Penalty-driven enhanced self-supervised learning (Noise2Void) for CBCT denoising

Sungho Yun, Uijin Jeong, Taejin Kwon, Dain Choi, Taewon Lee, et al.

Sungho Yun, Uijin Jeong, Taejin Kwon, Dain Choi, Taewon Lee, Sung-Joon Ye, Gyuseong Cho, Seungryong Cho, "Penalty-driven enhanced self-supervised learning (Noise2Void) for CBCT denoising," Proc. SPIE 12463, Medical Imaging 2023: Physics of Medical Imaging, 1246327 (7 April 2023); doi: 10.1117/12.2652826

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

Penalty driven enhanced self-supervised learning (Noise2Void) for CBCT denoising

Sungho Yun*[a], Uijin Jeong[a], Taejin Kwon[a], Dain Choi[a], Taewon Lee[a], Sung-Joon Ye[b], Gyuseong Cho[a] , and Seungryon Cho[a]

[a]Department of Nuclear and Quantum Engineering, Korea Advanced Institution of Science and Technology, Daejeon, Republic of Korea; [b]Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea

ABSTRACT

Self-supervised learning for CT image denoising is a promising technique because it does not require clean target data that are usually unavailable in the clinic. Noise2void (N2V) is one of the famous methods to denoise the image without paired target data and it has been used to denoise optical images and also medical images such as MRI, and CT. However, the performance of the N2V is still limited due to the restricted receptive field of the network and it decreases the prediction performance for CT images that have complex image context and non-uniform Poisson random noise. Thus, we proposed enhanced N2V that utilizes penalty-driven network optimization to further denoise the images while preserving the important details. We used the total variation term to further denoise the image and also the laplacian pyramids term to preserve the important edges of the image. The degree of the influence of each penalty term is controlled by the hyperparameter value and they are optimized to achieve the best image quality in terms of noise level and structure sharpness. For the experiment, the real dental CBCT projection data were used to train the network in the projection domain. After the network training, the test results were reconstructed and compared at each different dose level. Meanwhile, PSNR, SNR, and a line profile were also evaluated to quantitatively compare the original FDK images, and proposed method. In conclusion, the proposed method achieved further denoises the image than N2V even preserving the details. By penalty-driven optimization, the network was able to learn the spectral features of the image while still the receptive field is limited to avoid identity mapping. We hope that our method would increase the practical utility of network-based CT images denoising that usually the target data are unavailable.

Keywords: self-supervised learning, noise2void, penalty term, CT, denoising and laplacian pyramid

1. INTRODUCTION

High noise in low-dose computed tomography (CT) is the main challenge because it often disrupts accurate diagnosis for medical treatment and surgical operation. The low number of incident photons causes the high Poisson random noise in measurement data due to the stochastic interaction nature between the imaging object and photons. Also, additional noise generated in the energy integrating detector contributed to resulting in poor image contrast (ex) an electrical noise). The low dose CT imaging is still a significant subject in the clinic because it would potentially decrease the biological damage due to the impact of ionizing radiation on patient health. To achieve the desired image quality without increment of scan dose, there have been lots of studies to denoise CT images. The filter-based denoising method [1-4] is a well-known approach with the penalty utilized iterative reconstruction method in CT image denoising [5-7]. Because its low computational cost well meets the requirement of CT image reconstruction in terms of speed, it has been preferred for practical application. Nowadays, lots of state-of-the-art deep-learning-based image denoising methods are being introduced in the field of computer vision and image processing. Due to the superior performance of the network in terms of denoising and edge-preserving to conventional filtering methods with efficient computational cost, they are now widely studied not only for optical images but also CT images [11-15]. However, the practical utility of the network in the medical image area is still limited due to the lack of available clinical data. Also, gathering noise-free data for the supervised learning-based method is nearly impractical. Self-supervised learning is one of the promising network training methods because there is no requirement for target data which is usually unavailable in the clinic. In recent, lots of self-supervised image-denoising methods have been introduced based on Noise2Noise, and Noise2Void [16-22]. In our study, we improved the performance of Noise2Void in the practical situation by employing additional penalty terms during network optimization. The total-variation (TV) term was used to expand the controllable degree of the smoothing

Medical Imaging 2023: Physics of Medical Imaging, edited by Lifeng Yu,
Rebecca Fahrig, John M. Sabol, Proc. of SPIE Vol. 12463, 1246327

range to further denoise the image [25]. The TV is designed to minimize a norm defined on the space of measures of bounded variation thereby it is useful to selectively kill the noise in the image. Meanwhile, to improve and conserve image sharpness, a laplacian pyramid was used [26-28]. The laplacian pyramid extracts the edges of the image using iterative subtraction and downsizing between the original image and the gaussian filtered image. Therefore, these two terms were complementarily conjugated to Noise2Void and optimized. To control the influence of penalty terms, the hyperparameters were multiplicated to them. Therefore, we were able to determine proper hyperparameters that generate the desired image quality which has high resolution and much denoised. The network training was fully conducted in the projection domain and it tested for both normal and low-dose dental CBCT datasets.

2. METHOD

2.1 Objective function with additional penalty terms

Before developing the algorithm, we considered whether penalty terms are able to be calculated by themselves to maintain the advantage of N2V. The total-variation term (TV) was first adopted to add because it may further denoise the images by minimizing the norm which is relevant to image sparsity. We intend to use this term with a hyperparameter to control the degree of the denoising level. Otherwise, image sharpness also is considered to preserve the details of images. The laplacian pyramids [26], is a useful edge extractor that uses iterative gaussian-filtered subtraction with image size modulation (Fig. 2). It is effective in selectively gathering the image connectivity feature that is relevant with edges among whole high-frequency components including the noise. In further, we additionally applied an anisotropic diffusion filter [29] in the first iteration step of the laplacian pyramid to prevent unwanted noise amplification. Therefore, the proposed objective function could be expressed

$$\begin{aligned}
&\arg\min_{\Theta} \sum_{j} \sum_{i} \left| f(x^{j}_{RF(i)};\Theta) - x^{j}_{i} \right|_{1} + \lambda_1 TV_{loss}(f(x^{j};\Theta)) + \lambda_2 LP_{loss}(f(x^{j};\Theta), x^{j}) \\
&s.t. \\
&TV_{loss}(x) = \frac{1}{N_u \cdot N_v} \sum_{u} \sum_{v} \left| x(u+1,v) - x(u,v) \right|_1 + \left| x(u,v+1) - x(u,v) \right|_1 \\
&LP_{loss}(x,y) = \sum_{iter} \left| \Phi_{iter}(x) - \Phi_{iter}(y) \right|_1
\end{aligned}$$

where TV_{loss} is the total-variation penalty term and LP_{loss} is the laplacian pyramid penalty term. *u, v* indicates image row and height. Φ_{iter} is Laplacian pyramid operator in certain iteration step *iter*. The fidelity and other penalty terms were constructed using l1-norm and conjugated with each hyperparameter value. The important thing is the absence of receptive fields in penalty terms. By doing so, the network was able to be penalized to generate desired spectral features by utilizing the excluded pixel information. In further, it is also helpful for accurate network prediction even with high irregularities by the complementation of vailed pixel information. Regarding identity mapping, the proposed penalty terms did not induce any unexpected identity mapping. This is because the TV loss term does not handle any image mapping and also the LP loss term only takes into account the sparse extracted edges of the image which is so far from the original one. Therefore, we were able to further denoise the images while preserving the details than the vanilla version of N2V. The hyperparameters were selectively chosen in the range of 0.001 to 0.1 to have the desired image quality in terms of noise level and sharpness. We also conducted an ablation study to confirm the effectiveness of each penalty term in the result section. To compare the performance of the proposed method with other methods (original FDK, vanilla N2V, and filter-based method), we measured SNR, PSNR, and line-profile for quantitative evaluation.

2.2 Network architecture & optimization technique

For the network architecture, we used a residual U-net [] which is consist of contracting and expanding paths with image size scaling. It has been widely used as a representative network architecture for semantic image segmentation or recovery in the medical imaging area. Each path is composed of multiple convolutional layers and non-linear activation functions with learnable weights that could be determined during the optimization process. In mathematical formulation, the network output follows the equation.

$$C_n(\mathbf{u}) = \mathrm{ReLU}\left(\mathbb{N}_B(\mathbf{W}_n * C_{n-1}(\mathbf{u}) + \mathbf{b}_n)\right)$$

where $\mathbf{u} \in \mathbb{R}$ is training input, $*$ means convolution, $\mathbf{W}_n$ and $\mathbf{b}_n$ denote weights and biases in the n^{th} (1,2,…,N) layer. $\mathbb{N}_B$ and ReLU denote batch normalization and non-linear activation function defined as $\mathrm{ReLU}(x) = \max(0, x)$. Over the optimization, the parameter $\{\mathbf{W}_1, \cdots, \mathbf{W}_N, \mathbf{b}_1, \cdots, \mathbf{b}_N\}$ sets were obtained to minimize the cost function. For the optimizer, Adam was used with a constant step learning rate schedular (step=10, gamma = 0.9) and the batch size was 4.

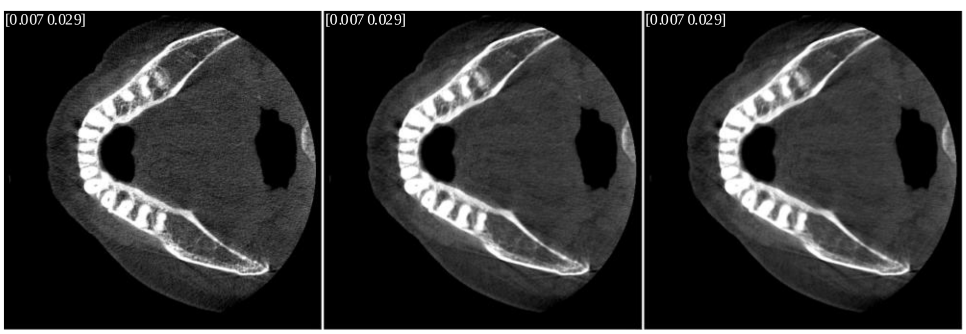
2.3 Dental CBCT data

To verify our algorithm in practical cases, we used the clinical dental CBCT projection data for network training and validation. Although the N2V also could be applied to reconstructed CT images, however, poor prediction performance and over-smoothing were often observed in this case. This is because amplified noise induced inaccurate network prediction, and it is much more critically represented in the image domain due to the presence of significant anatomical structures. Therefore, our study was conducted in the projection domain to avoid the above disadvantages. For the normal dose datasets (tube voltage: 95kVP and current: 11mA), we used 8 sets of patient data each consisting of 400 projections. Among them, only one set of patient data was used for the network training and others were used for the test. The projection data include normal teeth and soft tissue but also metals such as amalgam treatment or dental implants.

3. RESULTS

3.1 Effect of TV term

As mentioned earlier, the total variation was used to further denoise the projection data. To check the effectiveness of the total variation term, we prepared an ablation study by only adding the TV term into vanilla N2V. One of the patient data in the low-dose dataset (8mA) was tested for the study, and the results are shown in Fig 4. In Fig 4. (a), the reconstructed slice images from the vanilla N2V still have a high noise property, however, it was significantly reduced in (b) and (c) due to the increased influence of the TV term. But this term also blurred the structural details of the image as hyperparameters are increased. Therefore, we considered another penalty term to prevent the over smoothing thereby achieving an optimized noise and sharpness characteristic.



(a) (b) (c)

Figure 1. Reconstructed dental CBCT slice images with respective TV hyperparameter; (a) vanilla N2V (λ = 0.00), (b) N2V-TV (λ = 0.03), and (C) N2V-TV (λ = 0.05). The first and second row represents the slices of the axial and sagittal plane.

3.2 Effect of LP term

The Laplacian-pyramid term was utilized to increase the sharpness of structural details. To confirm the performance of the LP term, an ablation study was conducted by comparing the results between the N2V-TV and N2V-TV-LP (proposed method). Patient data from a low-dose dataset was tested in the comparison and results are shown in Fig 5. In Fig.5, the λTV was firstly determined to have the desired noise property and then the λLP was determined to enhance structure sharpness. As a result, it was able to obtain high image contrast by removing that much noise without blurring. About the scale of hyperparameter, [0.03,0.05] range of λTV was effective for this ablation study, and [0.03,0.07] range of λLP properly worked well to prevent blurring.

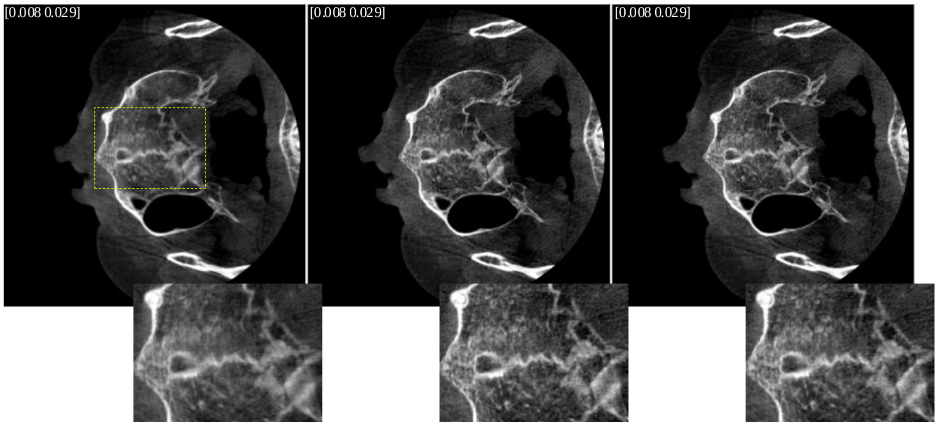


Figure 2. Reconstructed dental CBCT slice images with respective LP and the same TV hyperparameter; (a) N2V-TV (λTV = 0.05), (b) N2V-TV-LP (λTV = 0.05, λLP = 0.05), and (C) N2V-TV-LP (λTV = 0.05, λLP = 0.07). The first and second row represents the slices of the axial and sagittal plane.

3.3 Results of normal dose datasets

In this section, we tested our proposed method to normal dose patient datasets and compared them between original FDK, and vanilla N2V. Each hyperparameter value was carefully chosen (λTV = 0.03, λLP = 0.03) to generate a high-quality image in terms of noise and sharpness. In Fig. 3. the original FDK image is severely corrupted by high noise and the N2V shows a slightly mitigated one. The proposed method was able to further denoise the images than vanilla N2V while preserving microstructures of teeth. Similarly, the proposed method shows the best performance also in sagittal planes (in Fig. 7). The quantitative noise comparison is shown in Table 1. The SNR and PSNR were calculated between the results of the original FDK, vanilla N2V, and the proposed method. As a result, the proposed method shows the superior SNR and PSNR value among them (a higher value is better).

Table 1. The quantitative comparison between results of normal dose datasets. The SNR and PSNR are calculated to evaluate the noise statistics of original FDK, vanilla N2V, and the proposed method.

	Original FDK	Vanilla N2V	Proposed
SNR	17.46	24.30	31.82
PSNR	18.92	25.76	33.27

4. DISCUSSION

In this work, we proposed a way for enhancing the performance of N2V by using penalty-driven network optimization. N2V has great potential for medical image application because it makes it possible to denoise images without an additional clean target, however, the direct application was impractical due to its inaccurate prediction performance for high noise and complex image context. Through the proposed method, it was able to achieve the desired denoising performance without structure distortion or blurring. The total-variation effectively worked to remove severe noise without texture modification in the projection domain even in low-dose scanning conditions (in Fig. 1). The laplacian pyramid loss term was also effective to increase image sharpness and the complementary usage of both terms was able to generate high image contrast (in Fig. 2). The results of our proposed method show stable performance even in this case in Fig. 3. In quantitative comparison between the existing N2V method, the SNR increased 30.9% and the PSNR increased 29.1% in normal dose datasets. The heuristic hyperparameter selection could be a limitation of our method, however, there is no more requirement of parameter fitting in practical utilization when it is once pre-determined.

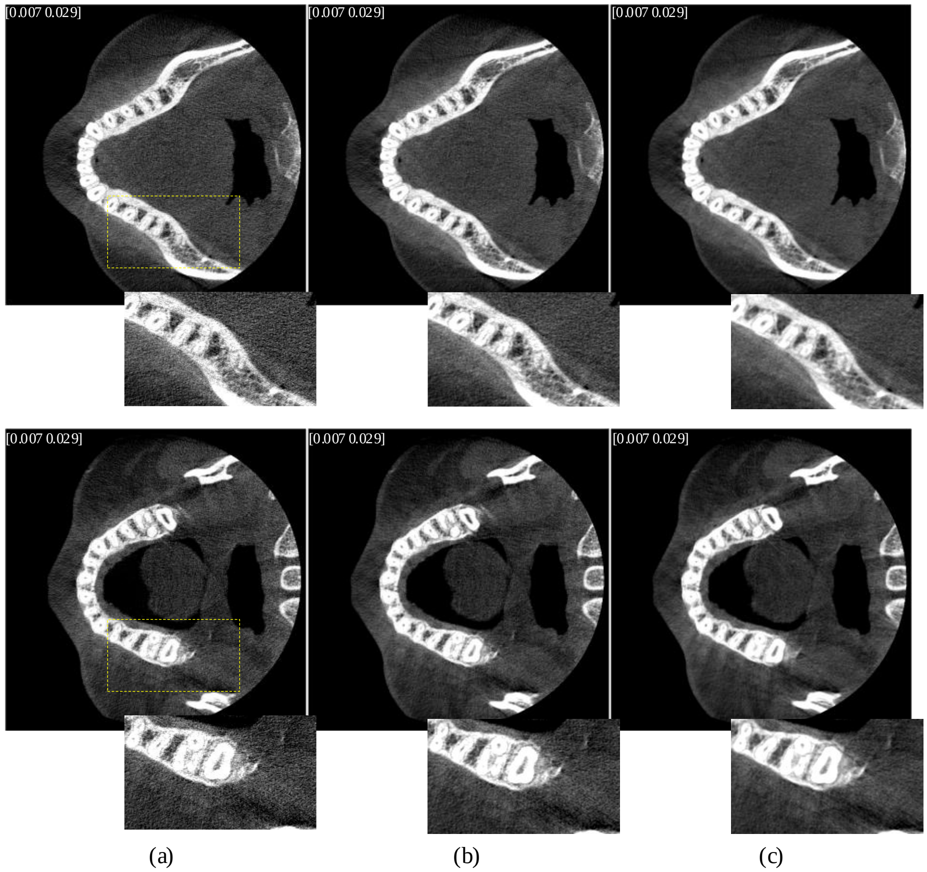


Figure 3. Reconstructed axial dental CBCT slice images of a patient; (a) Original, (b) vanilla N2V, and (C) N2V-TV-LP ($\lambda TV = 0.03$, $\lambda LP = 0.03$).

5. CONCLUSION

We have proposed an effective self-supervised learning-based denoising network with additional penalty terms for desired image quality. To increase the practical utility of N2V in the medical imaging area, we enhanced its prediction performance by TV terms and laplacian pyramid loss in the projection domain. The utilization of two terms was able to denoise further than the vanilla version while preserving the edges of the image. We found that our proposed method works well for real dental CBCT datasets even in low-dose cases. Finally, we hoped that our algorithm may increase the practical utility of the network for accurate diagnosis for patients.

REFERENCES

[1] Justusson B I 1981 Median filtering: statistical properties. Two-dimensional Digit. signal Process. II.
[2] Li Z, Yu L, Trzasko J D, Lake D S, Blezek D J, Fletcher J G, McCollough C H and Manduca A 2014 Adaptive nonlocal means filtering based on local noise level for CT denoising Med. Phys. 41
[3] Zhao T, Hoffman J, McNitt-Gray M and Ruan D 2019 Ultra-low-dose CT image denoising using modified BM3D scheme tailored to data statistics Med. Phys. 46
[4] Manduca A, Yu L, Trzasko J D, Khaylova N, Kofler J M, McCollough C M and Fletcher J G 2009 Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT Med. Phys. 36
[5] Tian Z, Jia X, Yuan K, Pan T and Jiang S B 2011 Low-dose CT reconstruction via edge-preserving total variation regularization Phys. Med. Biol. 56

[6] Sagara Y, Hara A K, Pavlicek W, Silva A C, Paden R G and Wu Q 2010 Abdominal CT: Comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients Am. J. Roentgenol. 195
[7] Kim Y, Kim Y K, Lee B E, Lee S J, Ryu Y J, Lee J H and Chang J H 2015 Ultra-low-dose CT of the thorax using iterative reconstruction: Evaluation of image quality and radiation dose reduction Am. J. Roentgenol. 204
[8] Dabov K, Foi R, Katkovnik V and Egiazarian K 2009 BM3D image denoising with shape-adaptive principal component analysis Proc. Work. Signal Process. with Adapt. Sparse Struct. Represent.
[9] ROBINSON E A and TREITEL S 1967 PRINCIPLES OF DIGITAL WIENER FILTERING Geophys. Prospect. 15
[10] Buades A, Coll B and Morel J-M 2011 Non-Local Means Denoising Image Process. Line 1
[11] Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J and Wang G 2017 Low-dose CT denoising with convolutional neural network Proceedings - International Symposium on Biomedical Imaging
[12] Yi X and Babyn P 2018 Sharpness-Aware Low-Dose CT Denoising Using Conditional Generative Adversarial Network J. Digit. Imaging 31
[13] Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra M K, Zhang Y, Sun L and Wang G 2018 Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss IEEE Trans. Med. Imaging 37
[14] Gu J and Ye J C 2021 AdaIN-Based Tunable CycleGAN for Efficient Unsupervised Low-Dose CT Denoising IEEE Trans. Comput. Imaging 7
[15] Li M, Hsu W, Xie X, Cong J and Gao W 2020 SACNN: Self-Attention Convolutional Neural Network for Low-Dose CT Denoising with Self-Supervised Perceptual Loss Network IEEE Trans. Med. Imaging 39
[16] Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M and Aila T 2018 Noise2Noise: Learning image restoration without clean data 35th International Conference on Machine Learning, ICML 2018 vol 7
[17] Krull A, Buchholz T O and Jug F 2019 Noise2void-Learning denoising from single noisy images Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition vol 2019-June
[18] Wu D, Kim K and Li Q 2021 Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning Med. Phys. 48
[19] Fang W, Wu D, Kim K, Kalra M K, Singh R, Li L and Li Q 2021 Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior Phys. Med. Biol. 66
[20] Song T A, Yang F and Dutta J 2021 Noise2Void: Unsupervised denoising of PET images Phys. Med. Biol. 66
[21] Niu C and Wang G 2020 Noise2Sim – Similarity-based self-learning for image denoising arXiv
[22] Papkov M, Roberts K, Madissoon L A, Shilts J, Bayraktar O, Fishman D, Palo K and Parts L 2021 Noise2Stack: Improving Image Restoration by Learning from Volumetric Data Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) vol 12964 LNCS
[23] Bai T, Wang B, Nguyen D and Jiang S 2021 Probabilistic self-learning framework for low-dose CT denoising Med. Phys. 48
[24] Liang K, Zhang L and Xing Y 2021 Training a low-dose CT denoising network with low-dose CT dataset: comparison of DDLN and Noise2Void
[25] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms Phys. D Nonlinear Phenom. 60
[26] Burt P J and Adelson E H 1983 The Laplacian Pyramid as a Compact Image Code IEEE Trans. Commun. 31
[27] Lai W S, Huang J Bin, Ahuja N and Yang M H 2017 Deep laplacian pyramid networks for fast and accurate super-resolution Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 vol 2017-January
[28] Paris S, Hasinoff S W and Kautz J 2011 Local Laplacian Filters: Edge-aware Image Processing with a Laplacian Pyramid ACM Trans. Graph. 30