

Usefulness of a No-Reference Metric for Evaluation of Images in Nuclear Medicine -A Comparative Study with Visual Assessment

Shigeaki Higashiyama (✉ higashiya@med.osaka-cu.ac.jp)

Osaka City University, Graduate School of Medicine, <https://orcid.org/0000-0002-3976-4521>

Yutaka Katayama

Osaka City University Hospital: Osaka Shiritsu Daigakuk Igakubu Fuzoku Byoin

Atsushi Yoshida

Graduate School of Medicine, Osaka City University

Nahoko Inoue

Graduate School of Medicine, Osaka City University

Takashi Yamanaga

Graduate School of Medicine, Osaka City University

Hideki Kawahata

Osaka City University Hospital: Osaka Shiritsu Daigakuk Igakubu Fuzoku Byoin

Takao Ichida

Osaka City University Hospital: Osaka Shiritsu Daigakuk Igakubu Fuzoku Byoin

Yukio Miki

Graduate School of Medicine, Osaka City University

Joji Kawabe

Graduate School of Medicine, Osaka City University

Original research

Keywords: No-reference metric, PIQE, Visual image evaluation, NIQE, Pairwise comparison method

Posted Date: April 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-372416/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose The normalized mean square error (NMSE) method is widely used to evaluate images in nuclear medicine. However, its use in clinical practice requires creating target images over a long time, which is difficult. It also requires standard images and reference teacher data that can be used at facilities worldwide because methods and doses in nuclear medicine imaging are yet to be standardized. This study examined the validity of perception-based image quality evaluator (PIQE), a no-reference metric that does not require a target image for image evaluation in nuclear medicine.

Methods The Hoffman brain phantom with 18F-fluoro-2-deoxy-D-glucose (FDG) was imaged and evaluated on slices in which the frontal and temporal lobes, bilateral lateral ventricles, and basal ganglia were drawn. Sixteen images with different pixel numbers and acquisition times were created and evaluated from the same slice. For the images obtained, visual evaluation criteria were developed for two endpoints: separation of white and gray matter boundaries, and uniformity of accumulation. Five evaluators assessed all the images using a paired comparison method, and then scored and ranked them. The images were also physically evaluated by the PIQE and a natural image quality evaluator (NIQE). The rankings obtained from the visual and physical evaluations were statistically compared.

Results Based on Spearman's test of significance, visual evaluation rankings showed a strong correlation with those obtained with PIQE ($r_s = 0.9559$, $p < 0.0001$) and no correlation with those obtained with NIQE ($r_s = 0.2324$ and $p = 0.3865$).

Conclusions Evaluation of images by PIQE is as effective as visual evaluation.

Introduction

Normalized mean square error (NMSE) is a widely used objective numerical index for setting the optimum collection and reconstruction conditions for nuclear medicine imaging [1-2]. It objectively evaluates the degree of similarity between the target and reference images. However, its use in clinical imaging is limited given the requirement for target images taken over time [1-5]. Hence, a new image evaluation method with the same reliability as the NMSE method is required.

In nuclear medicine, examinations are performed with many nuclides at various doses. Therefore, it is difficult to create a standard image or teacher data that can be used in all facilities. In recent years, diagnostic imaging support has adapted to various methods using artificial intelligence, among which the generative adversarial network (GAN), an unsupervised learning method, has been drawing attention [6-8]. In a study using GAN, Xuan et al. pointed out that it is difficult to evaluate images using peak signal to noise ratio (PSNR) based on the differences from the reference images [9]. They recommend using a no-reference metric, which utilizes the perceptual index as a metric for image evaluation [9].

To the best of our knowledge, there are no reports that statistically compare the results of visual evaluation of images by diagnostic imaging specialists and radiographers to the results of physical

evaluation using a no-reference metric. We created several images with varying acquisition times using the Hoffman brain phantom. These images were visually evaluated, scored, and ranked by five evaluators, including nuclear medicine physicians, radiologists, and nuclear medicine technologists. These rankings were compared to those from a physical evaluation using a no-reference metric, to determine the latter method's usefulness for image evaluation in nuclear medicine.

Materials And Methods

Image acquisition and analysis

A Hoffman 3D brain phantom (Data Spectrum) with 26 MBq of 18F-Fluoro-2-deoxy-D-glucose (FDG) was used as the head phantom. It was imaged as the target image for 1800 s in the list mode

The images were then reconstructed with frame times of 60, 120, 180, 300, 360, 450, 600, and 900 s. The reconstruction matrix was set to 440 and 880 pixels. Based on the clinical examination conditions of our hospital, the imaging conditions were:

Pixel size: 0.825×0.825

Slice thickness: 3

Reconstruction method: point-spread function (PSF) and time-of-flight (TOF)

Subset: 5, Iteration: 8

Filter: All-Pass

The images evaluated were slices from previous brain PET imaging studies using phantoms, depicting the frontal and temporal lobes, bilateral lateral ventricles, and the basal ganglia [10]. Biograph Vision (SIEMENS) was used for imaging and data acquisition. MATLAB (MathWorks) was used for image evaluation and calculation using a no-reference metric, and JMP (SAS Japan) was used for statistical analysis.

Visual evaluation and target images

Three qualified diagnostic radiologists and nuclear medicine specialists and four nuclear medicine technologists were selected to evaluate 16 images visually. Two of the four are nuclear medicine technologists with over 15 years of clinical experience. Two of the four are inexperienced nuclear medicine technologists. The images were acquired in the list mode with 440 pixels in the acquisition matrix at a frame time of 1800 s. Figure 1 shows the images reconstructed at frame times of 120, 180, 300, 360, 450, 600, and 900 s. Figure 2 shows the same images with 880 pixels.

Visual evaluation

Based on a study that evaluated the quality of brain PET images using 18F-FDG, the following two items were defined as the criteria for visual image evaluation[11]. (1) A clear delineation of the basal ganglia's limbic and its clear separation from the cerebral white and gray matter and (2) uniform accumulation of FDG in the basal ganglia and cerebral white matter.

We used the paired comparison method, which allows the evaluators' order of evaluation to be ranked [12].

Pairwise comparison method

A collaborator other than the visual evaluator established the procedure and acquired the images. Two different images from a total of 16 images were selected and displayed on either side of the monitor screen. A total of 240 images were prepared to be randomly displayed on the monitor. The identity of the two images to be presented was not disclosed to the evaluator.

Images in Figures 3 and 4

Figure 3 shows one of the images displayed on the monitor. On the left is a 440-pixel image at 180 s and on the right is a 900-pixel image at 880 s. The numbers corresponding to all 240 images evaluated are listed in Figure 4. Figure 3 corresponds to square 29 in Fig. 4.

Ten evaluation score sheets (Figure 4) were prepared for 5 evaluators to assess items 1 and 2. The table in Figure 4 was not shown to the evaluators, who visually evaluated the 240 images displayed in random order. If the display on the right side was better than that on the left side, one point was assigned to the cell in Figure 4 for that image.

For the display in Figure 3, if the image on the right showed a more uniform accumulation of FDG in the basal ganglia and white matter, the cell in square 29 of the evaluation score sheet for item 2 was assigned a score of 1. Two images identical to Figure 3 are included in the presentation but arranged in opposite directions, i.e., the 900 s image with 880 pixels is presented on the left side, and the 180 s image with 440 pixels is presented on the right side corresponding to square 212 in Figure 4. In this case, if the image on the left side was better, a score of 0 was assigned to square 212.

Previous reports have visually scored PET images with different acquisition times and the degree of glucose metabolism and malignancy in thyroid tumors on a 5-point scale [12-14]. Based on these reports, we scored the images for their acquisition times and pixel numbers.

Ranking images following visual evaluation

The higher the total score in the rightmost column of Figure 4, the better the result. Also, the lower the total score in the bottom row, the better the result. These scores were totaled, and the average values were calculated to obtain the visual evaluation scores and ranks.

Evaluation by NMSE

For the physical evaluation, we used a physical index based on the NMSE method. It has been conventionally used to calculate the similarity between target and reference images. The ideal and captured images are used as the reference and target images, respectively [3-4].

The NMSE method normalizes the target images using the maximum number of pixels. The smaller the calculated value, the closer it is to the ideal target image [7,15]. The calculation method is based on the formula shown in Figure 5 [2-3,15].

In this study, the target images were 440- and 880-pixels taken with a frame time of 1800 s. Therefore, the remaining images were used to calculate the physical index.

Evaluation using a no-reference metric

A perception-based quality evaluator (PIQE) is a non-reference perception-based image quality evaluation method for real-world images. It uses the mean subtraction contrast normalization coefficient to calculate the image quality score [16]. PIQE is an unsupervised method that does not require a learning model [8,16-18]. By contrast, natural image quality evaluator (NIQE) is an existing blind image quality evaluation method that relies on opinion-based supervised learning to predict quality scores [6,8].

PIQE is inspired by the following principles of how humans perceive image quality: First, human visual attention is strongly directed to prominent points in an image or spatially active areas; this property is adapted by estimating distortion only in the spatially prominent areas [16]. Second, local quality at the block/patch level is the overall quality of the image that humans perceive, and this property is addressed by calculating the distortion level at the local block level of size $n \times n$, where $n = 16$ [16].

Figure 6 shows a block diagram of the proposed method. The input image is preprocessed, followed by a block-level analysis to identify the distortion [16]. Each distorted block is assigned a score based on the distortion type, and the block-level scores are then pooled to determine the overall image quality. In addition to the quality score, it also generates a spatial quality map that can be effectively used in other applications.

NIQE, by contrast, uses only measurable deviations from the statistical regularities observed in natural images to calculate image quality scores in a completely blind manner [18]. It builds a collection of "quality-aware" statistical features based on a simple and successful spatial domain natural scene statistics (NSS) model [8]. Distorted image quality is expressed as a simple distance metric between the model statistic and distorted image statistic [8,18]. Lower PIQE and NIQE scores indicate better imaging evaluation [16-18].

Evaluation and statistical analysis

Figure 5 presents the NMSE scores and rankings for the other images. The visual evaluation and NMSE rankings were compared for evaluation.

The correlations between the PIQE/NIQE scores and the visual assessment rankings were evaluated. The 18F-FDG PET/CT results of the artificial intelligence-based method were compared to those of the expert reader (19), and Spearman's rank difference test was performed.3.

Results

Results of visual evaluation

Tables 1 and 2 show the scores of the two items for each image from all the raters using the paired comparison method. About the Table1 score, higher score men the better the evaluation result. About the Table2 score, lower score men the better the evaluation result. The average visual evaluation results using the pairwise comparison method for all the items is shown in Table 3. Evaluators 6 and 7 are inexperienced nuclear medicine technologists. There result of visual evaluations were tended to differ from the distribution from other evaluators. Therefore, regarding the examination of the ranking this time, we decided to exclude their evaluation results,.

The longer the shooting time, the better the score. For images with frame times other than 120 s, the score was better when the number of pixels was 880.

Results using the NMSE

Tables 4 and 5 present the NMSE evaluation results and ranked scores, respectively. In most images, the NMSE value improved and approached 0 as the acquisition time increased. In addition, the physical evaluation results of images with 880 pixels were better than those of images with 440 pixels.

Evaluation by PIQE

The results of the physical evaluation using the PIQE are shown in Table 5. The results were better for images with (a) a lower no-reference metric value, (b) longer acquisition time, and (c) 880 pixels. Spearman's significance test of the visual evaluation results and PIQE rankings showed a rank correlation coefficient (rs) of 0.9559 and p value < 0.0001, indicating a strong correlation between the two methods (Figure 7).

Results from NIQE

Table 6 presents the results of the physical evaluation by NIQE.

Spearman's significance test of the visual evaluation and NIQE rankings yielded an rs of 0.2324 and a p value of 0.3865, indicating no correlation between the two methods (Figure 8).

Discussion

In this study, we investigated the usefulness of a no-reference metric for evaluating images in nuclear medicine. The visual evaluation of the images by five raters was compared to that using the NMSE, and a statistical correlation was determined. Image evaluation using the PIQE showed a good correlation with visual evaluation, suggesting that these two methods are equally good.

Of the seven evaluators, the inexperienced evaluators 6 and 7 were inconsistent in the ranking of the results compared to the other evaluators. Therefore, regarding the examination of the ranking this time, we decided to exclude their evaluation results. This shows the validity of the evaluator's choice in this study. It also shows that the evaluation criteria are not easy evaluation criteria or evaluation targets that can be evaluated by anyone.

NMSE has been widely used for image evaluation and determination of imaging conditions in nuclear medicine [1-2]. We used NMSE to evaluate images of 1800 s as the target images. Images with longer acquisition times up to 900 s ranked better, corresponding to an image quality between 440 and 880 pixels. Although the number of images evaluated visually and using NMSE was different, the resultant rankings corresponded almost exactly.

Visual evaluation of images from 1800 to 180 s showed that longer acquisition times resulted in better evaluation scores (Table 3). The results and rankings obtained using NMSE were similar (Table 5). For images with the same acquisition time, the 880-pixels image scored better than the 440-pixels image (Table 3).

For images with a capturing time of 120 s, the difference in ranking between 440 and 880 pixels was less than 0.2 points. A minute difference compared with the other rankings; however, it reversed the visual evaluation rankings (Table 3).

Evaluators 4 and 5 evaluated the ranking of 440 and 880-pixel images with a 120 s acquisition time, reversing the rating order for items 1 and 2. They were diagnostic radiologists with more than 10 years of clinical experience and nuclear medicine specialists. This evaluation reversed the average rankings for the 440- and 880-pixel images at 120 s. For item 1, both evaluators found that the boundary between the white matter and gray matter of the temporal lobe and the peripapillary thalamus was clearer in the 440-pixels image because it had a wider area without accumulation. For item 2, the 440-pixels image more uniform accumulation because of a denser accumulation in the frontotemporal white matter, thalamus, and caudate nucleus in general. Noisy images acquired at 120 s were not of optimal quality to be used in clinical imaging.

The rankings obtained by visual evaluation and a no-reference metric method were compared by five evaluators. Generally, a supervised method outperforms an unsupervised method [17]. However, when creating the dataset for supervised learning in nuclear medicine, which is not well standardized, generating a standard image is not easy. It is a reality to perform a general-purpose quantitative evaluation using supervised learning, not the target [21-22]. In this study, PIQE, an unsupervised method that does not require learning data to evaluate image quality (20), yielded better results. Moreover, since

PIQE does not depend on learning data, it is considered a less environmentally dependent metric that can be handled on the same scale at all facilities conducting nuclear medicine examinations and imaging. Hence, PIQE may be an efficient image evaluation method.

The NIQE results showed no correlation with the visual evaluation results. This could be because NIQE, being a supervised method, employs a learning model using natural scene statistics [8,10,17]. Like natural images, PET images follow the Poisson distribution for image generation [23]. Since non-reference quality metrics match subjective human quality scores over fully-referenced quality metrics [6-7], PET image evaluation by a no-reference metric was expected to be useful. This is another reason why PIQE is more consistent with visual evaluation than NIQE.

There are several limitations to this study. Based on the brain phantom images, this study did not use clinical imaging. However, our findings suggest that PIQE, which does not depend on learning data and is less environment-dependent, maybe comparable to visual evaluation by radiologists and specialists. It is expected that PIQE will be applied for clinical image evaluation of the brain and other parts of the body.

Conclusions

The evaluation of nuclear medicine images using PIQE has the same capability as visual evaluation by qualified diagnostic radiologists and nuclear medicine technologists. PIQE, therefore, has the potential as a useful method for clinical image evaluation.

Declarations

Funding

The authors did not receive support from any organization for the submitted work.

Conflicts of interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Availability of data and material

The datasets analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Shigeaki Higashiyama and Yutaka Katayama. The first draft of the manuscript was written by Shigeaki Higashiyama and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval.

Ethical approval was waived by the local Ethics Committee of Osaka City University Hospital in view of the retrospective nature of the study and all the procedures being performed were part of the routine care.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Patients signed informed consent regarding publishing their data and photographs.

Acknowledgements

We are grateful to the radiological technicians at the Department of Radiology, Osaka City University Hospital.

References

1. Pierre-Olivier K, Denis Mariano-Goulart, Marion Quiquere, Françoise Lyonnet, Michel Zanca, Michel Rossi. Optimum tomographic reconstruction parameters for HMPAO brain SPET imaging: a practical approach based on subjective and objective indexes. *Eur J Nucl Med.*1995;22:671-677. DOI: 10.1007/BF01254569.
2. Kyohei Okuda, Susumu Fujii, Shota Sakimoto. Impact of Novel Incorporation of CT-based Segment Mapping into a Conjugated Gradient Algorithm on Bone SPECT Imaging: Fundamental Characteristics of a Context-specific Reconstruction Method. *Asia Ocean J Nucl Med Biol.*2019;7:49-57. DOI:10.22038/AOJNMB.2018.31711.1219.
3. Peter Händel. Understanding Normalized Mean Squared Error in Power Amplifier Linearization. *IEEE Microwave and Wireless Components Letters.*2018;28: 1047–1049. DOI: 10.1109/LMWC.2018.2869299.
4. Mario Carmelo Cirillo. On the use of the normalized mean square error in evaluating dispersion model performance. *Atmospheric Environment Part A General Topics.*1993;27:2427-2434. DOI: 10.1016/0960-1686(93)90410-Z.
5. Bill C. Penney, Michael A. King, Ronald B. Schwinger, Stephen P. Baker, Peter Stritzke, Paul W. Doherty. Constrained least-squares restoration of nuclear medicine images: Selecting the coarseness function. *MEDICAL PHYSICS.*1987;14:849-858. DOI: <https://doi.org/10.1118/1.596123>
6. Anish Mittal, Anush Krishna Moorthy, Alan Conrad Bovik. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing.*2012;21: 4695 – 4708. DOI: 10.1109/TIP.2012.2214050.

7. H.R.Sheikh, A.C.Bovik, L.Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Transactions on Image Processing. 2005;14:1918-1927. DOI: 10.1109/TIP.2005.854492.
8. Anish Mittal, Rajiv Soundararajan, Alan C. Bovik. Making a “Completely Blind” Image Quality Analyzer. IEEE Signal Processing Letters. 2019;20:209-212.DOI: 10.1109/LSP.2012.2227726
9. Xuan Zhu, Yue Cheng, Jinye Peng, Rongzhi Wang, Mingnan Le, Xin Liu.Super-Resolved Image Perceptual Quality Improvement via Multi-Feature Discriminators.Computer Vision and Pattern Recognition.DOI: 10.1117/1.JEI.29.1.013017
10. Frederic Fahey, Paul Christian, Katherine Zukotynski. Briana Sexton-Stallone¹, Christina Kiss, Bonnie Clarke, Arzu Onar-Thomas, Tina Young Poussaint. Use of a Qualification Phantom for PET Brain Imaging in a Multicenter Consortium: A Collaboration Between the Pediatric Brain Tumor Consortium and the SNMMI Clinical Trials Network. J Nucl Med. 2019;60:677–682 DOI: 10.2967/jnumed.118.219998
11. Julien Salvadori¹, Laetitia Imbert, Mathieu Perrin, Gilles Karcher, Zohra Lamiral, Pierre-Yves Marie, Antoine Verger. Head-to-head comparison of image quality between brain 18F-FDG images recorded with a fully digital versus a last-generation analog PET camera. EJNMMI Research. 2019:61 DOI: <https://doi.org/10.1186/s13550-019-0526-5>
12. Thomas C. Brown, George L. Peterson. An enquiry into the method of paired comparison: reliability, scaling, and Thurstone's Law of Comparative Judgment. General Technical Report. DOI: <https://doi.org/10.2737/RMRS-GTR-216>
13. Philippe Thuillier, David Bourhis, Nathalie Roudaut. Geneviève Crouzeix, Zarrin Alavi, Ulrike Schick, Philippe Robin, Véronique Kerlan, Pierre-Yves Salaun, Ronan Abgral. Diagnostic Value of FDG PET-CT Quantitative Parameters and Deauville-Like 5 Point-Scale in Predicting Malignancy of Focal Thyroid Incidentaloma. Frontiers in Medicine .2019;12 DOI: <https://doi.org/10.3389/fmed.2019.00024>.
14. Garry M. McDermott, Fahmid U. Chowdhury, Andrew F. Scarsbrook. Evaluation of noise equivalent count parameters as indicators of adult whole-body FDG-PET image quality. Ann Nucl Med.2013;27:855–861.DOI:10.1007/s12149-013-0760-2.
15. Jessica Chani-Cahuana, Christian Fager, Thomas Eriksson.Lower Bound for the Normalized Mean Square. Error in Power Amplifier Linearization.IEEE Microwave and Wireless Components Letters.2018;28:435-427.DOI: 10.1109/LMWC.2018.2817021
16. Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Sumohana S. Channappayya, warup S. Medasani.Blind image quality evaluation using perception based features.2015 Twenty First National Conference on Communications.DOI: 10.1109/NCC.2015.7084843.
17. Pina Marziliano¹, Frederic Dufaux, Stefan Winkler, Touradj Ebrahimi.Perceptual blur and ringing metrics: application to JPEG2000.Image Communication.2004;19:163–172.DOI: 10.1016/j.image.2003.08.003
18. H.R. Sheikh, M.F.Sabir, A.C.Bovik.A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms.IEEE Transactions on Image Processing. 2006;15:3440-3451.DOI:

10.1109/TIP.2006.881959

19. Nicolo *Capobianco*, Michel *Meignan*, A \cap e - *Sego* \leq ne Cottureau. Laetitia Vercellino, Ludovic Sibille, Bruce Spottiswoode, Sven Zuehlsdorff, Olivier Casasnovas, Catherine Thieblemont, Irène Buvat. Deep-Learning 18F-FDG Uptake Classification Enables Total Metabolic Tumor Volume Estimation in Diffuse Large B-Cell Lymphoma. *J Nucl Med* .2021; 62:30–36 DOI: 10.2967/jnumed.120.242412.
20. Dan Yang, Veli-Tapani Peltoketo, Joni-Kristian Kämäräinen. 2019 IEEE/CVF International Conference on Computer Vision Workshop. DOI: 10.1109/ICCVW.2019.00485.
21. Andrea Varrone, Susanne Asenbaum, Thierry Vander Borghet, Jan Booij, Flavio Nobili, Kjell Någren, Jacques Darcourt, Özlem L. Kapucu, Klaus Tatsch, Peter Bartenstein, Koen Van Laere. EANM procedure guidelines for PET brain imaging using [18F]FDG, version 2. *Eur J Nucl Med Mol Imaging*. 2009;36:2103-2110. DOI: 10.1007/s00259-009-1264-0.
22. Ronald Boellaard, Roberto Delgado-Bolton, Wim J. G. Oyen, Francesco Giammarile, Klaus Tatsch, Wolfgang Eschner, Fred J. Verzijlbergen, Sally F. Barrington, Lucy C. Pike, Wolfgang A. Weber, Sigrid Stroobants, Dominique Delbeke, Kevin J. Donohoe, Scott Holbrook, Michael M. Graham, Giorgio Testanera, Otto S. Hoekstra, Josee Zijlstra, Eric Visser, Corneline J. Hoekstra, Jan Pruim, Antoon Willemsen, Bertjan Arends, Jörg Kotzerke, Andreas Bockisch, Thomas Beyer, Arturo Chiti, Bernd J. Krause. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0 *Eur J Nucl Med Mol Imaging*. 2015;42:328–354 DOI: 10.1007/s00259-014-2961-x.
23. A. Teymurazyan, T. Riauka, H.-S. Jans, D. Robinson. Properties of Noise in Positron Emission Tomography Images Reconstructed with Filtered-Backprojection and Row-Action Maximum Likelihood Algorithm. *J Digit Imaging*. 2013;26:447–456. DOI: 10.1007/s10278-012-9511-5

Tables

Due to technical limitations the Tables are available as a download in the Supplementary Files.

Figures

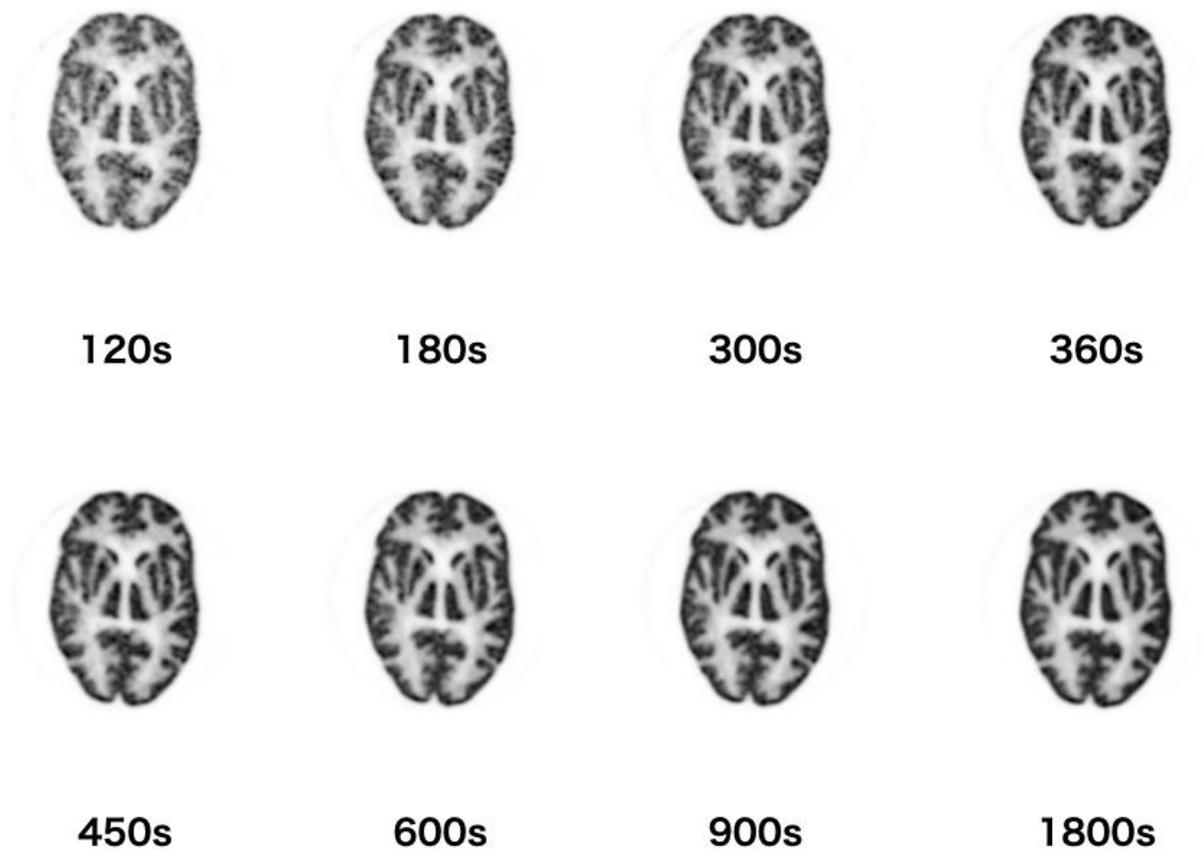


Figure 1

Hoffman 3D brain phantom images with 440-pixels A Hoffman 3D brain phantom with 26 MBq of 18F-fluoro-2-deoxy-D-glucose (FDG) was imaged and acquired with Biograph Vision (SIEMENS). The images were acquired at a frame time of 1800 s in the list mode with a 440-pixel acquisition matrix using slices depicting the frontal and temporal lobes, bilateral lateral ventricles, and basal ganglia. Images were reconstructed at frame times of 120, 180, 300, 360, 450, 600, and 900 s.

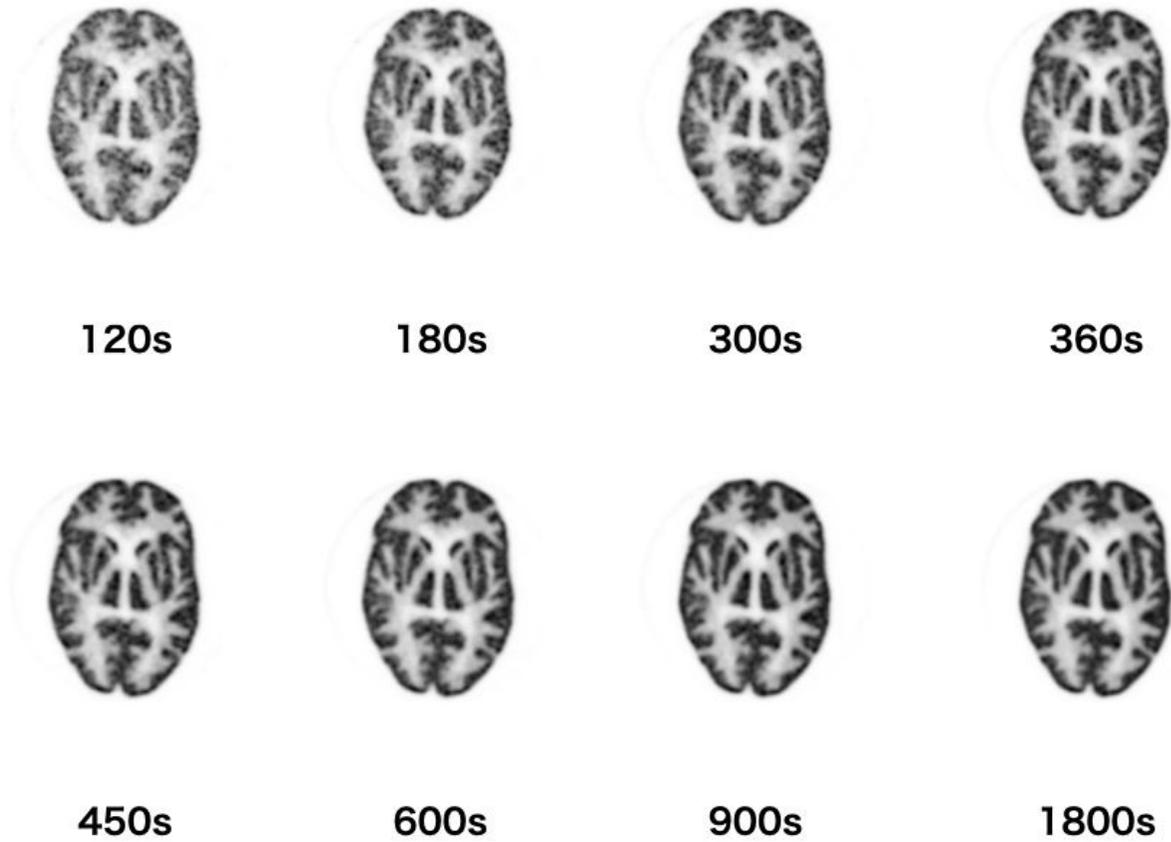


Figure 2

Hoffman 3D brain phantom images with 880-pixels A Hoffman 3D brain phantom with 26 MBq of ^{18}F -fluoro-2-deoxy-D-glucose (FDG) was imaged and acquired with Biograph Vision (SIEMENS). The images were acquired at a frame time of 1800 s in the list mode with an 880-pixel acquisition matrix using slices depicting the frontal and temporal lobes, bilateral lateral ventricles, and basal ganglia. Images were reconstructed at frame times of 120, 180, 300, 360, 450, 600, and 900 s.



Figure 3

An image presented to the evaluator Shown is one of the images presented to the evaluator. On the left is a 180-second image with 440 pixels, and on the right is a 900-second image with 880 pixels. This image corresponds to the square number 29 shown in Fig. 4.

	440-120s	440-180s	440-300s	440-360s	440-450s	440-600s	440-900s	440-1800s	880-120s	880-180s	880-300s	880-360s	880-450s	880-600s	880-900s	880-1800s	score
440-120s		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
440-180s	16		17	18	19	20	21	22	23	24	25	26	27	28	29	30	
440-300s	31	32		33	34	35	36	37	38	39	40	41	42	43	44	45	
440-360s	46	47	48		49	50	51	52	53	54	55	56	57	58	59	60	
440-450s	61	62	63	64		65	66	67	68	69	70	71	72	73	74	75	
440-600s	76	77	78	79	80		81	82	83	84	85	86	87	88	89	90	
440-900s	91	92	93	94	95	96		97	98	99	100	101	102	103	104	105	
440-1800s	106	107	108	109	110	111	112		113	114	115	116	117	118	119	120	
880-120s	121	122	123	124	125	126	127	128		129	130	131	132	133	134	135	
880-180s	136	137	138	139	140	141	142	143	144		145	146	147	148	149	150	
880-300s	151	152	153	154	155	156	157	158	159	160		161	162	163	164	165	
880-360s	166	167	168	169	170	171	172	173	174	175	176		177	178	179	180	
880-450s	181	182	183	184	185	186	187	188	189	190	191	192		193	194	195	
880-600s	196	197	198	199	200	201	202	203	204	205	206	207	208		209	210	
880-900s	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225		
880-1800s	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240		
score																	

Figure 4

Evaluation of all images shown is the evaluation sheet for the images (out of 240 images) presented to the evaluator. The rating is the total score in the leftmost column and the bottom row.

$$\text{NMSE} = \frac{\sum (g(x,y) - f(x,y))^2}{\sum (f(x,y))^2}$$

Figure 5

NMSE formula NMSE: Normalized mean square error shown is the formula used in the NMSE method. The mean square error of the evaluated image is for the reference image. Reference image: $f(x,y)$
Evaluation image: $g(x,y)$

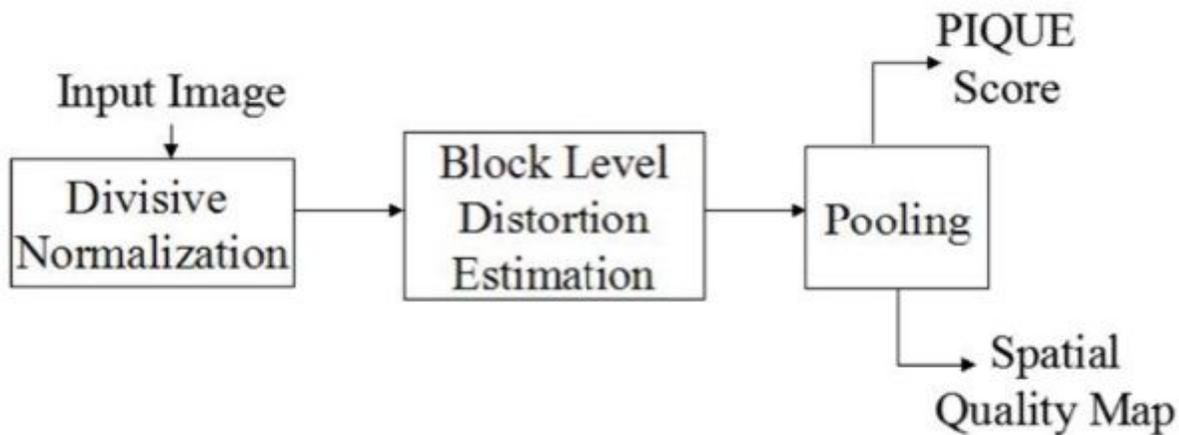


Figure 6

Block diagram of the proposed method The input image is subjected to a preprocessing step. A block-level analysis is performed to identify the distortion, and each distorted block is assigned a score based on the distortion type. The block-level scores are finally pooled to determine the overall image quality.

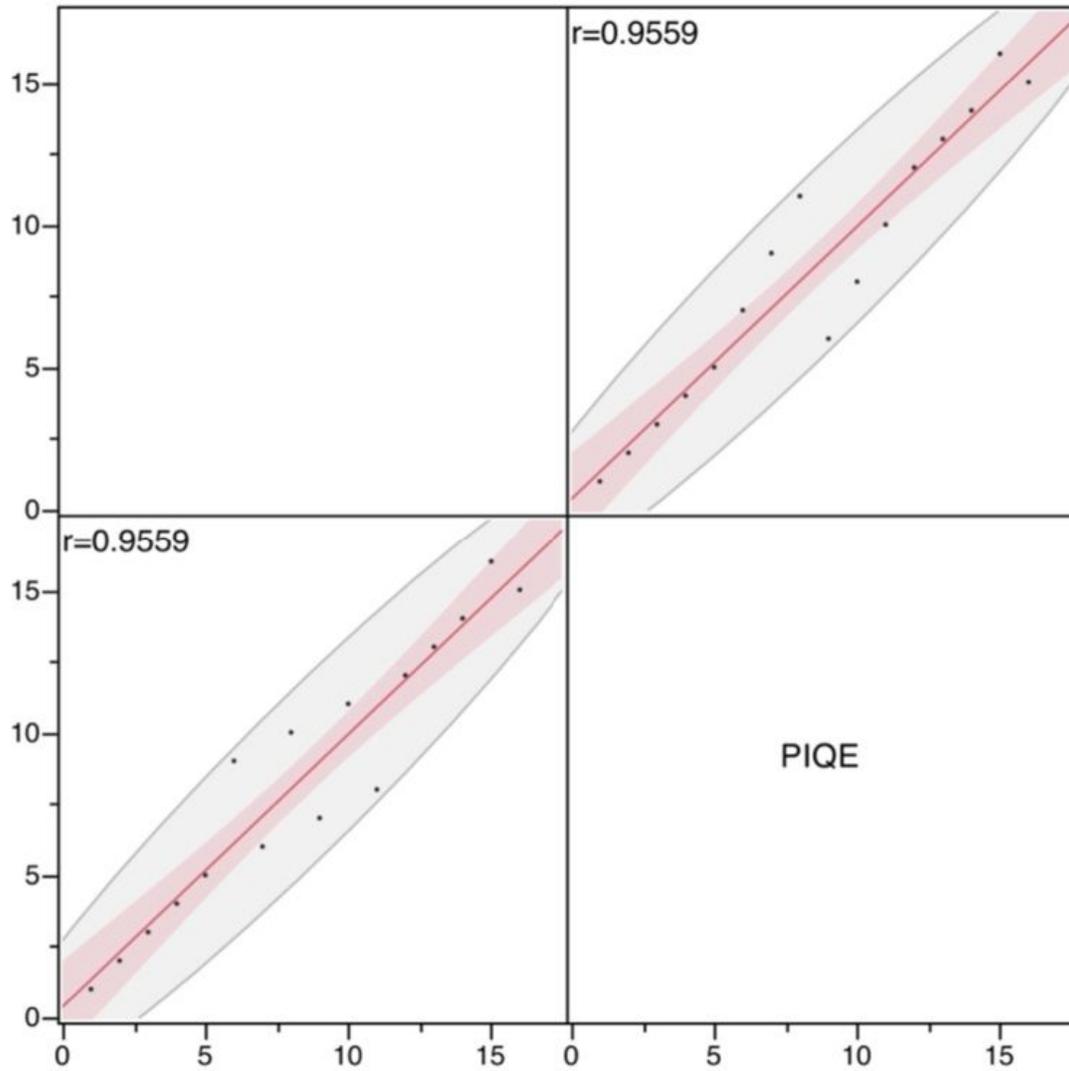


Figure 7

Correlation between the rankings of visual assessment and PIQE PIQE: Perception-based image quality evaluator Spearman's significant difference test between the rankings of visual assessment and PIQE reveals an r_s of 0.9559 and a p value < 0.0001 , indicating a strong correlation.

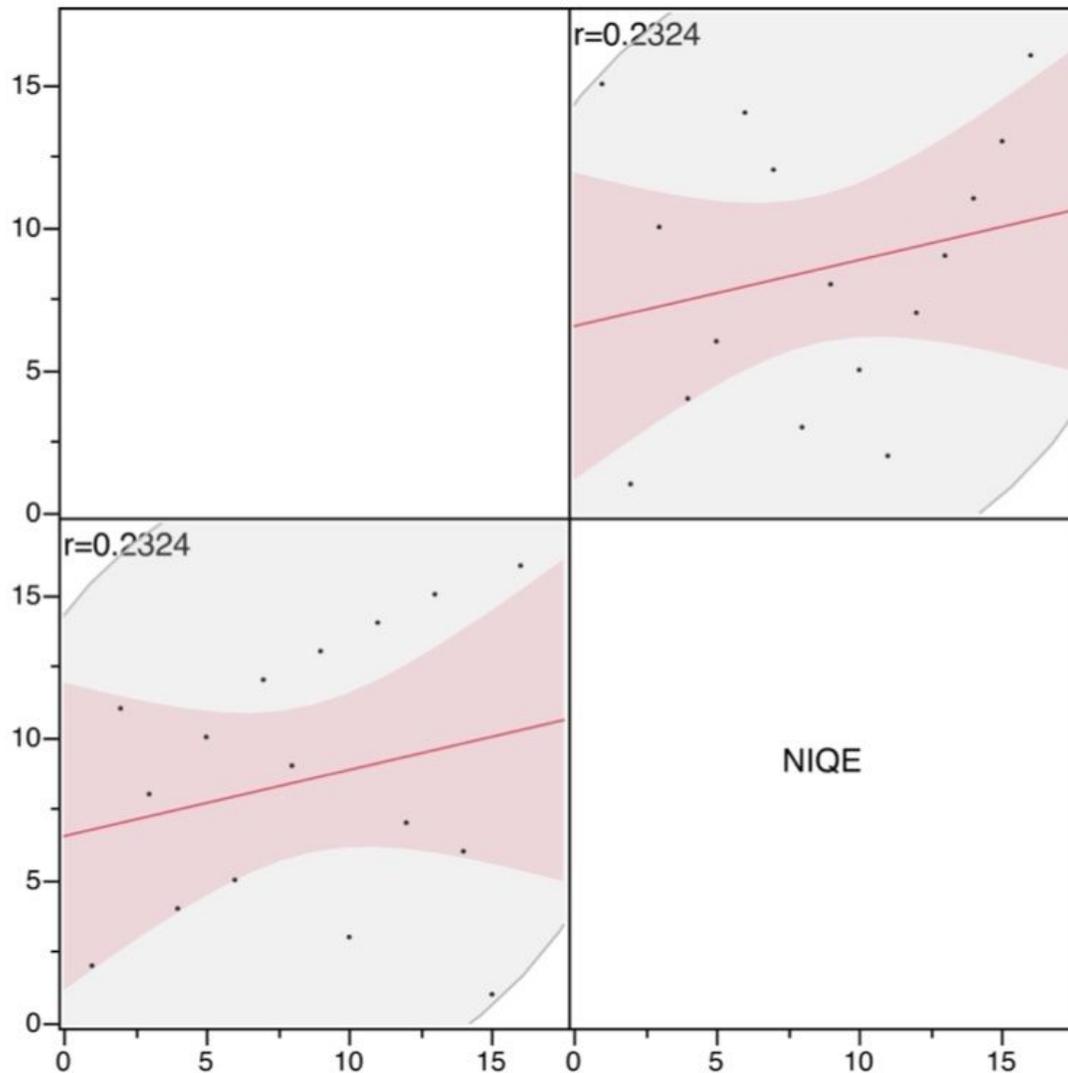


Figure 8

Correlation between the rankings of visual assessment and NIQE NIQE: Natural Image Quality Evaluator Spearman's significant difference test between the rankings of visual assessment and NIQE reveals a r_s of 0.2324 and a p value of 0.3865, indicating no strong correlation between the two methods.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.pdf](#)