MS-IQA: A Multi-Scale Feature Fusion Network for PET/CT Image Quality Assessment

Siqiao Li[1], Chen Hui[1,2⋆], Wei Zhang[1], Rui Liang[3], Chenyue Song[1], Feng Jiang[2], Haiqi Zhu[1⋆], Zhixuan Li[4], Hong Huang[5], and Xiang Li[1]

[1] Harbin Institute of Technology, China
[2] Nanjing University of Information Science and Technology, China
[3] Harbin Medical University, China
[4] Nanyang Technological University, Singapore
[5] Sichuan University of Science and Engineering, China
siqiaoli@stu.hit.edu.cn, chui@nuist.edu.cn, haiqizhu@hit.edu.cn

Abstract. Positron Emission Tomography / Computed Tomography (PET/CT) plays a critical role in medical imaging, combining functional and anatomical information to aid in accurate diagnosis. However, image quality degradation due to noise, compression and other factors could potentially lead to diagnostic uncertainty and increase the risk of misdiagnosis. When evaluating the quality of a PET/CT image, both low-level features like distortions and high-level features like organ anatomical structures affect the diagnostic value of the image. However, existing medical image quality assessment (IQA) methods are unable to account for both feature types simultaneously. In this work, we propose MS-IQA, a novel multi-scale feature fusion network for PET/CT IQA, which utilizes multi-scale features from various intermediate layers of ResNet and Swin Transformer, enhancing its ability of perceiving both local and global information. In addition, a multi-scale feature fusion module is also introduced to effectively combine high-level and low-level information through a dynamically weighted channel attention mechanism. Finally, to fill the blank of PET/CT IQA dataset, we construct PET-CT-IQA-DS, a dataset containing 2,700 varying-quality PET/CT images with quality scores assigned by radiologists. Experiments on our dataset and the publicly available LDCTIQAC2023 dataset demonstrate that our proposed model has achieved superior performance against existing state-of-the-art methods in various IQA metrics. This work provides an accurate and efficient IQA method for PET/CT. Our code and dataset are available at https://github.com/MS-IQA/MS-IQA/.

Keywords: PET/CT · Image Quality Assessment · Multi-Scale Feature Fusion · Attention Mechanism · Deep Learning

⋆ Chen Hui and Haiqi Zhu are both corresponding authors.
This work was supported in part by the Central Guidance for Local Science and Technology Development Fund Projects under grant 2024ZYD0266, in part by the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology under grant 2025r029.

1 Introduction

Positron Emission Tomography / Computed Tomography (PET/CT) is a widely used imaging modality that integrates metabolic information from PET and anatomical details from CT, providing essential insights for clinical diagnosis and treatment planning [1]. The quality of PET/CT images significantly impacts diagnostic accuracy, influencing clinical decision-making in oncology, cardiology, and neurology. However, PET/CT images often suffer from noise, compression artifacts and scan-related distortions due to limited radio tracer dose, short acquisition time and data compression during transmission [2]. Low-quality images can obscure critical anatomical structures, leading to misinterpretation and increased diagnostic uncertainty. The need for an automated and accurate PET/CT image quality assessment (IQA) method is therefore crucial to enhance clinical efficiency and reliability.

IQA methods are generally categorized into Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) approaches [3]. In clinical scenarios, where reference images are unavailable, NR-IQA is the most applicable method for PET/CT IQA [4]. Traditional NR-IQA techniques like NIQE [5], BRISQUE [6] and DIIVINE [7] rely on hand-crafted statistical features that are optimized for natural images but unable to account for the unique characteristics of medical imaging, such as anatomical structure visibility and diagnostic value. Deep learning-based NR-IQA models have demonstrated significant advancements in natural image quality assessment (NIQA) [8–12], but their direct application to medical imaging remains challenging [13–15]. First, PET/CT IQA is determined by both low-level distortions (e.g., noise, compression artifacts) and high-level anatomical structures (e.g., organ boundaries, lesion detectability). CNN-based [16] models excel at extracting local texture features but struggle with long-range contextual information, while Transformer-based [17, 18] models provide global feature modeling but lack fine-grained local feature extraction [19, 20]. Therefore, a robust PET/CT IQA model should integrate multi-scale features to capture both local artifacts and global anatomical structures simultaneously. Additionally, the scarcity of large-scale labeled PET/CT IQA datasets also significantly limits the performance of deep learning models.

To address these challenges, we propose MS-IQA, a multi-scale feature fusion network for PET/CT IQA, which integrates both local and global features for robust quality assessment. Our experiments on PET-CT-IQA-DS and the LDCTIQAC2023 [21] dataset demonstrate that MS-IQA outperforms existing NR-IQA models, achieving superior correlation with radiologist assessments. Our key contributions are summarized as follows:

- We propose MS-IQA, a novel multi-scale feature fusion network for PET/CT IQA, which leverages both local and global features from ResNet and Swin Transformer to achieve accurate and efficient PET/CT IQA.
- We propose a multi-scale feature fusion mechanism that combines high-level and low-level intermediate features through a dynamically weighted channel attention mechanism, effectively enhancing the perceptual capability.

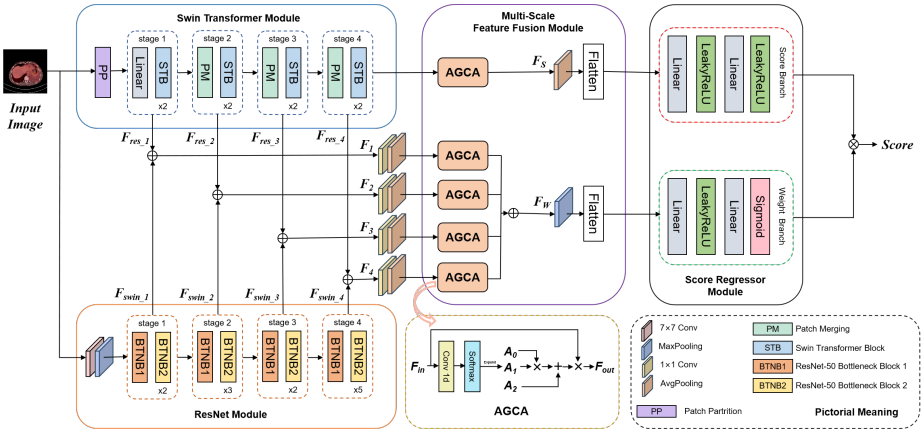


Fig. 1: The architecture of our proposed model: MS-IQA. It consists of 4 modules: RM, STM, MSFFM and SRM. The detailed structure of AGCA block is shown in the lower corner, with pictorial meaning on its right side.

- We construct PET-CT-IQA-DS, a dataset containing 2,700 varying-quality PET/CT images with quality scores assigned by radiologist. To the best of our knowledge, this is the first PET/CT IQA dataset.

2 Methodology

The proposed MS-IQA model leverages both local feature extraction capabilities of ResNet and global contextual modeling capabilities of Swin Transformer through a dual-branch structure. It consists of four main components: the ResNet Module (RM), the Swin Transformer Module (STM), the Multi-Scale Feature Fusion Module (MSFFM), and the Score Regressor Module (SRM) as illustrated in Fig. 1. First, the input image is passed through RM and STM simultaneously, obtaining F_{res_1} to F_{res_4} and F_{swin_1} to F_{swin_4} from 4 intermediate stages of each module. Then these features are sent into MSFFM for fusion through an Adaptive Graph Channel Attention (AGCA) mechanism [24], which refines the attention over the feature channels. Finally, the fused features are used to predict the quality score in SRM.

2.1 ResNet Module

We use ResNet-50 [22] as one of our feature extraction backbones. ResNet-50 demonstrates strong feature extraction capabilities, particularly in extracting local features. It consists of 50 layers, which can be divided into four stages, as shown in Fig. 1. The output features of different stages have varying scale. Lower-level features have larger scale and retain rich spatial information like edge textures. Higher-level features have smaller scale and represent semantic

information like organ contours. The input image Img is passed through ResNet-50 and extract features outputted by stage 1 to 4 as F_{res_1} to F_{res_4} as shown in Eq. 1. Then we send them into the multi-scale feature fusion module.

$$F_{res\_i} = ResNet_{stagei}(Img) \tag{1}$$

2.2 Swin Transformer Module

The other feature extraction backbone we use is Swin Transformer Tiny [23], a Vision Transformer (ViT) [25] based model. The Shifted Window Mechanism it used enhances global feature extracting while maintaining the computational efficiency of local self-attention, making it suitable for modeling long-range dependencies and preserving global information.It also adopts a hierarchical feature representation mechanism. As shown in Fig. 1, it consists of four stages (stage 1 to 4), where the feature scale gradually decreases as the depth increases. Similar to the ResNet module, We pass the input image Img through this module and extract features outputted by stage 1 to 4 as F_{swin_1} to F_{swin_4} as shown in Eq. 2. Then we also send them into the multi-scale feature fusion module.

$$F_{swin\_i} = Swin_{stagei}(Img) \tag{2}$$

2.3 Multi-Scale Feature Fusion Module

In this module, we aim to combine the strong local feature perception capabilities of ResNet with the long-range dependencies modeling capabilities of Swin Transformer through a feature fusion mechanism, effectively utilizing the information embedded in multi-scale features. First, we combine F_{res_i} from the i-th stage of ResNet with F_{swin_i} from the i-th stage of Swin Transformer. It can be found that F_{res_i} and F_{swin_i} have the same width and height, so they can be concatenated along the channel dimension to obtain F_i as shown in Eq. 3. Before entering the AGCA block, F_i first passes through a 1×1 convolution layer for dimensionality reduction to reduce computational complexity, and then an average pooling layer is applied to obtain the average value of each channel.

$$F_i = F_{res\_i} \oplus F_{swin\_i} \tag{3}$$

AGCA introduces graph theory into channel attention, treating each channel as a vertex of a graph, with the relationships between channels represented by the graph's adjacency matrix. The structure of the AGCA block is shown in Fig. 1. For f_{in}, the input of AGCA, a 1×1 convolution layer, followed by a sigmoid operation are applied, resulting in f as shown in Eq. 4.

$$f = \sigma(Conv(f_{in})) \tag{4}$$

The attention calculation involves three matrices: A_0, A_1, and A_2. First, f is transposed and replicated C times to expand it into a $C \times C$ shape, resulting in A_1, as shown in Eq. 5.

$$A_1 = \underbrace{[f^T \ f^T \ \cdots \ f^T]}_{C \text{ times}} \tag{5}$$

The calculation process of attention matrix A is shown as Eq. 6. A_0 is an identity matrix, and by performing matrix multiplication between A_0 and A_1, the resulting matrix has its diagonal elements set to the diagonal elements of A_1, with all other elements set to 0, which represents the self attention of each channel. A_2 is the adjacency matrix of the graph where the vertices correspond to channels, capturing the attention relationships between channels. A_2 is added to the multiplication result between A_0 and A_1 to obtain the final attention matrix A, which combines both the self attention and the inter-channel attention.

$$A = (A_0 \times A_1) + A_2 \tag{6}$$

Subsequently, A, which now represents the attention weights of each channel, is multiplied with f_{in} to apply the channel attention on it, including both self-attention and inter-channel attention.

F_1 to F_4 are separately sent into AGCA block for attention computation and then concated to obtain the final fused feature F_W, which is then passed through a max pooling layer for dimensionality reduction and flattened into a one-dimensional feature, denoted as f_w. Additionally, the output of the last stage of Swin Transformer, F_{swin_4}, is processed through an AGCA module, producing F_S, which is then passed through a same process as F_W to obtain f_s. Finally, f_w and f_s are sent into the score regression module.

2.4 Score Regresssor Module

Our score regression module contains two branches as shown in Fig. 1. f_S is passed through the score branch consisting of 2 linear layers and 2 LeakyReLU to obtain score S as shown in Eq. 7, while f_W is passed through the weight branch consisting of 2 linear layers, 1 LeakyReLU and 1 Sigmoid to obtain weight W as shown in Eq. 8. The final quality score is obtained by multiplying S and W, as shown in Eq. 9. This balancing mechanism between two branches serves as a regularization constraint, effectively mitigating model overfitting [31].

$$S = LeakyReLU(Linear_{S2}(LeakyReLU(Linear_{S1}(f_S)))) \tag{7}$$

$$W = \sigma(Linear_{W2}(LeakyReLU(Linear_{W1}(f_W)))) \tag{8}$$

$$Score = S \times W \tag{9}$$

2.5 Loss Function

Eq. 10 to Eq. 12 show the loss function we employed, where n represents the total number of samples, t_i and t_j represent the ground truth values of the i-th and j-th samples, y_i and y_j represent the predicted values of the i-th and j-th samples, α is a hyperparameter used to control the scaling factor and we set it to 2.0 in practice, σ represents Sigmoid activation function, λ_1 and λ_2 represent

the weights of two losses and we set λ_1=λ_2=0.5 in this work. The loss function simultaneously optimizes the model's ranking ability and regression ability.

$$L_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - t_i)^2 \tag{10}$$

$$L_{\text{Ranking}} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\sigma\left(\alpha(t_i - t_j)\right) - \sigma\left(\alpha(y_i - y_j)\right)\right)^2 \tag{11}$$

$$L_{\text{Total}} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{Ranking}} \tag{12}$$

3 Experiments

3.1 Dataset and Experiment Details

To the best of our knowledge, we are the first to fill the blank of PET/CT IQA dataset. We construct PET-CT-IQA-DS, which contains 2,700 varying-quality PET/CT images with quality score labels. The pristine images are sourced from 20 patients of the publicly available dataset Lung-PET-CT-Dx [26]. For each patient, 5 images were selected from different anatomical locations. Due to factors such as the dose of radioactive tracers, thermal noise and compression during transmission, mixed distortions may be introduced to PET/CT [27]. Therefore, we add mixed distortions composed of poisson noise, gaussian noise and JPEG compression in pristine images to simulate varying-quality images, each distortion with 3 levels, resulting in $3 \times 3 \times 3 \times 5 \times 20 = 2,700$ distorted images in total. Samples of part of the distorted images are shown in (a) of Fig. 2.

Subsequently, 5 radiologists participated in our subjective experiment, evaluating the image quality from its distortion and diagnostic value. The scoring was based on a 5-point Likert scale (0 to 4). Higher scores indicate better PET/CT image quality. The average score of these 5 radiologists was computed as the final score. We split PET-CT-IQA-DS into training (80%) and testing (20%) sets. To avoid data leakage, images in these two sets are from different patients.

We also conducted an experiment on LDCTIQAC2023, a publicly available CT IQA dataset used in the Low-dose Computed Tomography Perceptual Image Quality Assessment Grand Challenge 2023, consisting of a training set with 1000 CT images and a testing set with 300 CT images.

We trained our MS-IQA model using an Adam optimizer with a learning rate of 1e-5 and a batch size of 16. The experiments were conducted on 2 NVIDIA GeForce RTX 3060 GPUs. Spearman's rank order correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC) are used as the metrics to evaluate the performance of IQA models. SROCC and PLCC indicate the correlation between the model's predicted scores and the ground-truth scores. Both metrics range from -1 to 1. Higher values reflect better performance.

Table 1: Comparison of IQA methods. Entries in red present the best performances. Results of the teams in challenge LDCTIQAC2023 are from [21].

Dataset	Method	SROCC	PLCC
PET-CT-IQA-DS	NIQE [5]	-0.2548	-0.2460
	BRISQUE [6]	0.6506	0.6601
	DIIVINE [7]	0.7412	0.7491
	HyperIQA [28]	0.8965	0.9047
	MUSIQ [29]	0.9118	0.9223
	CONTRIQUE [30]	0.8626	0.8758
	MANIQA [31]	0.9335	0.9476
	TOPIQ [32]	0.9278	0.9372
	Re-IQA [33]	0.9037	0.9093
	ARNIQA [34]	0.9390	0.9463
	LoDa [35]	0.9433	0.9517
	QCN [36]	0.9468	0.9596
	MS-IQA (ours)	0.9559	0.9701
LDCTIQAC2023	gabybaldeon	0.9096	0.9143
	Team Epoch	0.9232	0.9278
	FatureNet	0.9338	0.9362
	CHILL@UK	0.9387	0.9402
	RPI_AXIS	0.9414	0.9434
	agaldran	0.9495	0.9491
	MS-IQA (ours)	0.9542	0.9543

3.2 Experiment Results

Table 1 shows the comparison of our method with several NR-IQA methods on PET-CT-IQA-DS, including SOTA methods from recent two years. Our method achieved the best performance among all the comparison methods on both evaluation metrics, demonstrating the superiority of our method. Compared to the second-best method, our method is 0.0091 higher in SROCC and 0.0105 higher in PLCC. That is because our MSFFM, which incorporates dynamic weighted channel attention, effectively enhances valuable multi-scale information.

Experiment results on LDCTIQAC2023 are also shown in Table 1. Our method outperforms all participating teams in the challenge on both evaluation metrics. Compared to the second-best team, our method achieves 0.0047 higher in SROCC and 0.0052 higher in PLCC. This result demonstrates the generalizability of our model on medical images from different modality.

3.3 Ablation Study

To validate the effect of each module in our model, we conducted an ablation experiment on PET-CT-IQA-DS. The results in Table 2 highlight that removing MSFFM results in the largest performance drop (SROCC $\downarrow$ 0.0534, PLCC $\downarrow$ 0.0568), demonstrating its importance in effectively integrating multi-scale features. Moreover, all ablated versions is inferior to the full model, which demon-

Table 2: Ablation study. RM: ResNet Module, STM: Swin Transformer Module, MSFFM: Multi-Scale Feature Fusion Module, SRM: Score Regressor Module.

RM	STM	MSFFM	SRM	SROCC	PLCC
✓	✓	✓	✓	0.9559	0.9701
✓	✓	✓		0.9440	0.9546
✓	✓		✓	0.9025	0.9133
✓		✓	✓	0.9174	0.9257
	✓	✓	✓	0.9380	0.9473
Stage1	Stage2	Stage3	Stage4	SROCC	PLCC
✓	✓	✓	✓	0.9559	0.9701
✓	✓	✓		0.9457	0.9590
✓	✓			0.9306	0.9421
✓				0.9263	0.9379

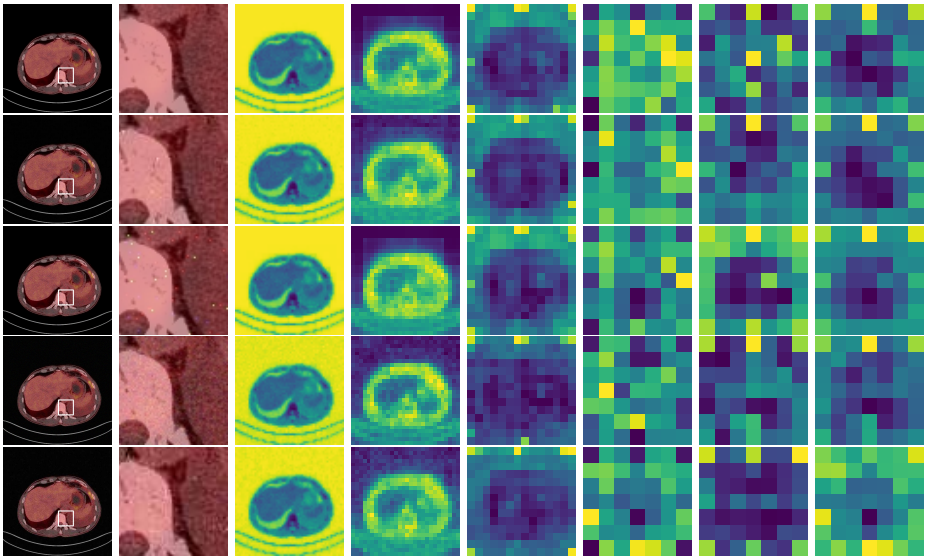
strates the contribution of each module. Table 2 also shows the comparison of different stages used in the model. The performance deteriorates as fewer stages are used (SROCC ↓ 0.0296, PLCC ↓ 0.0322), indicating that both high-level and low-level information contribute to IQA and our feature fusion mechanism effectively enhances the model's performance. Removing stage3 feature results in the largest performance drop (SROCC ↓ 0.0151, PLCC ↓ 0.0169), as it's from the middle of model, better balancing edge details and semantic information.

3.4 Visualization

To intuitively illustrate the process of our model, we visualized the feature maps of key components as shown in Fig. 2. (a) shows several images from our dataset, which are generated by adding varying levels of distortions to a same reference image. (b) shows a magnified detail of the white-boxed region in (a), highlighting the impact of different distortions. (c) to (f) display feature maps F_1 to F_4. It can be seen that lower-level features contain richer edges and noise information, while higher-level features extract more abstract semantic information. (g) and (h) show F_W and F_S, the inputs of weight and score branch. These two features complement each other well, effectively enhancing our model's performance.

4 Conclusion

In this work, We propose MS-IQA, a multi-scale feature fusion network for PET/CT IQA, which combines and utilizes both low-level and high-level features of ResNet and Swin Transformer. We also introduce PET-CT-IQA-DS, which is, to the best of our knowledge, the first PET/CT IQA dataset. Our experiments demonstrate the superiority of our model over existing SOTA NR-IQA methods. Despite the excellent performance on PET/CT and CT IQA, our model lacks a broader evaluation on medical images from more modalities, which will be our next direction of work. In conclusion, we provide an accurate and efficient solution for future PET/CT IQA task.



(a) Image (b) Detail (c) F_1 (d) F_2 (e) F_3 (f) F_4 (g) F_W (h) F_S

Fig. 2: Visualization Samples

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Sureshbabu, Waheeda, and Osama Mawlawi. "PET/CT imaging artifacts." Journal of nuclear medicine technology 33.3 (2005): 156-161.
2. Qi, Chi, et al. "An artificial intelligence-driven image quality assessment system for whole-body [18F] FDG PET/CT." European Journal of Nuclear Medicine and Molecular Imaging 50.5 (2023): 1318-1328.
3. Wang, Zhou, and Alan Conrad Bovik. Modern image quality assessment. Diss. Morgan & Claypool Publishers, 2006.
4. Chow, Li Sze, and Raveendran Paramesran. "Review of medical image quality assessment." Biomedical signal processing and control 27 (2016): 145-154.
5. Mittal, Anish, Rajiv Soundararajan, and Alan C. Bovik. "Making a “completely blind” image quality analyzer." IEEE Signal processing letters 20.3 (2012): 209-212.
6. Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." IEEE TIP 21.12 (2012): 4695-4708.
7. Zhang, Yi, et al. "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes." Signal processing: image communication 29.7 (2014): 725-747.
8. Bosse, Sebastian, et al. "Deep neural networks for no-reference and full-reference image quality assessment." IEEE TIP 27.1 (2017): 206-219.
9. Bianco, Simone, et al. "On the use of deep learning for blind image quality assessment." Signal, Image and Video Processing 12 (2018): 355-362.

10. Bianco, Simone, et al. "On the use of deep learning for blind image quality assessment." Signal, Image and Video Processing 12 (2018): 355-362.
11. Huang, Yipo, et al. "Explainable and generalizable blind image quality assessment via semantic attribute reasoning." IEEE TMM 25 (2022): 7672-7685.
12. Li L, Huang Y, Wu J, et al. Predicting the quality of view synthesis with color-depth image fusion[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(7): 2509-2521.
13. Pal, Debashish, Bhavik Patel, and Adam Wang. "SSIQA: Multi-task learning for non-reference CT image quality assessment with self-supervised noise level prediction." 2021 IEEE 18th ISBI. IEEE, 2021.
14. Lee, Wonkyeong, et al. "Performance evaluation of image quality metrics for perceptual assessment of low-dose computed tomography images." Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment. Vol. 12035. SPIE, 2022.
15. Lee, Wonkyeong, et al. "Low-dose computed tomography perceptual image quality assessment." Medical Image Analysis 99 (2025): 103343.
16. LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
17. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
18. Golestaneh, S. Alireza, Saba Dadsetan, and Kris M. Kitani. "No-reference image quality assessment via transformers, relative ranking, and self-consistency." Proceedings of the IEEE/CVF WACV. 2022.
19. Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep learning in medical image analysis." Annual review of biomedical engineering 19.1 (2017): 221-248.
20. Suk, Heung-Il, and Dinggang Shen. "Deep learning-based feature representation for AD/MCI classification." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16. Springer Berlin Heidelberg, 2013.
21. Lee, Wonkyeong, et al. "Low-dose computed tomography perceptual image quality assessment." Medical Image Analysis 99 (2025): 103343.
22. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
23. Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF ICCV. 2021.
24. Xiang, Xin, et al. "AGCA: An adaptive graph channel attention module for steel surface defect detection." IEEE TIM 72 (2023): 1-12.
25. Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
26. Li, P., Wang, S., Li, T., Lu, J., HuangFu, Y., & Wang, D. (2020). A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis (Lung-PET-CT-Dx) [Data set].
27. Elbakri, Idris A., and Jeffrey A. Fessler. "Statistical image reconstruction for polyenergetic X-ray computed tomography." IEEE TMI 21.2 (2002): 89-99.
28. Su, Shaolin, et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network." Proceedings of the IEEE/CVF CVPR. 2020.
29. Ke, Junjie, et al. "Musiq: Multi-scale image quality transformer." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
30. Madhusudana, Pavan C., et al. "Image quality assessment using contrastive learning." IEEE Transactions on Image Processing 31 (2022): 4149-4161.

31. Yang, Sidi, et al. "Maniqa: Multi-dimension attention network for no-reference image quality assessment." Proceedings of the IEEE/CVF CVPR. 2022.
32. Chen, Chaofeng, et al. "Topiq: A top-down approach from semantics to distortions for image quality assessment." IEEE Transactions on Image Processing (2024).
33. Saha, Avinab, Sandeep Mishra, and Alan C. Bovik. "Re-iqa: Unsupervised learning for image quality assessment in the wild." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.
34. Agnolucci, Lorenzo, et al. "Arniqa: Learning distortion manifold for image quality assessment." Proceedings of the IEEE/CVF WACV. 2024.
35. Xu, Kangmin, et al. "Local distortion aware efficient transformer adaptation for image quality assessment." arXiv preprint arXiv:2308.12001 (2023).
36. Shin, Nyeong-Ho, Seon-Ho Lee, and Chang-Su Kim. "Blind image quality assessment based on geometric order learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.