|^2) + \lambda (E_x [1 - D(G(x))] + E_y [D(y)]). \quad (3.2)$$

Entire networks are trained to solve an objective function using the Adam solver with the learning rate of 0.0001 and the momentum parameters of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. For the training of UIGAN, standard GAN training scheme [20] is employed, wherein the discriminator and the generator are trained alternately.

IV. RESULTS AND DISCUSSION

We compare the performance of the UINet and the UIGAN to that of TVAL3 under two-facing transducer array configuration. Additionally, image reconstruction performance of conventional USCT under circular transducer array configuration is presented for comparison. The networks are evaluated using 600 object profiles from 20 different brains. For

Table 1 Quantitative performance comparison of model-driven and data-driven approaches in phantom image reconstruction

	Model driven approach		Data driven approach		
	Circular array*		Proposed both-sided array		
	Uniform sample	Compressed sensing	TVAL3	UINet	UIGAN
MNAE	0.2~1.2	0.18~0.65	0.e0	0.26	0.3
SSIM	-	-	0.7242	0.8217	0.8075
Time	-	-	432s/iter	0.0316s	0.03628

*MNAE depends on the acquisition reduction factor [3] The evaluated object profile are 64x64 resolution.

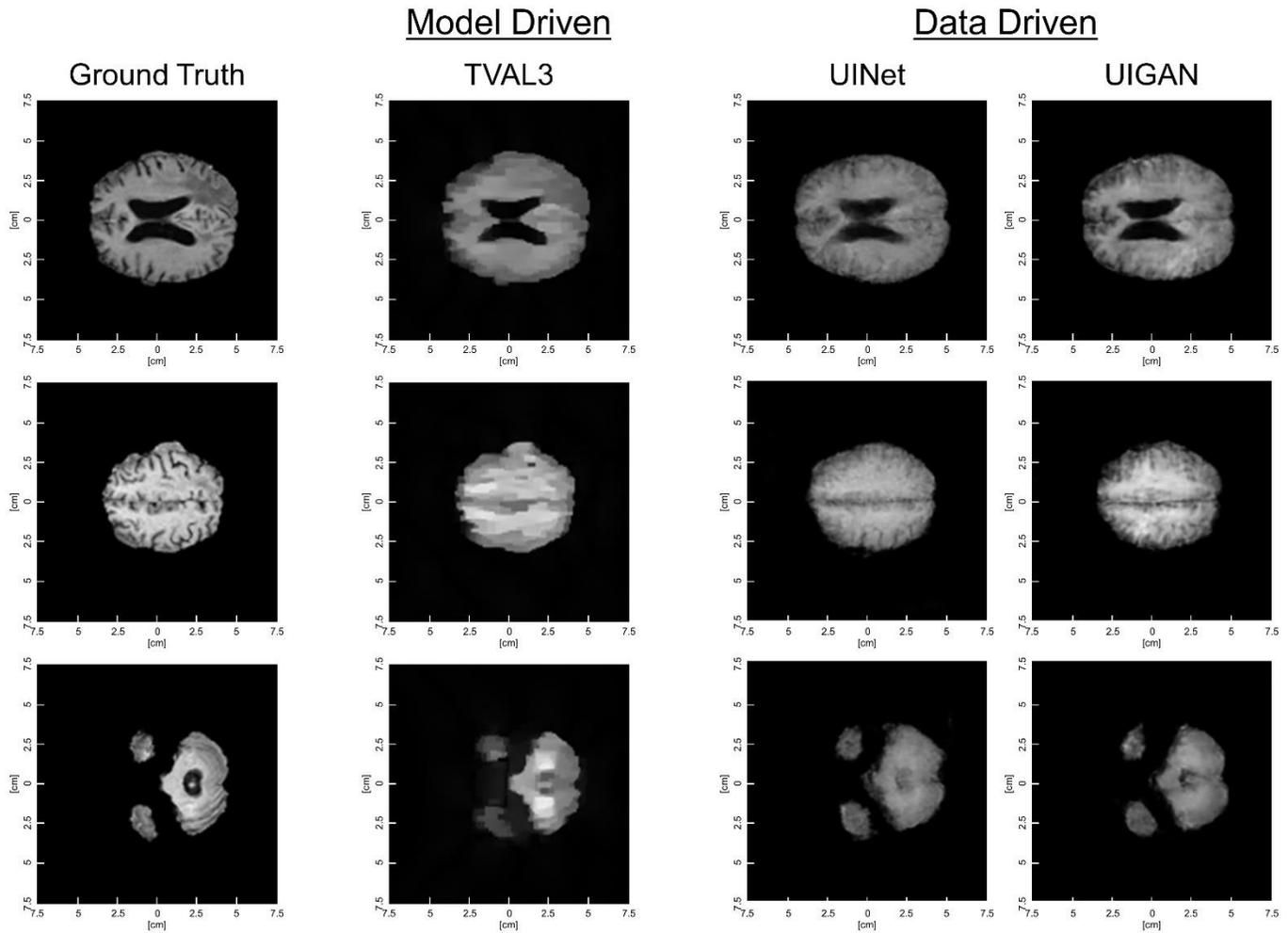


Figure 3 Reconstruction results of model driven approach and data driven approaches in MRI derived numerical brain phantoms

the quantitative assessment of the network, the mean normalized absolute error (MNAE) and the structural similarity (SSIM) index metrics are investigated. Table I shows that two-facing transducer arrays can reconstruct reliable object profile even with limited angular information. The proposed UCT system shows similar MNAE as compared to the existing complex circular array UCT for the restoration of higher resolution (128x128) images while offering versatility without usage restrictions. TVAL3 algorithm outperforms NN approaches in MNAE criteria but the underperforms in SSIM, which indicates that proposed network structure is effective in extracting the high-frequency details of the image. Moreover, the NN approaches achieves significant reduction in the reconstruction time of less than 0.03628s, which opens the new possibility of real-time UCT imaging.

Figure 3 shows the reconstruction results of the analytical and NN-based approaches. TVAL3 solver results in blurry image without revealing the brain crease. On the contrary, the proposed network approaches recover the details of the brain phantom. It is also clear from Fig. 3 that the UIGAN synthesizes high contrast images as compared to the UINet. Such performance improvement stems from the fact that UINet selects a final image with minimal Euclidean distance by

averaging the possible output candidates. The high contrast realistic output of the UIGAN is attributed to the fact that the discriminator of the UIGAN ignores the low-contrast images as fake images.

V. CONCLUSION

In this work, a learning-based high-resolution UCT scheme using two facing transducer arrays is presented. The proposed learning-based approaches reconstruct similar quality image as compared to conventional complex USCT schemes based on a circular transducer array while significantly reducing the reconstruction time. Moreover, the progressive refinement network structure is promising in overcome the high-frequency loss issue in iterative analytic reconstruction schemes.

REFERENCES

- [1] Ranger, Bryan. "Breast ultrasound tomography versus magnetic resonance imaging for clinical display of anatomy and tumor rendering: Preliminary results." (2016).
- [2] Wang, Kun, et al. "Waveform inversion with source encoding for breast sound speed reconstruction in ultrasound computed tomography." *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 62.3 (2015): 475-493

- [3] van Sloun, Ruud JG, et al. "Compressed sensing for beamformed Ultrasound computed tomography." 2015 IEEE International Ultrasonics Symposium (IUS). IEEE, 2015.
- [4] van Sloun, Ruud, et al. "Compressed sensing for ultrasound computed tomography." IEEE Transactions on Biomedical Engineering 62.6 (2015): 1660-1664.
- [5] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. 2014.
- [6] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [7] Gupta, Prateek, et al. "A modified PSNR metric based on HVS for quality assessment of color images." 2011 International Conference on Communication and Industrial Application. IEEE, 2011.
- [8] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing 13.4 (2004): 600-612.
- [9] Robinson, Brent S., and James F. Greenleaf. "The scattering of ultrasound by cylinders: Implications for diffraction tomography." The Journal of the Acoustical Society of America 80.1 (1986): 40-49.
- [10] Saad, Youcef, and Martin H. Schultz. "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems." SIAM Journal on scientific and statistical computing 7.3 (1986): 856-869.
- [11] Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." IEEE transactions on medical imaging 34.10 (2014): 1993-2024.
- [12] C. Li, W. Yin, and Y. Zhang, "TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithm," 2009, [online] available at <http://www.caam.rice.edu/~optimization/L1/TVAL3/>.
- [13] Zhang, Jian, et al. "Improved total variation based image compressive sensing recovery by nonlocal regularization." 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013). IEEE, 2013.
- [14] A. B. L. Larsen, S. K. Sønderby, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. In ICML, 2016
- [15] Wang, Ting-Chun, et al. "High-resolution image synthesis and semantic manipulation with conditional gans." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
- [16] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [17] Wang, Ting-Chun, et al. "High-resolution image synthesis and semantic manipulation with conditional gans." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
- [18] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016..
- [19] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift."
- [20] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.