Cross-Domain AI-Generated Image Quality Assessment via Content-Distortion Awareness

Shun Zhu[1], **Xichen Yang**[1*], **Dechun Zhao**[1], **Tianshu Wang**[2], **Yan Zhang**[1], **Tianyin Li**[1], **Xiaobo Shen**[3]
[1]School of Computer and Electronic Information/Artificial Intelligence, Nanjing Normal University
[2]College of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine
[3]School of Computer Science and Engineering, Nanjing University of Science and Technology
{shunzhu, xichen_yang}@njnu.edu.cn

Abstract

With the expanding use of artificial intelligence generated images (AGIs) in scenarios such as gaming, art, and film production, evaluating their quality is essential to ensure their practical utility. To guarantee effective quality measurement, both content and distortion must be considered. However, existing image quality assessment (IQA) methods fail to sufficiently combine the two aspects in their assessments. Meanwhile, the AGIs annotated with credible content and distortion labels are lack, thus, employing cross-domain methods to utilize typically public image datasets is meaningful. Based on the above conclusions, this paper proposed a cross-domain AI-generated IQA via content-distortion awareness (CDAQA). Firstly, the large language model has been utilized to obtain content and distortion expressions of images. Secondly, a dual-stream contrastive language-image pre-training module (DCLIP) has been designed to gain both quality-aware features in images and the corresponding expressions that related to content and distortion, respectively. Thirdly, a graph knowledge transferring module (GKT) that constructs graphs via inter-domain content relevance has been proposed. GKT can update distortion information in the source domain according to the target domain representations. Finally, the cross-domain framework has been designed to update the existing IQA model for AGIs. And, the contrastive loss of DCLIP, feature update loss of GKT, and quality prediction loss are coordinated to jointly update different modules, enabling sufficient fusion of target domain knowledge. Experimental results demonstrate that CDAQA achieves higher accuracy and stability in cross-domain AGIs tasks.

Code-https://github.com/dart-into/CDAQA

1 Introduction

Artificial intelligence generated images (AGIs) are widely used in fields such as entertainment, education, and social media [Aktay, 2022]. However, AGIs can be effected by distortions, such as artifacts, color cast, and texture coarseness [Li *et al.*, 2023]. Measuring the quality of AGIs is crucial for screening data suitable for different application scenarios. Most image quality assessment (IQA) methods [Min *et al.*, 2025; Liu *et al.*, 2022; Ke *et al.*, 2021] require sufficient images with reliable labels for model training. The limited scale of available AGI datasets can not support model updates for AGI Quality Assessment (AQA). Considering that public datasets in classic scenarios [Lin *et al.*, 2019] can offer effective knowledge for AQA, designing cross-domain IQA methods to utilize them for updating AGI models is meaningful.

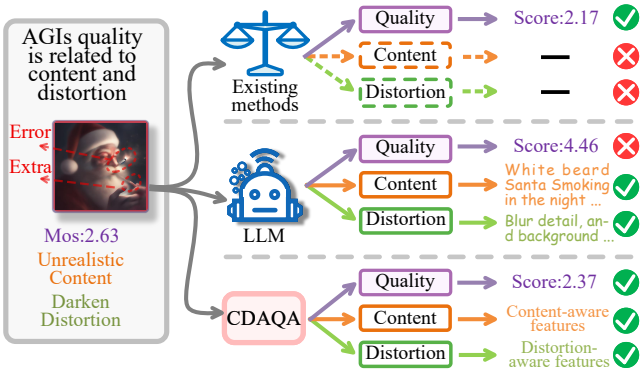

Figure 1: Comparisons of quality, content, and distortion analysis in AGIs tasks.

The quality fluctuations of AGIs are related to both content and distortion information [Li *et al.*, 2023]. Sufficiently combining the content and distortion information is important to improve the accuracy of AQA. However, existing cross-domain IQA methods [Li *et al.*, 2024a] mainly focus on analyzing the relationship between the source domain and target domain in single perspective. Moreover, obtaining content and distortion labels in various scenarios is challenging, making it difficult for model training to learn related knowledge. Large language models (LLMs) [Naveed *et al.*, 2025] have advantage in analyzing semantics of images across different scenarios, thus can provide descriptions from perspectives such as content, attributes, and quality. Employing LLMs to obtain content and distortion descriptions can offer label support for cross-domain AQA methods. As illustrated in Fig-

*Corresponding author

ure 1, the proposed method can effectively utilize the advantage of LLM for content and distortion analysis, which solve the shortcomings of existing IQA metrics.

The effectiveness of extracted content and distortion information requires the model to comprehensively combine both images and related text. As a classic vision-language processing method, contrastive language-image pre-training (CLIP) is widely applied to vision-language interactive tasks. CLIP-based IQA methods [Tang *et al.*, 2024; Zhang *et al.*, 2023] can accurately analyze quality fluctuations. Combining CLIP with cross-domain AQA methods can effectively extract content and distortion information from images and the text.

Knowledge transferring between domains is essential to utilize the source domain to address target domain tasks. As a classic strategy for knowledge sharing, graph neural networks (GNNs) are widely applied in vision scenarios [Zhang, 2025; Yang *et al.*, 2025]. The special structure of GNNs ensures dynamically updating features [Yang *et al.*, 2024; Fang *et al.*, 2023]. Some GNN-based methods [Li *et al.*, 2022] can effectively enhance features through knowledge learning. Employing GNNs to cross-domain AQA tasks can theoretically achieve knowledge transferring between domains.

Moreover, the update and learning strategies of cross-domain AQA models affect the performance. Existing methods [Li *et al.*, 2024a; Agnolucci *et al.*, 2024] mainly focus on updating the IQA model, neglecting the impact of assistant modules in IQA tasks. By constructing the joint updates of different modules, some methods [Zhang *et al.*, 2023; Ou *et al.*, 2021] have gained improvement. Therefore, comprehensively consider the update requirements of different modules is meaningful to cross-domain AQA methods.

Based on the above conclusions, this paper proposes a cross-domain AI-generated IQA via content-distortion awareness (CDAQA). Firstly, the LLM has been utilized to obtain content and distortion expressions of images. Secondly, a dual-stream CLIP module (DCLIP) has been designed to gain both content and distortion information in images and the corresponding expressions. Thirdly, a graph knowledge transferring module (GKT) that constructs graphs via inter-domain content relevance has been proposed to update distortion information in the source domain according to the target domain. Finally, the contrastive loss of DCLIP, feature update loss of GKT, and quality prediction loss are coordinated to jointly update different modules to suit for the target domain. Experimental results demonstrate that CDAQA achieves higher accuracy and stability in cross-domain AGIs tasks. Our work contributions are summarized as follows.

- A DCLIP module is proposed to extract both content and distortion information in images and the corresponding expressions that obtained by LLM.
- A GKT module is designed to update inter-domain distortion information, which constructs graphs via content relevance between the source domain and target domain.
- The contrastive loss, feature update loss, and quality prediction loss are combined to jointly update different modules to suit for AGIs.
- Experimental results demonstrate that CDAQA achieves enhanced performance in cross-domain AGIs tasks.

2 Related Work

2.1 AQA methods

Classic IQA methods utilized handcrafted features to simulate human subjective evaluation [Mittal *et al.*, 2012a]. The development of deep learning have advanced the application of deep networks to IQA [Zhang *et al.*, 2023; Zhu *et al.*, 2020; Su *et al.*, 2020]. These methods mainly focus on perceptual information to process classical scenarios datasets.

With the widespread adoption of generative artificial intelligence, the application scenarios of IQA extend to AGIs. However, the quality fluctuation of AGIs is related to both content and distortion information [Li *et al.*, 2023], which is overlooked by existing methods. As a classic vision-language processing method, CLIP has been employed in some IQA methods to extract related features with the guidance of text. LIQE [Zhang *et al.*, 2023] leverages the CLIP combined with textual quality descriptions to enhance the perception of quality features. CLIP-AGIQA [Tang *et al.*, 2024] utilizes CLIP to effectively analyze keywords in text to extract features of subjects. Employing CLIP in AQA methods to combine images and text for comprehensive content and distortion information extraction is feasible.

In conclusion, designing AQA methods to measure AGIs quality fluctuation is meaningful. Meanwhile, CLIP can effectively analyze content and distortion information for AGIs.

2.2 Cross-domain Learning

The various scenarios require IQA models to address cross-domain tasks. In computer vision applications, cross-domain methods primarily employ strategies such as discrepancy measurement [Long *et al.*, 2015; Long *et al.*, 2017], adversarial generation [Ajakan *et al.*, 2014], and feature reconstruction [Lu *et al.*, 2024]. Employing effective cross-domain strategies for IQA methods is important to improve the performance in cross-domain tasks.

The small scale AGI datasets require cross-domain IQA methods to learn more effective knowledge from classic scenarios. Among classical cross-domain IQA methods, UCDA [Chen *et al.*, 2021b] focuses on the confidence discrepancy in quality perception between domains. DGQA [Li *et al.*, 2024a] aligns and filters images from both domains based on their distortion distributions. RankDA [Chen *et al.*, 2021a] and StyltAM [Lu *et al.*, 2024] leverage quality-aware features across domains for domain-adaptive updates. These methods mainly focus on domain information in single perspective. However, the quality fluctuation in AGIs is related to both content and distortion information, which need cross-domain methods to comprehensively analyze quality-aware features.

Knowledge sharing between domains is crucial for models to address cross-domain tasks. GNNs possess strong capabilities for information sharing and feature enhancement. The spatio-temporal attention module proposed by GESTA [Cheng *et al.*, 2023] utilizes GNNs to enhance the consistency of spatial-temporal representations. GraphIQA [Li *et al.*, 2022] employs distortion types and levels to construct graph features, thereby strengthening the representation of quality-related information. Employing GNNs to cross-domain IQA methods can effectively fuse information across domains.

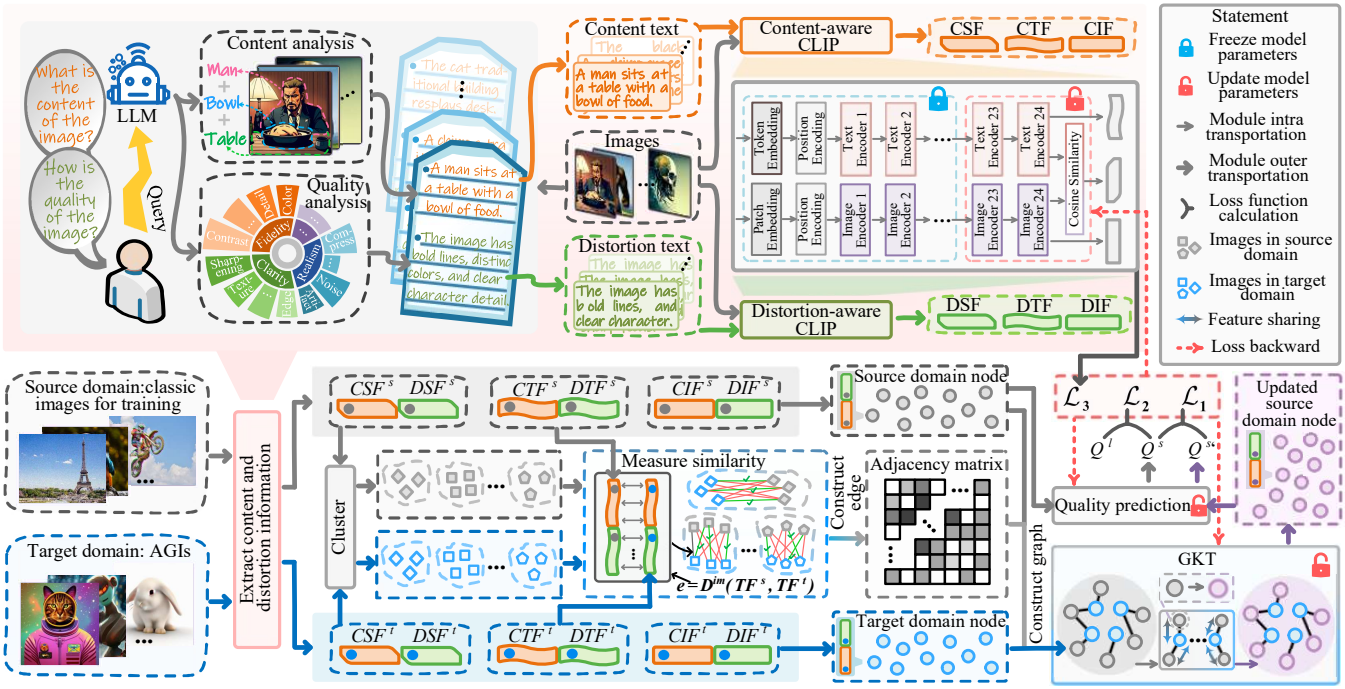

Figure 2: The illustration of CDAQA. CDAQA consists of content and distortion expression, feature extraction, GKT, and model training.

Moreover, the learning and update strategies of different modules also impact cross-domain performance. DGQA [Li *et al.*, 2024a] considers only the update of the quality assessment model, overlooking the co-adaptation of the integrated distortion classifier. The joint updates of different modules in LIQE [Zhang *et al.*, 2023] and SDD-FIQA [Ou *et al.*, 2021] lead to the improvement of performance. Therefore, comprehensively considering the update requirements of different modules is essential for cross-domain IQA.

In conclusion, employing GNNs to share knowledge between source domain and target domain is meaningful to improve the performance of models in cross-domain tasks. Meanwhile, effective update strategies can guide different modules for sufficient learning.

3 Method

CDAQA consists of content and distortion expression, feature extraction, GKT, and model training. The overall structure of CDAQA is illustrated in Figure 2.

3.1 Content and Distortion Expression

CDAQA utilizes LLMs to annotate images with textual description to solve the lack of related labels. Due to the strong performance in image and semantics analysis, ByteDance's generative AI model is employed to obtain content and distortion textual labels. We utilize ByteDance's generative AI model to analyze the logical relationships between the main subject and components in images, thus can obtain a comprehensive description as content textual label. Additionally, ByteDance's generative AI model is used to analyze distortions of images from various perspectives such as clarity and fidelity, thus can obtain description as distortion textual label. Specifically, the length and style of generated text are standard. Some research [Li *et al.*, 2023] proposed that the text length can reflect the performance of models. When the text is too long, the subsequent model hardly analyze the semantics. Based on the preliminary comparison, the text length is limit to between 10 and 20 words. To ensure the effective expression in text, at least one primary subject and one secondary subject are included in content textual label. And, the modifier is needed to subjects. Two to four most prominent distortion angles of the image are described in distortion textual label. The details and lines are also noticed to better express AGIs. The content and distortion textual labels of examples are illustrated in Figure 3.

3.2 Feature Extraction

CDAQA employs DCLIP to extract content and distortion information. DCLIP consists of content-aware CLIP and distorion-aware CLIP, and each following the standard CLIP architecture [Hafner *et al.*, 2021]. The inputs of content-aware CLIP are images and corresponding content textual labels, while the inputs of distortion-aware CLIP are images and corresponding distortion textual labels. The patch embedding, token embedding, position embedding, and the first 22 encoder layers in each stream are frozen to ensure the preservation of features and reduce catastrophic forgetting. The last two encoder in each stream are trained to learn knowledge. The content-aware CLIP outputs content text features CTF, content similarity features CSF, and content image features CIF. Correspondingly, the distortion-aware CLIP outputs a distortion text feature DTF, a distortion similarity feature DSF, and a distortion image feature DIF.

Question	What is the content of the iamge?	How is the quality of the image?
	A girl with braids sits on green grass in a forest with small mushrooms.	The image shows clear details, soft colors, and distinct lines.
	Three people in red Santa hats sit around a decorated Christmas table.	The image has vivid colors, clear lines, but with unrealistic details.
	A blonde anime character stands in a smoky, fiery environment.	The image has warm colors, soft focus, but blur and hazy background.
	The image shows abstract shapes with bright yellow and brown colors.	The image has blurred edges, and the dark brown abstract forms.

Figure 3: The illustration of content and distortion text.

3.3 GKT

CDAQA proposes GKT to construct graph and update the source domain to suit for the target domain. CTF, CSF, CIF, DTF, DTF, and DSF are utilized for graph construction. The images $\{\mathcal{I}_i, i \in (1,m)\}$ in classic scenes and AGIs $\{\mathcal{I}_j, j \in (1,n)\}$ are used as the source domain $\mathcal{D}om_s$ and the target domain $\mathcal{D}om_t$, respectively. The source domain features $\{CTF_i^s, CSF_i^s, CIF_i^s\}$ and target domain features $\{CTF_j^t, CSF_j^t, CIF_j^t\}$ are obtained by DCLIP. Both the source and target domain concatenate the content and distortion features. Specifically, CSF_i^s and DSF_i^s are concatenated for SF_i^s. CTF_i^s and DTF_i^s are concatenated for TF_i^s. CIF_i^s and DIF_i^s are concatenated for IF_i^s. Correspondingly, concatenated features of the target domain are SF_j^t, TF_j^t, and IF_j^t. SF_i^s and SF_i^t are clustered via k-means to generate c clusters C_{si} and C_{tj} of the source and target domain images, respectively. This enables analyze the distributions of domains according to semantics. Then, distances $\mathcal{D}_{a,b}^{cls}$ between source domain clusters centers $f^a, a \in (1,c)$ and target domain clusters centers $f^b, b \in (1,c)$ are calculated as follows:

$$\mathcal{D}_{a,b}^{cls} = ||f^a - f^b||. \tag{1}$$

The source domain cluster closest to the target domain cluster is considered the same cluster. In the same cluster, the distances between the source domain images and the target domain images are calculated by TF_i^s and TF_j^t:

$$\mathcal{D}_{i,j}^{im} = \sqrt{(v_i - v_j)^T (\frac{co_i - co_j}{2})^{-1} (v_i - v_j)}, \tag{2}$$

where v_i and v_j represent the mean vectors of TF_i^s and TF_j^t, respectively, and co_i and co_j represent the covariance matrices of TF_i^s and TF_j^t, respectively. For each target domain image, the two closest source domain images are selected as its neighboring nodes, and the corresponding distances $\mathcal{D}_{est,j}^{im}$ serve as edges $e_{est,j}$. Edges between source and target domain images that are not identified as neighbors are assigned the value of 0. The adjacency matrix $\mathcal{A}_{st}$ consisting of all edges is as follows:

$$\mathcal{A}_{st} = \{e_{est,j} = \mathcal{D}_{est,j}^{im} | j \in (1,n)\}. \tag{3}$$

IF^s and IF^t are used as node features of the source domain and target domain, respectively. The whole node feature matrix $\mathcal{H}_{st}$ is as follows:

$$\mathcal{H}_{st} = \{\{IF_i^s\}, \{IF_j^t\} | i \in (1,m), j \in (1,n)\}. \tag{4}$$

The GKT module is used for updating $\mathcal{H}_{st}$ according to $\mathcal{A}_{st}$, which consists of graph convolutions and activation function σ. The update process in the graph convolution is as follows:

$$\mathcal{H}_{st}^{'} = \sigma(\mathcal{A}_{st}\mathcal{H}_{st}\mathcal{W}_1 + \mathcal{B}_1), \tag{5}$$

where $\mathcal{H}_{st}^{'}$ and $\mathcal{W}_1$ are the updated node feature matrix and weight matrix of the first graph convolution. $\mathcal{B}_1$ represents the bias matrix. The updated source domain node feature $\{IF_i^{s'}\}$ is separated from $\mathcal{H}_{st}^{'}$.

Algorithm 1 The process of CDAQA

Input: $\{\mathcal{I}_i, i \in (1,m)\}$, and $\{\mathcal{I}_j, j \in (1,n)\}$.
Output: $\{\mathcal{Q}_j^t\}$.

$\{SF_i^s, TF_i^s, IF_i^s\} = DCLIP[\{\mathcal{I}_i\}, LLM(\{\mathcal{I}_i\})]$.
$\{SF_j^t, TF_j^t, IF_j^t\} = DCLIP[\{\mathcal{I}_j\}, LLM(\{\mathcal{I}_j\})]$.
$\{\{C_{si}\}, \{C_{tj}\}\} \overset{k-means}{\longleftarrow} \{\{SF_i^s\}, \{SF_j^t\}\}$.
if $\{C_{si}\}$ close to $\{C_{tj}\}$ **do**
  Calculate $\mathcal{D}_{i,j}^{im}(TF_i^s, TF_j^t)$.
  $e_{est,j} \overset{top\ \ two}{\longleftarrow} \{\mathcal{D}_{i,j}^{im}\}$.
$\mathcal{A}_{st} = \{e_{est,j} | j \in (1,n)\}$.
$\mathcal{H}_{st} = \{\{IF_i^s\}, \{IF_j^t\} | i \in (1,m)\}$.
$\mathcal{H}_{st}^{'} = \sigma(\mathcal{A}_{st}\mathcal{H}_{st}\mathcal{W}_1 + \mathcal{B}_1)$, select $\{IF_i^{s'}\} \in \mathcal{H}_{st}^{'}$.
$\{\{\mathcal{Q}_i^s\}, \{\mathcal{Q}_i^{s'}\}\} = Linear(\{\{IF_i^s\}, \{IF_i^{s'}\}\})$.
$\mathcal{L}_1 = \sum_{i=1}^{m} \left\| \mathcal{Q}_i^s - \mathcal{Q}_i^{s'} \right\|^2$.
$\mathcal{L}_2 = \sum_{i=1}^{m} \left\| \mathcal{Q}_i^s - \mathcal{Q}_i^l \right\|^2$.
$\mathcal{L}_3 = \sum_{1}^{4} \{\mathcal{L}_{SF}\}/4$.
$\mathcal{L}_{total} = \alpha * \mathcal{L}_1 + \beta * \mathcal{L}_2 + \gamma * \mathcal{L}_3$.
Update $\mathcal{G}_{IQA}$ and obtain θ_{IQA}.
$\{\mathcal{Q}_j^t\} = \mathcal{G}_{IQA}(\theta_{IQA}, \{\mathcal{I}_j\})$.
return $\{\mathcal{Q}_j^t\}$

3.4 Model Training

CDAQA utilizes loss feedback information of different modules to joint update the whole framework. The difference between $\{IF_i^s\}$ and $\{IF_i^{s'}\}$ is analyzed to obtain loss feedback to guide different modules to transfer to the target domain. $\{IF_i^s\}$ and $\{IF_i^{s'}\}$ are input to the quality prediction module which consists of four linear layers to get the quality score $\{\mathcal{Q}_i^s\}$ and $\{\mathcal{Q}_i^{s'}\}$, respectively. $\{\mathcal{Q}_i^{s'}\}$ are related the quality fluctuation knowledge of the target domain. Then, the differences between $\{\mathcal{Q}_i^s\}$ and $\{\mathcal{Q}_i^{s'}\}$ are calculated as loss information $\mathcal{L}_1$ to obtain the knowledge of the target domain:

$$\mathcal{L}_1 = \sum_{i=1}^{m} \left\| \mathcal{Q}_i^s - \mathcal{Q}_i^{s'} \right\|^2. \tag{6}$$

	Train	KADID				TID2013					
	Test	AGIQA-3k		AGIQA-20k		AGIQA-3k		AGIQA-20k		Average	
	Methods *(year)*	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Hand-craft methods	BRISQUE *(2012)*	0.210	0.244	0.118	0.154	0.148	0.180	0.120	0.139	0.149	0.179
	NIQE *(2012)*	0.326	0.371	0.416	0.425	0.225	0.263	0.306	0.324	0.318	0.346
	PIQE *(2015)*	0.176	0.249	0.164	0.180	0.253	0.301	0.145	0.147	0.184	0.219
Classic deep-learning methods	RankIQA *(2017)*	0.317	0.376	0.402	0.385	0.326	0.357	0.290	0.334	0.334	0.363
	MetaIQA *(2020)*	0.510	0.546	0.528	0.517	0.429	0.450	0.464	0.479	0.483	0.498
	DBCNN *(2023)*	0.457	0.492	0.324	0.328	0.364	0.402	0.268	0.300	0.353	0.381
	HyperIQA *(2020)*	**0.725**	0.750	**0.712**	**0.725**	**0.712**	**0.740**	0.359	0.374	0.627	0.647
	AKD-IQA *(2025)*	0.422	0.468	0.405	0.423	0.535	0.567	0.502	0.509	0.466	0.492
	DiffV^2IQA *(2025)*	0.540	0.579	0.358	0.359	0.475	0.502	0.471	0.478	0.461	0.480
	MCOLE *(2024)*	0.489	0.531	0.680	0.694	0.505	0.538	0.424	0.437	0.525	0.550
Cross-domain methods	DANN *(2014)*	0.327	0.378	0.410	0.425	0.460	0.491	0.432	0.449	0.407	0.436
	UCDA *(2021)*	0.221	0.257	0.234	0.250	0.574	0.607	0.397	0.418	0.357	0.383
	RankDA *(2021)*	0.562	0.580	0.573	0.575	0.480	0.529	0.531	0.548	0.537	0.558
	StyleAM *(2024)*	0.619	0.649	0.611	0.618	0.547	0.563	**0.574**	**0.590**	0.588	0.605
	DGQA *(2024)*	0.716	**0.755**	0.707	0.716	0.612	0.647	0.512	0.526	**0.637**	**0.661**
	CDAQA	**0.778**	**0.795**	**0.748**	**0.775**	**0.741**	**0.775**	**0.640**	**0.659**	**0.727**	**0.751**

Table 1: Cross-domain performance comparisons on AGIs.

The distance between labels $\{\mathcal{Q}_i^l\}$ and $\{\mathcal{Q}_i^s\}$ is also part of loss feedback information. In this way, different modules can be balanced to better fuse the target domain and the source domain. The whole framework can learn accurate knowledge from the quality prediction in source domain. Thus, loss information $\mathcal{L}_2$ is calculated as follows:

$$\mathcal{L}_2 = \sum_{i=1}^{m} \left\| \mathcal{Q}_i^s - \mathcal{Q}_i^l \right\|^2. \tag{7}$$

Moreover, the contrastive loss of DCLIP is employed to improve the overall framework to learn content-aware and distortion-aware knowledge. The DCLIP can also better extract features by combining images and text. The calculation of $\mathcal{L}_3$ is as follows:

$$\begin{aligned}\mathcal{L}_{SF} = -\frac{1}{2N}[\sum_{x=1}^{N} \log \frac{\exp(SF_{x,y}/\tau)}{\sum_{y=1}^{N} \log\{\exp(SF_{x,y}/\tau)\}} \\ +\sum_{y=1}^{N} \log \frac{\exp(SF_{x,y}/\tau)}{\sum_{x=1}^{N} \log\{\exp(SF_{x,y}/\tau)\}}],\end{aligned} \tag{8}$$

$$\mathcal{L}_3 = \sum_{1}^{4}\{\mathcal{L}_{SF}\}/4, \tag{9}$$

where N represents the batch size of images. x and y are the row and column indices of similarity matrix, respectively. τ is a learnable temperature parameter. And, SF represents content and distortion similarity matrix of the source domain and target domain, $SF \in \{CSF^s, DSF^s, CSF^t, DSF^t\}$.

Combining the $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_2$, the $\mathcal{L}_{total}$ is obtained as follows:

$$\mathcal{L}_{total} = \alpha * \mathcal{L}_1 + \beta * \mathcal{L}_2 + \gamma * \mathcal{L}_5, \tag{10}$$

where α, β and γ represent the weight of $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$, respectively. $\mathcal{L}_{total}$ is used to jointly update DCLIP, GKT, and quality prediction module. The whole IQA framework $\mathcal{G}_{IQA}$ is updated with new parameters θ_{IQA}. The quality scores $\{\mathcal{Q}_j^t\}$ of AIGs $\{\mathcal{I}_j\}$ can be obtained by $\mathcal{G}_{IQA}$ and θ_{IQA}. The process of CDAQA is described in Algorithm 1.

4 Experiments

4.1 Experimental Settings

This section includes the expression of datasets, metrics, and parameters details.

Datasets and Metrics. We conducted cross-domain IQA experiments on two AGIs datasets (AGIQA-3k [Li *et al.*, 2023] and AGIQA-20k [Li *et al.*, 2024b]). And, 4 natural scenarios datasets, namely, (KADID [Lin *et al.*, 2019], CSIQ [Larson and Chandler, 2010], TID2013 [Ponomarenko *et al.*, 2013], and LIVEC [Ghadiyaram and Bovik, 2015]) were employed for cross-domain training and test. Moreover, 2 underwater datasets (SAUD [Jiang *et al.*, 2022] and UIED [Zheng *et al.*, 2022]) were utilized for cross-domain test in more scenarios. Two typical metrics, namely, Spearman rank correlation coefficient (SRCC) and Pearson linear correlation coefficient (PLCC), were employed to evaluate the monotonicity and accuracy of the results.

Parameters Details. During training, a random 224×224 patch was sampled from each image, and data augmentation was performed by random horizontal flipping. The RGB features of these patches were input into the quality assessment net. The network was trained with a batch size of 16, and the learning rate η was set to 1×10^{-4}. Additionally, weight decay of the learning rate was applied every 10 epochs, and the weight decay is 0.8. The value of parameters c, α, β, and γ are 5, 0.30, 0.75, and 0.25 in the baseline.

4.2 Cross-domain Performance on AGIs

To demonstrate the performance in cross-domain AGIs tasks, CDAQA was compared with 15 IQA methods, including 3 handcrafted methods, namely, BRISQUE [Mittal *et al.*, 2012a], NIQE [Mittal *et al.*, 2012b] and PIQE [Venkatanath *et al.*, 2015], 7 classic deep learning-based methods, namely, RankIQA [Liu *et al.*, 2017], MetaIQA [Zhu *et al.*, 2020], DBCNN [Zhang *et al.*, 2023], HyperIQA [Su *et al.*, 2020],

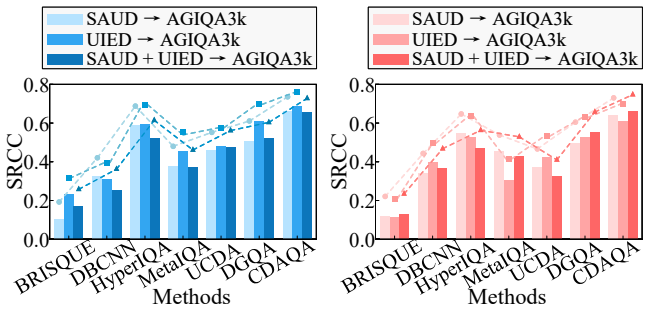

Figure 4: Cross-Domain performance comparisons from underwater to AGIs datasets.

AKD-IQA [Hu *et al.*, 2025], DiffV^2IQA [Wang *et al.*, 2025], and MCOLE [Jiang *et al.*, 2024], and 5 cross-domain methods, namely, DANN [Ajakan *et al.*, 2014], UCDA [Chen *et al.*, 2021b], RankDA [Chen *et al.*, 2021a], StyleAM [Lu *et al.*, 2024] and DGQA [Li *et al.*, 2024a]. Specifically, handcrafted methods are compared to validate the advantage of deep-learning. Classic deep learning-based methods are employed to validate the effectiveness of cross-domain applications. Cross-domain methods are compared to validate the advantage of our cross-domain framework. All methods trains on KADID and TID2013, and then tests on AGIQA-3k and AGIQA-20k, respectively. The results are illustrated in Table 1 and the best results are in bold. The experimental results demonstrate that the SRCC and PLCC of CDAQA are at least 4.73% and 6.89% higher than these methods on AGIQA-3k and AGIQA-20k, respectively. The average values of the SRCC and PLCC of CDAQA are also the best. Therefore, CDAQA perform best in cross-domain AGIs and classic scenarios.

To further demonstrate the performance in cross-domain AGIs tasks, experiments of transferring underwater datasets to AGIs are conducted in Figure 4. Six methods are employed for comparisons. Results demonstrate that compared to other methods, our method improves by at least 10.69% and 13.28% in AGIQA-3k and AGIQA-20k, respectively. Although underwater and AGIs scenarios differ more significantly, CDAQA performs best in cross-domain AGIs tasks.

4.3 Ablation Study

This section includes experiments for demonstrating the effectiveness of DCLIP, GKT, and joint learning, respectively. All experiments are trained on KADID and tested on AGIQA-3K and LIVEC. The best results are in bold.

Effectiveness of CACLIP, DACLIP, and GKT. To validate the effectiveness of DCLIP and GKT, Table 2 is illustrated. In Table 2, scheme #1, scheme #2, only utilize one of the CACLIP, DACLIP, respectively. Scheme #3, scheme #4, and scheme #5 only remove one of the CACLIP, DACLIP, and GKT, respectively. The scheme #7 is the baseline of CDAQA. The experimental results demonstrate that one of three modules can improve the model by 17.82%, 6.38%, and 14.31%, respectively. Therefore, the CACLIP, DACLIP, and GKT in CDAQA are effective in measuring image quality.

Effectiveness of Joint Learning. To validate the effectiveness of joint learning, Table 3 are illustrated. In Table 3, scheme #1, scheme #2, and scheme #3 only utilize one of the $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$, respectively. Scheme #4, scheme #5, and scheme #6 only remove one of the $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$, respectively. The experimental results demonstrate that one of three loss can improve the model by 11.32%, 20.98%, and 17.82%, respectively. Therefore, the $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ are effective in cross-domain tasks.

#	CACLIP	DACLIP	GKT	SRCC/PLCC AGIQA-3k	SRCC/PLCC LIVEC
1	✓	✗	✗	0.425/0.448	0.417/0.436
2	✗	✓	✗	0.567/0.574	0.624/0.637
3	✓	✗	✓	0.654/0.679	0.721/0.723
4	✗	✓	✓	0.646/0.671	0.651/0.659
5	✓	✓	✗	0.660/0.705	0.671/0.676
6	✓	✓	✓	**0.778/0.795**	**0.767/0.770**

Table 2: Effectiveness of CACLIP, DACLIP, and GKT.

#	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	SRCC/PLCC AGIQA-3k	SRCC/PLCC LIVEC
1	✓	✗	✗	0.327/0.358	0.306/0.311
2	✗	✓	✗	0.628/0.651	0. 634/0.639
3	✗	✗	✓	0.401/0.436	0.371/0.382
4	✓	✓	✗	0.662/0.689	0.651/0.676
5	✗	✓	✓	0.714/0.749	0.689/0.694
6	✓	✗	✓	0.420/0.431	0.394/0.407
7	✓	✓	✓	**0.778/0.795**	**0.767/0.770**

Table 3: Effectiveness of $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$.

Test	Method	KADID	CSIQ	TID2013
SAUD	BRISQUE	0.216	0.159	0.243
	DBCNN	0.418	0.351	0.338
	HyperIQA	0.497	0.529	**0.591**
	MetaIQA	0.586	0.510	0.569
	UCDA	0.623	0.527	0.498
	DGQA	**0.627**	**0.531**	0.524
	CDAQA	**0.694**	**0.701**	**0.674**
UIED	BRISQUE	0.253	0.105	0.117
	DBCNN	0.348	0.271	**0.438**
	HyperIQA	**0.497**	0.325	0.369
	MetaIQA	0.481	0.397	0.329
	UCDA	0.409	0.317	0.358
	DGQA	0.434	**0.401**	0.409
	CDAQA	**0.563**	**0.505**	**0.493**

Table 4: Cross-domain performance comparisons from natural to underwater datasets.

4.4 Comparison of Cross-Domain Scenarios

To prove the performance of CDAQA in more different scenarios, six methods are compared in Table 4. In Table 4, 3 natural scenario dataset are for training and test on two underwater datasets, namely, SAUD and UIED. The best two results are in bold. The experimental results demonstrate that, compared to other methods, our method improves by

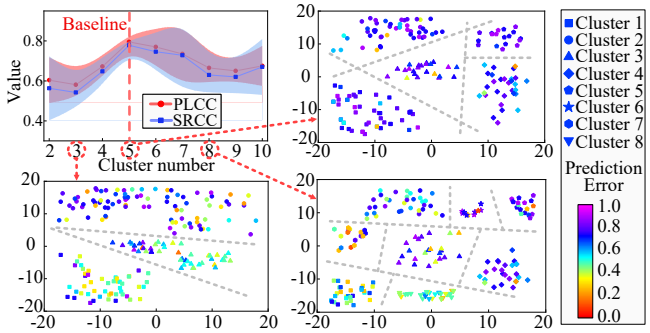

Figure 5: Analysis of different cluster number.

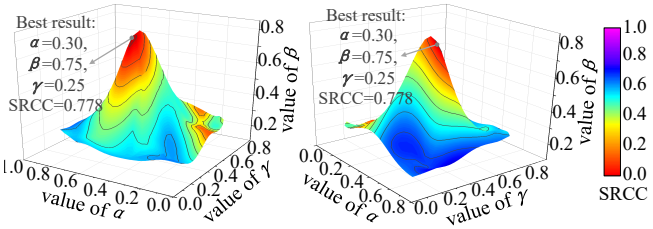

Figure 6: Analysis of different loss weights.

at least 10.69% and 13.28% in SAUD and UIED, respectively. Therefore, CDAQA performs best in cross-domain tasks, which meets various IQA scenarios.

4.5 Parameter Discussions

This section includes experiments for analyzing the parameters of c, α, β, and γ. All experiments are trained on KADID and tested on AGIQA-3K.

Analysis of Cluster Number c. To demonstrate the effectiveness of different cluster number c, Figure 5 is illustrated. In Figure 5, the SRCC, PLCC, and corresponding error band of employing different value of c are compared. The cluster results are illustrated when c is set to 3, 5, 8, respectively. The prediction error of different examples are reflect with rainbow color. Results demonstrate that the result is the best when c is set to 5, and the predicted scores has smallest error. Therefore, the baseline of setting c to 5 is effective.

Analysis of Loss Weights α, β, and γ. To demonstrate the effectiveness of different loss weights, Figure 6 is illustrated. In Figure 6, the SRCC of setting different value of α, β, and γ are recorded in 3D mapping chart. The values of SRCC are reflect with rainbow color. Results demonstrate that higher β and γ improve the performance. And, the performance get best when α, β, and γ are 0.30, 0.75, and 0.25, respectively. Therefore, the baseline of α, β, and γ is effective. The $\mathcal{L}_2$ which related to quality fluctuation significantly effects the performance of CDAQA.

4.6 Visual Analysis

This section includes experiments for analyzing the accuracy and graph construction of CDAQA.

Accuracy Analysis. To demonstrate the accuracy of CDAQA, the predicted scores and labels are plotted as red

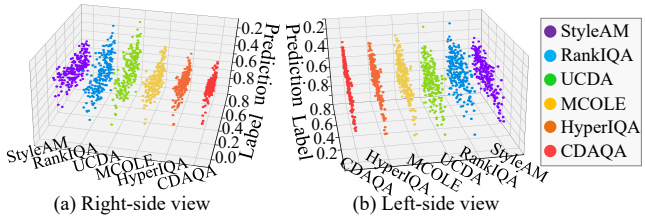

Figure 7: The illustration of scatter plots.

Figure 8: The illustration of graph construction.

scatters in Figure 7, when LIVEC is for test. The scatters of 5 methods are illustrated for comparisons. The results demonstrate that scatters of CDAQA is the most concentrated and predicted scores are close to the labels, which demonstrate that CDAQA is more accurate.

Graph Construction Analysis. To demonstrate the effectiveness of graph construction, several samples are illustrated for visual expression. As illustrated in Figure 8, the edge between the source domain images and target domain images represents similarity. The results demonstrate that the source domain images has relationship with the target domain images in content and distortion.

5 Conclusion

In this paper, CDAQA utilizes both content and distortion information to measure the quality of AGIs. Firstly, the LLM has been employed to gain the content and distortion descriptions of the AGIs. Secondly, a DCLIP module is designed to extract content-distortion-aware features. Thirdly, a GKT module based on inter-domain relevance has been proposed to adapt content and distortion features from the source domain to the target domain. Finally, the contrastive loss of DCLIP, feature update loss of GKT, and quality prediction loss are jointly optimized to update all modules for target domain. Experimental results demonstrate that CDAQA achieves competitive and stable performance in cross-domain AGI quality assessment tasks.

In future work, knowledge sharing in more dimensions will be conducted for measuring the quality of AGIs. Furthermore, the update strategy will be further refined to better handle cross-domain IQA tasks.

Contributions

Shun Zhu and Xichen Yang are co-first authors. And, Xichen Yang is the corresponding author.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62101268, 62472226), and Qinglan Project of Jiangsu Province of China.

References

[Agnolucci *et al.*, 2024] Lorenzo Agnolucci, Leonardo Galteri, Marco Bertini, and Alberto Del Bimbo. Arniqa: Learning distortion manifold for image quality assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 189–198, 2024.

[Ajakan *et al.*, 2014] Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, and Mario Marchand. Domain-adversarial neural networks. *arXiv preprint arXiv:1412.4446*, 2014.

[Aktay, 2022] Sayım Aktay. The usability of images generated by artificial intelligence (ai) in education. *International technology and education journal*, 6(2):51–62, 2022.

[Chen *et al.*, 2021a] Baoliang Chen, Haoliang Li, Hongfei Fan, and Shiqi Wang. No-reference screen content image quality assessment with unsupervised domain adaptation. *IEEE Transactions on Image Processing*, 30:5463–5476, 2021.

[Chen *et al.*, 2021b] Pengfei Chen, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Unsupervised curriculum domain adaptation for no-reference video quality assessment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5178–5187, 2021.

[Cheng *et al.*, 2023] Kai Cheng, Yang Liu, and Xinhua Zeng. Learning graph enhanced spatial-temporal coherence for video anomaly detection. *IEEE Signal Processing Letters*, 30:314–318, 2023.

[Fang *et al.*, 2023] Yin Fang, Qiang Zhang, Ningyu Zhang, Zhuo Chen, Xiang Zhuang, Xin Shao, Xiaohui Fan, and Huajun Chen. Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nature Machine Intelligence*, 5(5):542–553, 2023.

[Ghadiyaram and Bovik, 2015] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2015.

[Hafner *et al.*, 2021] Markus Hafner, Maria Katsantoni, Tino Köster, James Marks, Joyita Mukherjee, Dorothee Staiger, Jernej Ule, and Mihaela Zavolan. Clip and complementary methods. *Nature Reviews Methods Primers*, 1(1):20, 2021.

[Hu *et al.*, 2025] Bo Hu, Wenzhi Chen, Jia Zheng, Leida Li, Wen Lu, and Xinbo Gao. No-reference image quality assessment via inter-level adaptive knowledge distillation. *IEEE Transactions on Broadcasting*, 2025.

[Jiang *et al.*, 2022] Qiuping Jiang, Yuese Gu, Chongyi Li, Runmin Cong, and Feng Shao. Underwater image enhancement quality evaluation: Benchmark dataset and objective metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):5959–5974, 2022.

[Jiang *et al.*, 2024] Qiuping Jiang, Xiao Yi, Li Ouyang, Jingchun Zhou, and Zhihua Wang. Towards dimension-enriched underwater image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[Ke *et al.*, 2021] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021.

[Larson and Chandler, 2010] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006–011006, 2010.

[Li *et al.*, 2022] Wei Li, Junjie Wang, Yunhao Gao, Mengmeng Zhang, Ran Tao, and Bing Zhang. Graph-feature-enhanced selective assignment network for hyperspectral and multispectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[Li *et al.*, 2023] Chunyi Li, Zicheng Zhang, Haoning Wu, Wei Sun, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai, and Weisi Lin. Agiqa-3k: An open database for ai-generated image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[Li *et al.*, 2024a] Aobo Li, Jinjian Wu, Yongxu Liu, and Leida Li. Bridging the synthetic-to-authentic gap: Distortion-guided unsupervised domain adaptation for blind image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28422–28431, 2024.

[Li *et al.*, 2024b] Chunyi Li, Tengchuan Kou, Yixuan Gao, Yuqin Cao, Wei Sun, Zicheng Zhang, Yingjie Zhou, Zhichao Zhang, Weixia Zhang, Haoning Wu, et al. Aigiqa-20k: A large database for ai-generated image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6327–6336, 2024.

[Lin *et al.*, 2019] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019.

[Liu *et al.*, 2017] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1040–1049, 2017.

[Liu *et al.*, 2022] Jianzhao Liu, Xin Li, Shukun An, and Zhibo Chen. Source-free unsupervised domain adaptation for blind image quality assessment. *arXiv preprint arXiv:2207.08124*, 2022.

[Long *et al.*, 2015] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features

with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

[Long *et al.*, 2017] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.

[Lu *et al.*, 2024] Yiting Lu, Xin Li, Jianzhao Liu, and Zhibo Chen. Styleam: Perception-oriented unsupervised domain adaption for no-reference image quality assessment. *IEEE Transactions on Multimedia*, 2024.

[Min *et al.*, 2025] Xiongkuo Min, Yixuan Gao, Yuqin Cao, Guangtao Zhai, Wenjun Zhang, Huifang Sun, and Chang Wen Chen. Exploring rich subjective quality information for image quality assessment in the wild. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[Mittal *et al.*, 2012a] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[Mittal *et al.*, 2012b] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.

[Naveed *et al.*, 2025] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology*, 16(5):1–72, 2025.

[Ou *et al.*, 2021] Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. Sdd-fiqa: Unsupervised face image quality assessment with similarity distribution distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7670–7679, 2021.

[Ponomarenko *et al.*, 2013] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database tid2013: Peculiarities and preliminary results. In *European workshop on Visual Information Processing (EUVIP)*, pages 106–111. IEEE, 2013.

[Su *et al.*, 2020] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020.

[Tang *et al.*, 2024] Zhenchen Tang, Zichuan Wang, Bo Peng, and Jing Dong. Clip-agiqa: boosting the performance of ai-generated image quality assessment with clip. In *International Conference on Pattern Recognition*, pages 48–61. Springer, 2024.

[Venkatanath *et al.*, 2015] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 twenty first National Conference on Communications (NCC)*, pages 1–6. IEEE, 2015.

[Wang *et al.*, 2025] Zhaoyang Wang, Bo Hu, Mingyang Zhang, Jie Li, Leida Li, Maoguo Gong, and Xinbo Gao. Diffusion model-based visual compensation guidance and visual difference analysis for no-reference image quality assessment. *IEEE Transactions on Image Processing*, 2025.

[Yang *et al.*, 2024] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3091–3110, 2024.

[Yang *et al.*, 2025] Liang Yang, Xin Chen, Jiaming Zhuo, Di Jin, Chuan Wang, Xiaochun Cao, Zhen Wang, and Yuanfang Guo. Disentangled graph spectral domain adaptation. In *Forty-second International Conference on Machine Learning*, 2025.

[Zhang *et al.*, 2023] Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14071–14081, 2023.

[Zhang, 2025] Yiwei Zhang. Social network user profiling for anomaly detection based on graph neural networks. In *2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)*, pages 1197–1201. IEEE, 2025.

[Zheng *et al.*, 2022] Yannan Zheng, Weiling Chen, Rongfu Lin, Tiesong Zhao, and Patrick Le Callet. Uif: An objective quality assessment for underwater image enhancement. *IEEE Transactions on Image Processing*, 31:5456–5468, 2022.

[Zhu *et al.*, 2020] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2020.