



Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction *In Vivo*

Hengrong Lan¹, Kang Zhou^{1,2}, Changchun Yang¹, Jun Cheng²,
Jiang Liu^{2,3}, Shenghua Gao^{1(✉)}, and Fei Gao^{1(✉)}

¹ School of Information Science and Technology, ShanghaiTech University,
Shanghai 201210, China

{gaoshh, gaofei}@shanghaitech.edu.cn

² Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences,
Ningbo 315201, China

³ Department of Computer Science and Engineering,
Southern University of Science and Technology, Guangdong 518055, China

Abstract. Photoacoustic computed tomography (PACT) breaks through the depth restriction in optical imaging, and the contrast restriction in ultrasound imaging, which is achieved by receiving thermoelastically induced ultrasound signal triggered by an ultrashort laser pulse. The photoacoustic (PA) images reconstructed from the raw PA signals usually utilize conventional reconstruction algorithms, e.g. filtered back-projection. However, the performance of conventional reconstruction algorithms is usually limited by complex and uncertain physical parameters due to heterogeneous tissue structure. In recent years, deep learning has emerged to show great potential in the reconstruction problem. In this work, for the first time to our best knowledge, we propose to infuse the classical signal processing and certified knowledge into the deep learning for PA imaging reconstruction. Specifically, we make these contributions to propose a novel Knowledge Infusion Generative Adversarial Network (Ki-GAN) architecture that combines conventional delay-and-sum algorithm to reconstruct PA image. We train the network on a public clinical database. Our method shows better image reconstruction performance in cases of both full-sampled data and sparse-sampled data compared with state-of-the-art methods. Lastly, our proposed approach also shows high potential for other imaging modalities beyond PACT.

Keywords: Photoacoustic computed tomography · Generative adversarial network · Reconstruction · Knowledge infusion

Hengrong Lan and Kang Zhou contributed equally to this work.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-32239-7_31) contains supplementary material, which is available to authorized users.

© Springer Nature Switzerland AG 2019

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 273–281, 2019.

https://doi.org/10.1007/978-3-030-32239-7_31

1 Introduction

Photoacoustic imaging (PAI) is a hybrid imaging technique based on the photoacoustic (PA) effect that is excited by a laser pulse. The applications of PAI have covered many interesting biomedical imaging areas, such as imaging of oxyhemoglobin saturation, melanoma or chromophore for both cancer diagnostics and treatment monitoring [1, 2]. More specifically, photoacoustic computed tomography (PACT) acquires PA signals using multi-element ultrasound array, and rebuilds the image from PA signals using the reconstruction algorithm. Conventional reconstruction methods, e.g. back-projection and time reversal are widely applied in PA image reconstruction, which however suffers from artifacts and information loss.

Recently, deep learning based approach has been developed to resolve the inverse problem, which basically includes two schemes: training a model that map the raw signals to final image, or ameliorating the result of the conventional reconstruction algorithm (i.e. post-processing) [3, 4]. The input of the former scheme contains all of the physical informations about the target, but there is a huge gap between raw signals and final PA image, which lacks textural information to provide direct physical relationship. On the other hand, the input of the latter scheme reserves the direct physical relationship by conventional reconstruction method, which however only approximates the ground-truth and loses some detailed information. Furthermore, these methods rely on the brute force of big data.

To go beyond brute force and utilize the merits of both schemes, in this paper, we propose Knowledge Infusion Generative Adversarial Network (Ki-GAN) to boost reconstruction performance. The knowledge comes from two sources: (1) Traditional signal processing inspiration (e.g. raw PA signals); (2) Traditional certified reconstruction algorithm (e.g. PA images reconstructed from back-projection). We attempt to introduce signal processing knowledge for framework design and embedding certified knowledge into image feature for PA imaging reconstruction. We enlighten an innovative and effective convolutional kernel to bridge the gap between raw signals and image, and propose a novel Ki-GAN to reconstruct the PA image, which merges the conventional reconstruction algorithms (e.g. delay-and-sum) as a part of the architecture and is fed with the raw PA data as input. To infuse the knowledge, our primary contribution is to suggest a novel framework to infuse the signal processing knowledge and conventional reconstruction knowledge, and achieve better results compared with prior work. Hemoglobin is the main contrast of biological tissue in PAI, and we premise the vessel is the prime target for reconstruction. A set of vessels data is used for training and validating the architecture, which utilizes MATLAB to load the treated segmented blood vessels from the public clinical database [5]. To ensure the performance in different conditions, a set of sparse data is used to test our approach. Furthermore, the *in vivo* PA imaging experiments have also been performed to validate our approach further. The code is available at <https://github.com/chenyilan/MICCAI19-Ki-GAN>.

2 Methods

In this work, we mainly focus on how to infuse more knowledge into the deep learning based imaging framework in Fig. 1. In the following, we will introduce our proposed solution to bridge the gap between PA signals and image by designing a new Auto-Encoder (AE) with signal processing knowledge and embedding a certified algorithm into image feature.

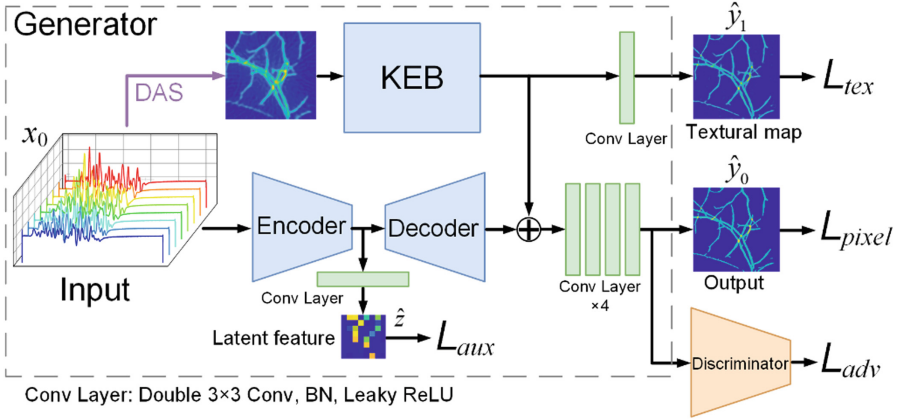


Fig. 1. The overall architecture of Ki-GAN; KEB is convolutional layers; DAS is a conventional reconstruction algorithm. KEB: knowledge embedding branch; DAS: delay-and-sum.

2.1 Introducing Signal Processing Knowledge for Auto-Encoder Design

As a backbone of our proposed Ki-GAN, our designed AE consists of two parts: (1) to adapt the AE to the physical PA signals, we propose the Auto-Encoder with PA Signal Sampling Inspired Kernel; (2) to further constrain the latent feature between PA signal and image, we introduce the Auto-Encoder with Image Feature Supervision.

PA Signal Sampling Inspired Kernel. As aforementioned, since the skip connection of U-Net for raw data is harmful to the decoder, we propose to adapt Auto-Encoder as our cornerstone rather than U-Net [6]. The input raw data has two dimensions: the transducer channel and the signal temporal distribution, where the dimension of signal temporal distribution is much larger than the transducer channel dimension. The large local receptive field is needed to identify the signal with larger length. Therefore, we apply convolutional kernel with 20×3 size to replace 3×3 size in the bottom layer of the encoder and kernel with 5×3 size to replace 3×3 size in other layers of the encoder, which is called PA Signal Sampling Inspired Kernel (PSSIK). We apply commonly-used MSE (mean square error) loss as our pixel loss, which is expressed as follow:

$$L_{pixel} = \|y - \hat{y}_0\|_F^2 \quad (1)$$

where y , \hat{y}_0 denote ground-truth and output image respectively.

Image Feature Supervision. To improve low-level features of the raw PA signal in the encoder, we assign auxiliary supervisions directly to the output of the encoder network (i.e. \hat{z} in Fig. 1). The auxiliary loss is computed as follow:

$$L_{aux} = \|z - \hat{z}\|_F^2 = \|f(y) - \hat{z}\|_F^2 \tag{2}$$

where \hat{z} denotes latent feature, and $z = f(y)$ denotes latent feature of ground-truth image y , where f denotes down-sampling operation on y . The more details of our proposed Auto-Encoder are provided in the supplementary materials.

2.2 Embedding Certified Knowledge into Image Feature

As mentioned above, we propose a solution to bridge the gap between the PA signal and image. Considering the raw PA signals are lacking texture information, we further introduce Knowledge Embedding Branch (KEB) to provide the textural information by converting the certified knowledge from DAS, as shown in Fig. 1. The result of DAS is an imperfect image which is confused by incident artifacts. To adapt the different size of vessels and eliminate the artifacts, three Texture Blocks containing two kernels with 3×3 and 1×1 constitute the KEB inspired by inception block [7] as shown in Fig. 2. Simultaneously, these blocks can maintain a fast computation and less parameters compared with a deep convolutional branch. To improve the KEB, we utilize the textural loss to restrict the textural maps, which is expressed as follow:

$$L_{tex} = \|y - \hat{y}_1\|_F^2 \tag{3}$$

where \hat{y}_1 denotes the textural maps.

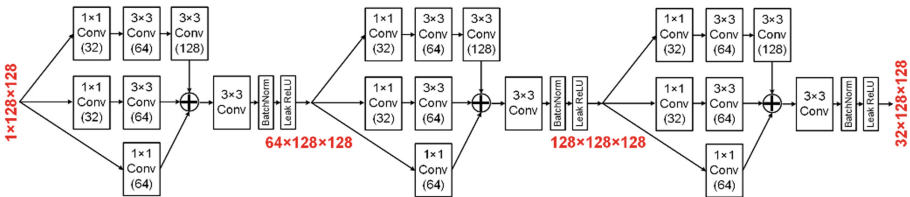


Fig. 2. Three Texture Blocks constitute the Knowledge Embedding Branch.

2.3 Ki-GAN: Knowledge Infusion GAN

As shown in Fig. 1, by integrating above two methods, we further propose to make use of adversarial learning [8] to improve the ineffectiveness of convolutional neural networks in modeling correlation between the PA signals and the detailed vessel reconstruction. On the other hand, pixel loss is an average operation, that is to say, the pixel loss will result in image smooth and blurring. Therefore, the generative adversarial

network is needed. The generator outputs the reconstructed image \hat{y}_0 , and is also restricted by adversarial loss, which is calculated as follow:

$$L_{advG} = \|D(y) - D(\hat{y}_0)\|_F^2 \quad (4)$$

where D is a function that outputs an intermediate layer of the discriminator. Our discriminator is inspired by *PatchGAN* [9] to penalize the texture at the scale of patches (see Fig. S2 in supplementary materials), and the adversarial loss can be expressed as:

$$L_{advD} = -\mathbb{E}_{x_0, y} [\log D(x_0, y)] - \mathbb{E}_{x_0, \hat{y}_0} [\log(1 - D(x_0, \hat{y}_0))] \quad (5)$$

Finally, the generator network is trained by minimizing the total loss:

$$L_{total} = \lambda_{adv} L_{advG} + \lambda_{pix} L_{pixel} + \lambda_{aux} L_{aux} + \lambda_{tex} L_{tex} \quad (6)$$

where λ_{adv} , λ_{pix} , λ_{aux} and λ_{tex} are hyper-parameters.

3 Experiments and Results

3.1 Datasets and Evaluation

The proposed approach requires a large amount of data for training the model, which is difficult to obtain since the current PAI equipment is still in the stage translating from preclinical to clinical application, and not available in the clinic. Therefore, we convert the publicly available datasets of fundus oculi [5] as the photoacoustic initial pressure distribution. The toolbox of k-wave [10] in MATLAB reads the initial pressure to generate the raw PA signals for training.

The vessels are surrounded by 120 channels' transducers with 18 mm radius. For each PA signal, the total recorded points are 2560 with 150 MHz sampling rate. The pixel size of the initial pressure map is 128×128 , and the output of the network is 128×128 as well. The center frequency of the ultrasound transducer is set as 5 MHz with 80% bandwidth, and the propagation velocity of ultrasound is 1500 m/s.

As the original dataset is small, we adopt some preprocessing to expand the data volume. Firstly, the complete blood vessel of fundus oculi is segmented into four equal parts; and then randomly transform (e.g. rotations and transpositions) and superpose two segmented blood vessels. After a series of treatment, we can acquire excessive initial pressure for PA imaging and generate sufficient training data.

The whole dataset is composed of 4300 training samples and 500 test samples. Simultaneously a set of sparse data are also utilized to evaluate the proposed approach, which compresses the signal channels from 120 to 40. In addition, we also explored rat thigh experimental data for the purpose of verifying the validity of the proposed approach in animal imaging *in vivo*.

3.2 Training Details

The framework is implemented in PyTorch. The network is fed the 2560×120 raw data as input with setting batch size as 32. The generator can be trained by Eq. (6), the λ_{aux} and λ_{tex} are set as 0.5 for both, which are optimum choices compared with different values (we also list the performance of different parameters values in supplementary materials). The λ_{adv} and λ_{pix} are set as 0.04 and 1 respectively. The discriminator can be trained by Eq. (5). In our evaluation, the simple delay-and-sum (DAS) algorithm is chosen as the image textural knowledge infused into Ki-GAN.

The computing platform we used is a high-speed graphics computing workstation consisting of Intel Xeon E5-2690 (2.6 GHz) \times 2 central processing unit (CPU) and NVIDIA GTX 1080ti \times 4 graphics processing units (GPU). The time consumption of every batch is 0.795 s in training stage.

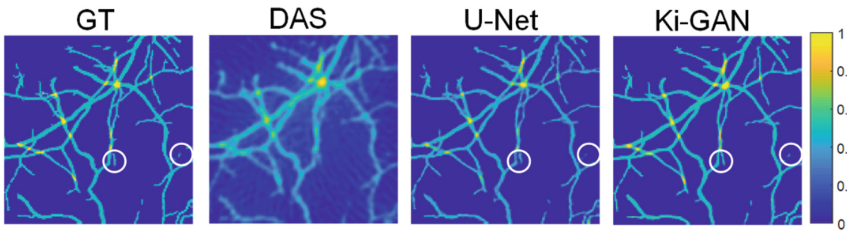


Fig. 3. Example of quantitative comparison using full-sampled data. From left to right: ground-truth, delay-and-sum, U-Net and Ki-GAN. The white circles indicate the local details.

3.3 Experimental Results

Full-Sampled Data. Figure 3 has shown a sample of results generated by different models, which comprise DAS, post-processing model (U-Net) and Ki-GAN. The images of DAS cannot avoid the artifacts albeit they preserve the vessel’s sketch indistinctly. The images from our proposed Ki-GAN are closer to ground-truth than U-Net. To further compare the performance of different approaches, we compare the results of the test set using three indexes of quantitative metrics: (1) structural similarity index (SSIM); (2) peak signal-to-noise ratio (PSNR); (3) signal-to-noise ratio (SNR). We also list the ablation studies beyond the methods shown in Fig. 3, including Auto-Encoder (AE#1), AE#1 with PSSIK (AE#2), AE#2 with Image Feature Supervision (AE#3), AE#3 with Embedded Certified Knowledge (AE#4), and U-Net inputting the raw data (U-Net¹). The quantitative comparison results are shown in Table 1, which indicate that our proposed Ki-GAN gets the upper hand compared with other models. Meanwhile, the ablation studies’ results also validate the effectiveness of different modules of our proposed method. More results of ablation studies and comparative experiments are provided in the supplementary materials (Fig. S4–5).

Table 1. Evaluation results of different models for the test sets (full-sampled data). U-Net¹: input the signals and resize to concatenation, U-Net²: input the result of DAS, AE#1: Auto-Encoder, AE#2: AE#1 with PSSIK, AE#3: AE#2 with Image Feature Supervision, AE#4: AE#3 with Embedded Certified Knowledge.

	DAS	U-Net ¹	U-Net ²	AE#1	AE#2	AE#3	AE#4	Ki-GAN
SSIM	0.2159	0.6453	0.8749	0.6519	0.6818	0.6931	0.9123	0.9285
PSNR	15.6176	18.4519	24.0175	18.7033	19.1193	19.1529	24.8951	25.5115
SNR	1.6386	4.3237	10.1285	4.7243	5.1403	5.1739	10.6161	11.5324

Sparse-Sampled Data. We further compare the performance of U-Net (post-processing) and Ki-GAN using the sparse data, which include only 40 channels' raw data to reconstruct the vessel image. It is noteworthy that we fill the zero data to keep the number of channel in 120 due to the fixed size of the input data. The qualitative comparison is shown in Fig. 4, which indicates that the performance of the proposed method is better than U-Net. The white circles marked three details in Fig. 4, showing that the result of Ki-GAN identifies with ground-truth image more closely compared with U-Net. The quantitative evaluation results also agree well with Fig. 4 as Table 2 showed.

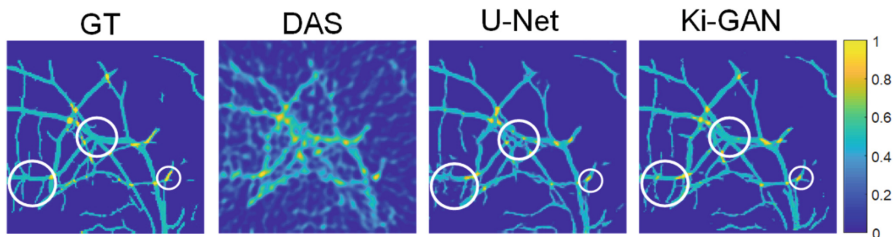


Fig. 4. Example of quantitative comparison using sparse-sampled data. From left to right: ground-truth, delay-and-sum, U-Net and Ki-GAN. The white circles indicate the local details.

Table 2. Evaluation results of different models for the test sets (sparse-sampled data).

	DAS	U-Net	Ki-GAN
SSIM	0.1842	0.8174	0.8617
PSNR	15.5123	21.348	22.7398
SNR	1.5333	7.4689	8.7607

In Vivo Data. Last but not least, the *in vivo* PA imaging experiments of a rat thigh have also been performed to validate our approach. Three methods including conventional iteration-based reconstruction, U-Net and Ki-GAN, are performed to illustrate the results in Fig. 5. It shows that our proposed model possesses a stronger contrast and fewer artifacts compared with other two methods in an insufficient training data set. The U-Net has a poor generalization performance compared with Ki-GAN in *in vivo* data, and suffers inevitable artifacts.

The time consumption of the iterative reconstruction algorithm with 10 iterations, U-Net and Ki-GAN are 331.51 s, 0.01 s and 0.025 s. The iterative algorithm depends on the repetitive calculation of the forward and backward models in the loop cycle. The image quality and time consumption show inevitable compromise in the iterative algorithm. U-Net shows the fastest mapping from image to image that consumes least time in three methods. The proposed Ki-GAN infuses the conventional reconstruction and raw-data based feature map. Meanwhile, the 0.025 s time consumption still sufficiently satisfies the requirement of real-time imaging for most clinical demands.

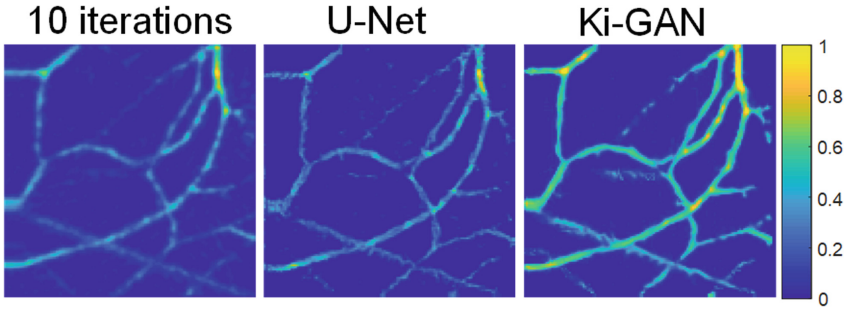


Fig. 5. The vessel imaging of rat thigh. From left to right: iterative algorithm with 10 iterations, U-Net and Ki-GAN.

4 Conclusion

Fast and accurate image reconstruction is a significant problem in PACT. In this paper, we propose a novel framework of knowledge infusion for reconstructing the PA image, which merges the conventional reconstruction with burgeoning deep learning. A novel Ki-GAN is proposed to rebuild the initial PA pressure of vessels. Ablation studies and comparative experiments show that the proposed model can perform very well in full-sampled data, sparse-sampled data, and *in vivo* experimental data. In the future work, we will try to exploit the real-time imaging system based on this method and extend the imaging dimension from 2D to 3D.

References

1. Wang, L.V., Hu, S.: Photoacoustic tomography: *in vivo* imaging from organelles to organs. *Science* **335**, 1458–1462 (2012)
2. Lan, H., Duan, T., Zhong, H., Zhou, M., Gao, F.: Photoacoustic classification of tumor model morphology based on support vector machine: a simulation and phantom study. *IEEE J. Sel. Top. Quantum Electron.* **25**, 1–9 (2019)
3. Cai, C., Deng, K., Ma, C., Luo, J.: End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt. Lett.* **43**, 2752–2755 (2018)
4. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**, 4509–4522 (2017)

5. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**, 501–509 (2004)
6. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
7. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence* (2017)
8. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680 (2014)
9. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134 (2017)
10. Treeby, B.E., Cox, B.T.: k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *J. Biomed. Opt.* **15**, 021314 (2010)