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

ZERO-SHOT MEDICAL IMAGE ARTIFACT REDUCTION

Yu-Jen Chen[1] *Yen-Jung Chang*[1] *Shao-Cheng Wen*[1] *Yiyu Shi*[2] *Xiaowei Xu*[3]
Tsung-Yi Ho[1] *Qianjun Jia*[3] *Meiping Huang*[3] *Jian Zhuang*[3]

[1]Department of Computer Science, National Tsing Hua University, Taiwan
[2]Department of Computer Science and Engineering, University of Notre Dame, USA
[3]Guangdong General Hospital, China

ABSTRACT

Medical images may contain various types of artifacts with different patterns and mixtures, which depend on many factors such as scan setting, machine condition, patients' characteristics, surrounding environment, etc. However, existing deep learning based artifact reduction methods are restricted by their training set with specific predetermined artifact type and pattern. As such, they have limited clinical adoption. In this paper, we introduce a "Zero-Shot" medical image Artifact Reduction (ZSAR) framework, which leverages the power of deep learning but without using general pre-trained networks or any clean image reference. Specifically, we utilize the low internal visual entropy of an image and train a light-weight image-specific artifact reduction network to reduce artifacts in an image at test-time. We use Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) as vehicles to show that ZSAR can reduce artifacts better than the state-of-the-art both qualitatively and quantitatively, while using shorter test time. To the best of our knowledge, this is the first deep learning framework that reduces artifacts in medical images without using *a priori* training set.

Index Terms— Image denoising, Deep learning, Zero-Shot

1. INTRODUCTION

Deep learning [1, 2] has demonstrated its great power in artifact reduction, a fundamental task in medical image analysis [3, 4, 5] to produce clean images for clinical diagnosis, decision making, and accurate quantitative image analysis. Existing deep learning based frameworks [6, 7, 8, 9] use training data sets that contain paired images (same images with and without artifacts) to learn the artifact features. Simulations are often needed to generate the data set for these methods, which may be different from clinical situations and lead to biased learning [6, 10]. To address this issue, [11] used cycle-consistent adversarial denoising network (CCADN) which no longer requires paired data.

However, all these methods still suffer from two mainstays: First, they require clean image references, which can be hard to obtain clinically. For example, motion artifacts in Magnetic Resonance Imaging (MRI) are almost always present due to the lengthy acquisition process [12]. In such situations, simulation is still the only way to generate the data set. Second, although the trained networks outperform non-learning based algorithms such as Block Matching 3D (BM3D) [13], they can only be applied to scenarios where the artifacts resemble what are in the training set, lacking the versatility that non-learning based methods can offer.

To attain the performance of deep learning based methods and the versatility of non-learning based ones, we introduce a "Zero-Shot" image-specific artifact reduction network (ZSAR), which builds upon deep learning yet does not require any clean image reference or *a priori* training data. Based on the key observation that most medical images have areas that contain artifacts on a relatively uniform background, the proposed approach could extract artifact pattern from image itself. At test-time, ZSAR extracts an artifact pattern directly and synthesizes paired image patches from input image to iteratively train a light-weight image-specific autoencoder for artifact reduction. Experimental results on clinical MRI and CT data with a variety of artifacts show that it outperforms the state-of-the-art methods using shorter execution time. To the best of our knowledge, ZSAR is the first deep learning based method that reduces artifacts in medical images without *a priori* training data.

2. METHODS

2.1. Overview

The main motivation of our work lies behind the fact that it is almost always possible to identify small regions of interests where significant artifacts exist over a relatively uniform background in any medical images. As such, it is possible to synthesize the paired dirty-clean patches from the exact image with artifacts to be reduced.

The overall framework of the proposed ZSAR is shown in Fig. 1, which is an iterative process. The framework works with 2D images, so 3D volumes are sliced first, similar to

This research was approved by the Research Ethics Committee of Guangdong General Hospital, Guangdong Academy of Medical Science with the protocol No. 20140316.

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

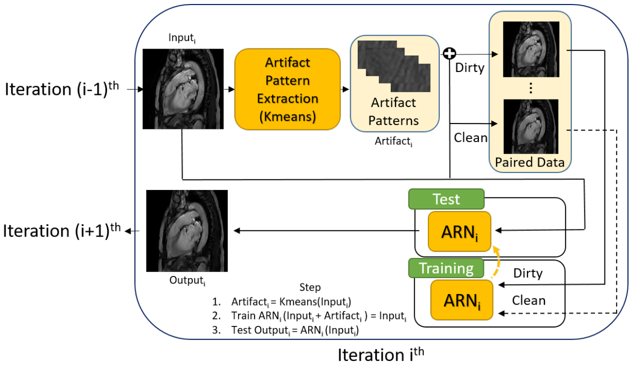


Fig. 1. The overall structure of ZSAR composed of Artifact Pattern Extraction and Artifact Reduction Network (ARN). Note that we treat the original input image as the output of "0^{th}" iteration.

many existing works [6]. For clarity, we call the phase where the model is trained to obtain the weights as "training", and the phase that applies the trained model to the input image to reduce artifacts as "test". Note that both phases are done on the spot for each specific input image and no pre-training is conducted.

For every iteration, ZSAR first extracts artifact patterns and synthesizes the paired dirty and clean images using the patterns (the details will be explained in Section 2.2). Note that the artifact pattern extraction in the 1^{st} iteration is different from those in the subsequent $(i+1)^{th}$ iterations $(i \geq 1)$. Later, the synthesized image is then used to train a lightweight artifact reduction network (ARN), which can reduce the artifact in input image (the details will be explained in Section 2.3). We terminate the iterative process when the artifact level (standard deviation) does not decrease. Our experiments show that the number of iterations needed is usually not more than four.

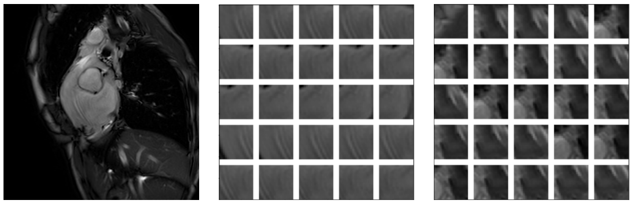
2.2. Artifact Pattern Extraction and Training Data Synthesis

For the 1^{st} iteration, since no clean reference image is provided, we extract the artifact pattern from the input image itself through an unsupervised approach. This is made possible based on the fact that for most artifacts in medical images, we can always identify areas where only artifacts exist [14, 15]. As such, we need to identify areas where the background is relatively uniform yet significant artifacts are present.

Towards this, we first crop the input image into patches with size 32×32. After that, an unsupervised clustering method, K-means [16] is applied. The main idea is to classify the patches into two clusters, one containing patches without structure boundaries (i.e., relatively uniform background), and the other containing patches with structure boundaries. Such a classification is possible as the patches in these two clusters will exhibit significant differences in terms of standard distributions of the pixel values: when structure boundaries are present, significant mean shift and large yet localized variations in pixel values can be observed. The feature of each patch is thus extracted as follows: the overall standard deviation of all the pixel values in the patch, and the mean value of all standard deviations extracted by a 8×8 sliding window. Fig. 2 shows an example of the clustering process. It can be clearly seen that one of the clusters contains patches with only uniform background (either with or without artifacts), while the other one contains all the patches with structure boundaries. As the patches in the former cluster always contain relatively uniform background, a zero-mean artifact pattern can be extracted by subtracting the mean pixel value of each patch. Note that in the patches without artifacts, the pattern extracted will just be empty. On the other hand, as long as some of the patches contain artifacts, our framework can utilize them to further synthesize the training data, which will be discussed later.

In the subsequent $(i+1)^{th}$ iteration $(i \geq 1)$, we observed that the difference between the clean image and the output image of $(i)^{th}$ iteration can be seen as the reduced artifact. The zero-mean artifact pattern is again generated by subtracting the mean pixel value of the difference.

To reflect artifacts of different intensities, we randomly scale each pattern following standard normal distribution. Those scaled artifact patterns are then superposed to random areas in the input image to form dirty images. Then we use the input image as the corresponding clean image so that paired dirty-clean data set, which contained only one dirty and one clean image is formed. Note that this synthesis process is conducted in every iteration.



(a) Input image (b) Artifact cluster (c) Other cluster

Fig. 2. An input image and examples of the two clusters after K-means is applied to the patches.

2.3. Artifact Reduction Network

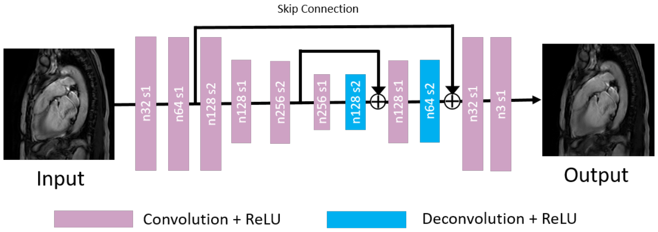


Fig. 3. Artifact Reduction Network (ARN) architecture. Note that n and s of each layer stand for the number of kernels and strides, respectively.

After the paired data is synthesized, it can be used to train any existing neural networks for artifact reduction. Considering the need of test-time training, we design a compacted network as shown in Fig. 3, which is formed by a 11-layer contextual autoencoder to reduce artifacts and restore the structural information. With the skip connection, these decoder layers can capture more contextual information extracted from different encoder layers. With such a light-weight network structure, it requires only a few epochs to converge. The pixel-wise mean square error (MSE) is used as the loss function to preserve structural and substance information:

$$Loss = L_{MSE}(O, G) \tag{1}$$

where O and G are the output of the contextual autoencoder and the clean image reference, respectively.

Through experiments, we find that ARN should be initialized and retrained in every iteration, which is more effective than incremental training based on the network from previous iterations. This is because each iteration is essentially a new artifact reduction procedure and the model only needs to learn the artifact level in the input of the current iteration. Also, in each iteration, only a single pair of images are used for training due to speed consideration. Since essentially, the same image is used during training and test, there is no overfitting concern.

3. EXPERIMENTS AND RESULTS

3.1. Cardiac Data Set and Evaluation Metrics

Our data set contains 17,844 2D MRI images (286 pulse sequences) from 11 patients and 48 3D cardiac CT volumes from 24 patients. Note that all MRI images are scanned by a 3T system.

All MRI and CT images are qualitatively evaluated by our radiologists on structural preservation and artifact level. For quantitative evaluation, due to the lack of ground truth, similarity based method cannot be applied in our case. For MRI, in addition to the mean of the pixel values in the most homogeneous areas, similar to [17, 18, 19] we divide the mean by the standard deviation of the pixel values in the area and use the resulting Signal-to-Noise ratio (SNR) as the metric. For CT, we follow existing work [20, 6] and select the most homogeneous area in regions of interest selected by our radiologists. The standard deviation (artifact level) of the pixels in the area should be as low as possible, and the mean (substance information) discrepancy after artifact reduction should not be too large to cause information loss.

3.2. Experimental Setup

ZSAR was implemented in Python3 with TensorFlow library. NVIDIA GeForce GTX 1080 Ti GPU was used to train and test the networks. For every convolution and deconvolution layer, Xavier initialization [21] was used for the kernels and the filter size is set to 3 and 4, respectively. Adam optimization [22] method was applied to train ARN by setting learning rate as 0.0005. Training phase was performed by minimizing loss function with the number of epoch and the number of iteration set to 1,000 and 4, respectively.

3.3. Comparisons with the state-of-the-art

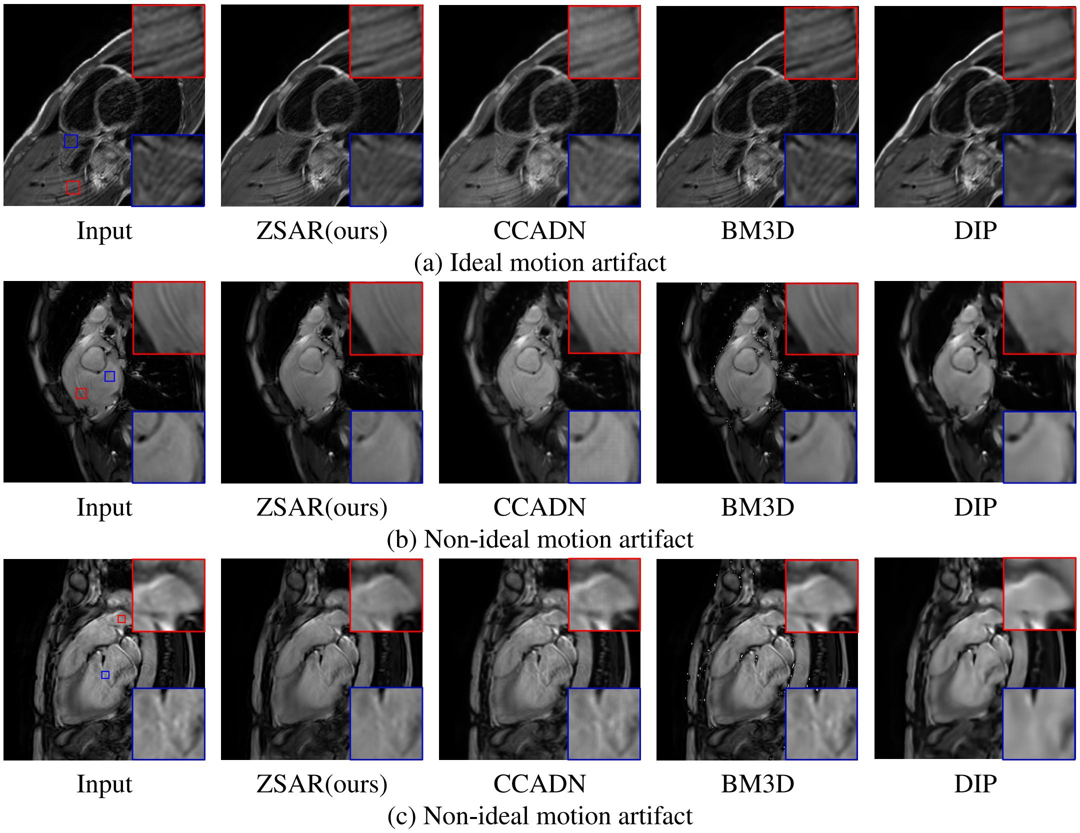
We compare ZSAR with CCADN, a state-of-the-art deep learning based method for medical image artifact reduction [11], which does not require paired training data. We also compare ZSAR with Deep image prior (DIP) [23], a state-of-the-art general-purpose denoising method, and a non-learning based algorithm BM3D. For CCADN and DIP, we follow the setting recommended in the paper. For each image, we tuned the parameters in BM3D to attain the best quality.

We start our experiments with the ideal scenario where the artifacts in both training set of CCADN and test MRI images contain motion artifact only. The qualitative results for CCADN, BM3D, DIP, and ZSAR are shown in Fig. 4 (a). All the methods preserve structures well, and ZSAR leads to the best motion artifact reduction. The corresponding statistics for the marked regions are reported in Table 1 (a). From the table, though CCADN has the largest SNR, it suffers from large mean discrepancy, which can be problematic. ZSAR achieves up to 50% higher SNR than BM3D. When comparing with DIP, ZSAR achieves similar SNR but less mean discrepancy.

Next, we study the non-ideal scenario where different artifact patterns or noise level of artifacts are absent from the training set of CCADN but appear in the test image. The results for MRI with different artifact patterns are shown in Fig. 4 (b-c) and Table 1 (b-c), respectively. Qualitatively, we can see that ZSAR outperforms CCADN and BM3D, while CCADN and DIP in case (b) result in brighter images (due to shifted mean pixel values). Quantitatively, ZSAR attains up to 77% and 74% higher SNR compared with CCADN and BM3D, respectively. In addition, since CCADN was trained in different scenario, in the region marked with red in case (c), it obtains SNR even smaller than the input image. Compared with DIP, ZSAR achieves lower SNR on region marked with red in case (b) and region marked with blue in case (c). However, this is in fact due to the large mean discrepancy in those two cases from DIP, which is not acceptable.

For CT, we follow similar setting as MRI in both scenarios. In ideal scenario, our experiments show that ZSAR achieves results comparable with that of CCADN and 14% lower standard deviation than BM3D. In non-ideal scenario, ZSAR beats CCADN and BM3D, achieving up to 19% and 25% lower standard deviation, respectively. On the other hand, DIP yields over 200% mean discrepancy in both scenarios, which is a critical problem for CT images.

Finally, to show that test-time training is feasible, we compare the average test time of ZSAR and the other three



(a) Ideal motion artifact

(b) Non-ideal motion artifact

(c) Non-ideal motion artifact

Fig. 4. Comparison using MRI test images under (a) ideal and both (b) and (c) non-ideal scenarios. Both contain motion artifacts but the pattern in (a) appeared in the training set but not in (b) and (c).

	(a) Red		(a) Blue		(b) Red		(b) Blue		(c) Red		(c) Blue	
	Mean	SNR	Mean	SNR	Mean	SNR	Mean	SNR	Mean	SNR	Mean	SNR
Input	477.3	6.2	381.9	7.6	994.4	9.5	1277.0	16.3	1265.0	13.7	1329.4	9.8
ZSAR	494.9	7.1	396.1	11.7	969.0	15.6	1279.8	25.9	1123.8	17.9	1206.9	17.6
CCADN	627.4	9.0	495.4	12.3	1280.4	12.8	1503.2	21.7	1204.3	13.0	1302.9	9.9
BM3D	476.1	6.2	380.3	7.8	989.0	9.8	1272.2	17.8	1260.7	14.2	1325.8	10.1
DIP	509.3	6.9	404.4	11.8	1089.5	18.7	1433.0	18.7	1333.7	15.8	1373.3	20.5

Table 1. Mean and SNR (Signal-to-Noise Ratio) for the largest homogeneous areas inside the marked regions of the MRI images in Fig. 4.

methods on the 3D MRI and CT images. The results are shown in Table 2. From the table, ZSAR requires less time than the other three methods despite the fact that it is trained on the spot for each input image. The fast speed of ZSAR is brought by two factors: 1) Its training usually converges within four iterations, and with few training data. Each iteration only takes about 1,000 epochs to converge. 2) It is much simpler than CCADN or DIP in structure and thus takes less time to test each 2D slice of the 3D images.

	ZSAR	CCADN	BM3D	DIP
MRI (360 slices)	192	1742	716	22308
CT (484 slices)	416	3533	1825	32641

Table 2. Test time comparison between ZSAR and the three methods, CCADN, BM3D, and DIP for 3D cardiac MRI (320×320) and CT (512×512) images (in sec.).

4. CONCLUSION

In this paper, we introduced ZSAR, a "Zero-Shot" medical image artifact reduction framework, which reduces artifacts in a medical image without using general pre-trained networks. Our method can be adapted for almost any medical images that contain varying or unknown artifacts, while previous state-of-the-art methods are restricted by the training data. Experimental results have shown that our framework can reduce artifacts qualitatively and quantitatively better than the state-of-the-art, using shorter test time.

5. REFERENCES

[1] Tianchen Wang, Jinjun Xiong, Xiaowei Xu, Meng Jiang, Haiyun Yuan, Meiping Huang, Jian Zhuang, and Yiyu Shi, “Msu-net: Multiscale statistical u-net for real-time 3d cardiac mri video segmentation,” in *MICCAI*. Springer, 2019, pp. 614–622.

[2] Zihao Liu, Xiaowei Xu, Tao Liu, Qi Liu, Yanzhi Wang, Yiyu Shi, Wujie Wen, Meiping Huang, Haiyun Yuan, and Jian Zhuang, “Machine vision guided 3d medical image compression for efficient transmission and accurate segmentation in the clouds,” in *CVPR*, 2019, pp. 12687–12696.

[3] Xiaowei Xu, Tianchen Wang, Yiyu Shi, Haiyun Yuan, Qianjun Jia, Meiping Huang, and Jian Zhuang, “Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching,” in *MICCAI*. Springer, 2019, pp. 477–485.

[4] Xiaowei Xu, Qing Lu, Lin Yang, Sharon Hu, Danny Chen, Yu Hu, and Yiyu Shi, “Quantization of fully convolutional networks for accurate biomedical image segmentation,” in *CVPR*, 2018, pp. 8300–8308.

[5] Xiaowei Xu, Tianchen Wang, Dewen Zeng, Yiyu Shi, Qianjun Jia, Haiyun Yuan, Meiping Huang, and Jian Zhuang, “Accurate congenital heart disease model generation for 3d printing,” *Proc. of IEEE International Workshop in Signal Processing Systems, Nanjing, China*, 2019.

[6] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang, “Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss,” *IEEE transactions on medical imaging*, 2018.

[7] Dongsheng Jiang, Weiqiang Dou, Luc Vosters, Xiayu Xu, Yue Sun, and Tao Tan, “Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network,” *Japanese journal of radiology*, vol. 36, no. 9, pp. 566–574, 2018.

[8] Nimu Yuan, Jian Zhou, and Jinyi Qi, “Low-dose ct image denoising without high-dose reference images,” in *IMF3DIRRNM*. International Society for Optics and Photonics, 2019, vol. 11072, p. 110721C.

[9] Xin Yi and Paul Babyn, “Sharpness-aware low-dose ct denoising using conditional generative adversarial network,” *Journal of digital imaging*, vol. 31, no. 5, pp. 655–669, 2018.

[10] Jelle Veraart, Dmitry S Novikov, Daan Christiaens, Benjamin Ades-Aron, Jan Sijbers, and Els Fieremans, “Denoising of diffusion mri using random matrix theory,” *NeuroImage*, vol. 142, pp. 394–406, 2016.

[11] Eunhee Kang, Hyun Jung Koo, Dong Hyun Yang, Joon Bum Seo, and Jong Chul Ye, “Cycle consistent adversarial denoising network for multiphase coronary ct angiography,” *arXiv preprint arXiv:1806.09748*, 2018.

[12] Maxim Zaitsev, Julian Maclaren, and Michael Herbst, “Motion artifacts in mri: a complex problem with many partial solutions,” *Journal of Magnetic Resonance Imaging*, vol. 42, no. 4, pp. 887–901, 2015.

[13] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian, “Video denoising by sparse 3d transform-domain collaborative filtering,” in *ESPC*. IEEE, 2007, pp. 145–149.

[14] F Edward Boas and Dominik Fleischmann, “Ct artifacts: causes and reduction techniques,” *Imaging Med*, vol. 4, no. 2, pp. 229–240, 2012.

[15] Katarzyna Krupa and Monika Bekiesińska-Figatowska, “Artifacts in magnetic resonance imaging,” *Polish journal of radiology*, vol. 80, pp. 93, 2015.

[16] AM Fahim, AM Salem, F Af Torkey, and MA Ramadan, “An efficient enhanced k-means clustering algorithm,” *Journal of Zhejiang University-Science A*, vol. 7, no. 10, pp. 1626–1633, 2006.

[17] Frank L Goerner and Geoffrey D Clarke, “Measuring signal-to-noise ratio in partially parallel imaging mri,” *Medical physics*, vol. 38, no. 9, pp. 5049–5057, 2011.

[18] Peter Kellman and Elliot R McVeigh, “Image reconstruction in snr units: a general method for snr measurement,” *MRM*, vol. 54, no. 6, pp. 1439–1447, 2005.

[19] N Rajalakshmi, K Narayanan, and P Amudhavalli, “Wavelet based weighted median filter for image denoising of mri brain images,” *IJEECS*, vol. 10, no. 1, pp. 201–206, 2018.

[20] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum, “Generative adversarial networks for noise reduction in low-dose ct,” *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[21] Xavier Glorot and Yoshua Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *ICAIS*, 2010, pp. 249–256.

[22] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[23] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, “Deep image prior,” in *CVPR*, 2018, pp. 9446–9454.