Semi-Supervised Blind Image Quality Assessment through Knowledge Distillation and Incremental Learning

Wensheng Pan[1 *], Timin Gao[1 *], Yan Zhang[1 †], Xiawu Zheng[1,2], Yunhang Shen[3], Ke Li [3], Runze Hu[4], Yutao Liu[5], Pingyang Dai[1]

[1]Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, Xiamen 361005, China
[2]Peng Cheng Laboratory, Shenzhen 518066, China
[3]Tencent Youtu Lab, Shanghai 200233, China
[4]School of Information and Electronics, Beijing Institute of Technology, Beijing 100086, China
[5]School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China
bzhy986@xmu.edu.cn

Abstract

Blind Image Quality Assessment (BIQA) aims to simulate human assessment of image quality. It has a great demand for labeled data, which is often insufficient in practice. Some researchers employ unsupervised methods to address this issue, which is challenging to emulate the human subjective system. To this end, we introduce a unified framework that combines semi-supervised and incremental learning to address the mentioned issue. Specifically, when training data is limited, semi-supervised learning is necessary to infer extensive unlabeled data. To facilitate semi-supervised learning, we use knowledge distillation to assign pseudo-labels to unlabeled data, preserving analytical capability. To gradually improve the quality of pseudo labels, we introduce incremental learning. However, incremental learning can lead to catastrophic forgetting. We employ Experience Replay by selecting representative samples during multiple rounds of semi-supervised learning, to alleviate forgetting and ensure model stability. Experimental results show that the proposed approach achieves state-of-the-art performance across various benchmark datasets. After being trained on the LIVE dataset, our method can be directly transferred to the CSIQ dataset. Compared with other methods, it significantly outperforms unsupervised methods on the CSIQ dataset with a marginal performance drop (-0.002) on the LIVE dataset. In conclusion, our proposed method demonstrates its potential to tackle the challenges in real-world production processes.

Introduction

Blind Image Quality Assessment (BIQA) aims to mimic human subjective systems of image quality. It is widely used as evaluation metrics (Wang et al. 2004) and loss functions (Huynh-Thu and Ghanbari 2008) for computer vision tasks. However, there is a great demand for labeled data. Due to the different subjective evaluation systems among observers and the different settings of the environment in each experiment. When the insufficient data problem arises, it is hard to make up for the problem in practice. To solve this problem, some researchers consider the BIQA task as an unsupervised problem (Mittal, Soundararajan, and Bovik 2012; Wu, Wang, and Li 2015; Venkatanath et al. 2015; Wu et al. 2020; Yang et al. 2021).

For example, Wu *et al.* (Wu, Wang, and Li 2015) proposed a novel unsupervised method by selecting statistical features named Local Pattern Statistics Index extracted from binary patterns of local image structures to evaluate image quality. Liu *et al.* (Liu et al. 2019) quantified the image quality degradation by measuring the structure, naturalness, and

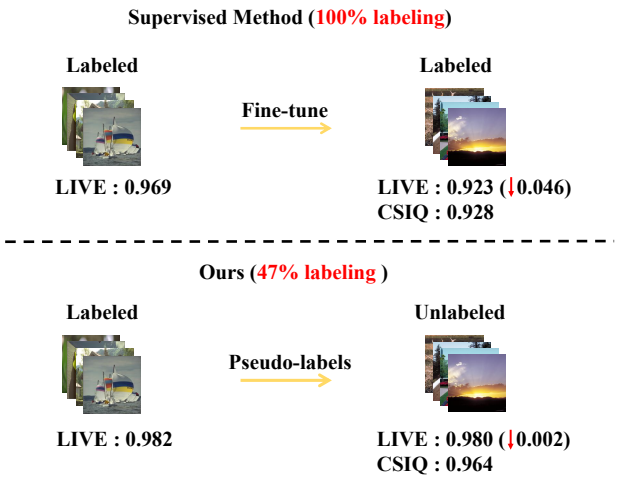

Figure 1: Comparison between the supervised method (HyperIQA) and the proposed method on the labeled data volume and performance in the LIVE $\rightarrow$ CSIQ task. "47% labeling" means that labeled data accounts for 47% of all training data. HyperIQA pre-trains on D^A and then fine-tunes on D^B resulting in a performance drop on D^A and poor performance on D^B. Our method involves pre-training on D^A and fine-tuning directly on ***UNLABELED*** D^B. As a result, the performance on D^A has been nearly maintained and surpassed HyperIQA's performance on D^B.

*These authors contributed equally.
†Corresponding Author.

perception quality variations of the distorted image from the pristine natural images. However, the performance of unsupervised methods is far from satisfactory.

With the development of deep learning, many researchers finetuned the pre-trained models to solve the insufficient data problem. CNN-based BIQA methods (Kang et al. 2014; Liu et al. 2022) directly used or fine-tuned a pre-trained CNN classification model as a feature extractor to further predict image quality scores. Ke *et al.* (Ke et al. 2021) used Vision Transformer as the backbone for feature extraction. In addition, Golestaneh *et al.* (Golestaneh, Dadsetan, and Kitani 2022) presented a hybrid architecture that combined CNN with Transformer to achieve better image representation. However, this pipeline not only has a high requirement for correlation between upstream tasks and BIQA but also fails to solve the insufficient data problem.

In this paper, we introduce a unified framework that combines semi-supervised and incremental learning to address the insufficient data problem. Specifically, the unlabeled data is uniformly partitioned into multiple subsets to simulate incremental learning. We treat each training phase as a semi-supervised learning problem, which is solved by the knowledge distillation algorithm. Meanwhile, we treat the performance degradation after each training as a catastrophic forgetting problem and leverage Experience Replay to handle it. More specifically, we use two datasets to simulate the process. Dataset D^A contains all the labeled data, while dataset D^B includes all the unlabeled data. Our task is to transfer the knowledge of dataset D^A to dataset D^B, namely $D^A \rightarrow D^B$. We first train a teacher model on the dataset D^A and then utilize the teacher model to generate pseudo-labels to dataset D^B, which are employed to overlay the manifold of dataset D^A onto dataset D^B to ensure that the model performs well on both the dataset D^A and D^B simultaneously. Meanwhile, the replay-based algorithm is leveraged during each training phase to prevent catastrophic forgetting by revisiting representative samples from previous phases. Fig.1 is the comparison between the supervised method (HyperIQA (Su et al. 2020)) and the proposed method on the labeled data volume and performance in the LIVE $\rightarrow$ CSIQ task, demonstrating that our method achieves better performance.

Our contributions are the following:

- To solve the insufficient data problem in BIQA, we introduce a unified framework that combines semi-supervised and incremental learning. In other words, the semi-supervised learning problem arises for each insufficient data problem, while the catastrophic forgetting problem arises after each incremental phase.
- To solve the semi-supervised learning problem, we propose a novel Kernel Ridge Regression (KRR) based knowledge distillation algorithm, which is leveraged to assign pseudo-labels to the unlabeled data. This allows us to transfer the manifold from dataset D^A to dataset D^B to ensure the performance of the proposed model on dataset D^B.
- To solve the catastrophic forgetting problem, we propose a novel replay-based approach, which is utilized to prevent degradation of the performance. By revisiting the representative examples of the previous phase, the analysis ability of the model is maintained.

Related Work

Blind Image Quality Assessment

The conventional BIQA researches can be further divided into distortion-specific (Wang, Sheikh, and Bovik 2002; Liu, Tanaka, and Okutomi 2013; Hu et al. 2021), Natural Scene Statistics based (Moorthy and Bovik 2011; Saad, Bovik, and Charrier 2012; Zhang, Zhang, and Bovik 2015), and Human Vision System (HVS) based metrics (Zhai et al. 2011; Gu et al. 2014). However, conventional BIQA methods restrict their ability to comprehensively represent image quality in complex real-world scenarios, especially for real-world images with diverse distortions and image contents. Recent years have witnessed the growing popularity of deep learning-based BIQA methods due to their ability to capture intricate image perceptual features. These approaches (Bosse et al. 2017; Liu, van de Weijer, and Bagdanov 2017; Ma et al. 2017; Zhang et al. 2018; Ying et al. 2020a; Zhu et al. 2020; Pan et al. 2022; Zhou et al. 2023) use Convolutional Neural Network as feature extractor, transforming the extracted features into quality scores. While CNN-based IQA models tackle complex distortion conditions and diverse image contents, they have some limitations, and the quality-aware features may remain highly abstract due to the scarcity of annotated BIQA images.

Recently, Vision Transformer (ViT) (Dosovitskiy et al. 2021) has shown impressive performance in various vision-related applications. ViT-based BIQA models can be categorized into hybrid Transformer (You and Korhonen 2021; Golestaneh, Dadsetan, and Kitani 2022) and pure ViT-based Transformer (Ke et al. 2021). Hybrid architectures combine CNNs with Transformer, while pure ViT-based methods focus solely on Transformer.

Semi-Supervised Blind Image Quality Assessment

While semi-supervised learning has been extensively studied, its application to IQA remains relatively scarce. SSL Ensemble (Wang, Li, and Ma 2021) utilized an ensemble learning approach to improve the diversity of model predictions for unlabeled data, thus enhancing the model's generalization performance. Differently, Prabhakaran et al. (Prabhakaran and Swamy 2023) presented a framework utilizing contrastive learning to develop feature representations, effectively pretraining an image encoder to cluster images based on their quality through synthetic distortions. By augmenting contrastive learning with downstream supervision, the study achieved more transferable representations suitable for IQA. In contrast to their approach, we employ knowledge distillation with kernel ridge regression (Welling 2013) to obtain pseudo-labels from the teacher model for unlabeled data.

Incremental Learning for Blind Image Quality Assessment

Incremental or continual learning research in the field of BIQA is still in its nascent stage. LwF-KG (Liu et al. 2022) pioneered the concept of continual learning in BIQA and introduced a simple yet effective approach. By building upon a shared backbone network, they appended a prediction head for a new dataset and imposed a regularizer, enabling all prediction heads to evolve with new data while mitigating catastrophic forgetting of old data. Ultimately, an aggregate quality score was computed by a weighted summation of predictions from all heads. R&R-Net (Ma et al. 2021) introduced a dynamic Remember and Reuse (R&R) network for efficient cross-task blind image quality assessment (BIQA) using a novel relevance-aware incremental learning strategy. The R&R network sequentially updates parameters for multiple evaluation tasks, preserving task-specific preferences while pruning and reusing parameters dynamically based on task relevance, achieving improved prediction accuracy. The distinction lies in the fact that prior work falls under supervised learning, while our approach involves incremental learning within a semi-supervised setting.

Proposed Method

Overview

In this article, we propose a novel Semi-Supervised BIQA framework (SS-IQA), based on knowledge distillation and incremental learning (Fig. 2). Our approach centers on the transfer of knowledge garnered from labeled data to those that lack labeling, with the aim of amplifying performance outcomes in unlabeled data while minimizing the performance drop of labeled data as much as possible.

Notations. In this study, we establish a set of notations to increase clarity and consistency. We denote the labeled dataset as $D^A = \{(x_i, y_i)\}_{i=1}^{M}$, and the unlabeled dataset as $D^B = \{(x_j)\}_{j=1}^{P}$. Therefore, we can define our task $D^A \rightarrow D^B$, which leverages the knowledge from D^A to improve the performance of D^B in a semi-supervised manner to ensure excellent performance on both D^A and D^B. We also introduce $\mathcal{E}^A$ and $\mathcal{E}^B$, which are representative examples derived from D^A and D^B, respectively.

Our training process. The teacher model is trained on weakly augmented images, and the student model is trained on strongly augmented images. Pseudo-labels generated by the teacher model supervise the student network on unlabeled datasets. Unlabeled datasets are divided into blocks, and the student model iteratively updates itself by training on each block of data in sequence. A sampling module selects representative samples for revising old knowledge to avoid catastrophic forgetting. This design enhances assessment quality, reduces data annotation costs, and improves generalization capability for addressing BIQA challenges.

Kernel Ridge Regression (KRR) - Distillation

We propose a distillation module to enable semi-supervised learning for BIQA and address data annotation challenges. We start by training a high-performance teacher model on a labeled dataset and then use it to generate pseudo-labels for unlabeled data to supervise the student model training. Initially, the student and teacher models are identical. We provide weak and strong augmentations to the teacher and student inputs, respectively, allowing the student to learn more effectively. To preserve image quality during augmentation, we use random cropping for the teacher and random horizontal flipping and cropping for the student. After each incremental phase, we replace the teacher with the student. This process iterates to improve the student's performance.

To address potential bias in pseudo-labels, we use Kernel Ridge Regression (KRR) to improve their reliability. KRR constructs a high-dimensional feature space and computes the similarity between data points using a kernel function. For unlabeled data, pseudo-labels are obtained by weighting labeled data points in the high-dimensional space based on kernel function values. This reduces bias and improves pseudo-label quality, leading to better model performance.

Specifically, we extract features from $D^{A'}$ and D^B using the teacher model to obtain feature matrices F^A and F^B. The set $D^{A'}$ is a subset of D^A, obtained through SDK-Sample. We use F^A and ground truths to fit a non-linear model for $D^{A'}$, minimizing the KRR loss as follows:

$$\mathcal{L} = \sum_{i \in D^{A'}} \left(y_i - \sum_{z \in D^{A'}} \alpha_z K\left(f_i, f_z\right) \right)^2 + \lambda \sum_{i,z \in D^{A'}} \alpha_i \alpha_z K\left(f_i, f_z\right), \tag{1}$$

where f_i and f_z denote the feature of the i-th and z-th sample, α_i and α_z are kernel ridge regression coefficients, describing the contribution of each feature point. λ balance model's fitting degree and model complexity, K is the Radial Basis Function (RBF) (Buhmann 2000). The definition of RBF can be written as:

$$K\left(f_i, f_z\right) = \exp\left(-\gamma \left\| f_i - f_z \right\|^2\right), \gamma > 0, \tag{2}$$

where γ defines the influence range of a single sample. This is known as the Gaussian Kernel (Keerthi and Lin 2003). Then, we utilize feature matrix F^B as input and obtain pseudo-labels for the unlabeled data D^B.

$$y_j = \sum_{i \in D^{A'}} \alpha_i K\left(f_i, f_j\right), \tag{3}$$

where y_j is the pseudo label of the j-th sample in D^B.

Incremental Learning

Our system consists of $N + 1$ phases, including one initial phase and N incremental phases. The unlabeled dataset D^B is uniformly partitioned into N subsets D_0^B, D_1^B, $\cdots$, D_{N-1}^B. In the initial phase, we train a teacher network on the labeled data D^A using the smooth $\mathcal{L}_1$ loss and save the resulting model Θ_0 and representative samples $\mathcal{E}^A$ to the system's memory. For each incremental phase (i-th phase), we retrieve Θ_{i-1} and representative exemplars $\mathcal{E}^A$ and $\mathcal{E}_{0:i-1}^B$

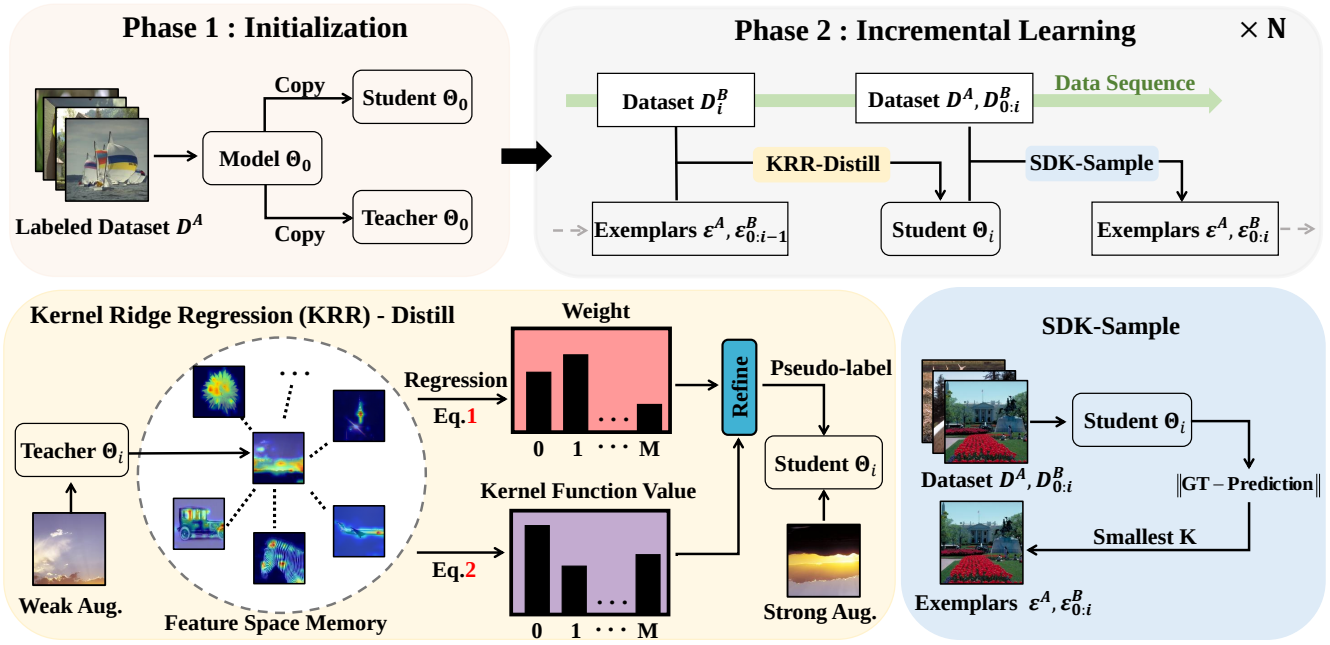

Figure 2: The overview of SS-IQA. The model comprises two phases. In the first phase, D^A is utilized to initiate both student and teacher. In the second phase, N incremental learning phases are carried out on D^B. The KRR-Distill module leverages the Kernel Ridge Regression to produce pseudo-labels. To prevent catastrophic forgetting during incremental learning, the SDK-Sample module selects the K most precise exemplars, serving as representative exemplars for replay.

from memory and train the student network Θ_i on representative exemplars and the new data D_i^B. Finally, we evaluate the performance of the student model Θ_N on the validation sets of both datasets D^A and D^B.

SDK - Sample

To prevent catastrophic forgetting and improve generalization performance, we propose a replay-based algorithm with the Smallest Difference topK (SDK)-Sample mechanism for representative sample selection. In the i-th incremental phase, we choose representative examples from the previous i-1 phases. This is achieved by validating the model Θ_{i-1} on datasets D^A and $D_{0:i-1}^B$, selecting the top K nearest examples to the corresponding ground truth and pseudo label, respectively, denoted as $\mathcal{E}^A$, $\mathcal{E}_{0:i-1}^B$. We then mix these representative samples with the incremental data D_i^B to train the model Θ_i, initialized by Θ_{i-1}. This approach ensures that the model does not excessively lean toward new knowledge and forget the learned knowledge. By avoiding catastrophic forgetting and enhancing the model's robustness and generalization, our method achieves improved performance in incremental learning.

Experiments

Datasets and Evaluation Protocols

We evaluate the performance of our proposed model on eight typical image quality evaluation datasets, including four synthetic datasets and four authentic datasets. The synthetic datasets we used are LIVE (Sheikh, Sabir, and Bovik 2006), CSIQ (Larson and Chandler 2010), TID2013 (Ponomarenko et al. 2015), and KADID (Lin, Hosu, and Saupe 2019). The authentic datasets we used are LIVEC (Ghadiyaram and Bovik 2015), KonIQ (Hosu et al. 2020), LIVEFB (Ying et al. 2020b), and SPAQ (Fang et al. 2020).

To evaluate the performance of our model, we use Pearson's Linear Correlation Coefficient (PLCC) and Spearman's Rank order Correlation Coefficient (SRCC) as evaluation metrics. Specifically, PLCC and SRCC are intended for assessing the accuracy of BIQA model predictions and the monotonicity of BIQA algorithm predictions, respectively. Both PLCC and SRCC range from 0 to 1, with higher values indicating better performance.

Implementation Details

We use the transformer structure as our model. Our transformer encoder is based on the ViT-S proposed in DeiT III (Touvron, Cord, and Jégou 2022) and pre-trained for 400 epochs on ImageNet-1K. Our baseline model does not employ KRR-based knowledge distillation or incremental learning. During training, we use the Adam optimizer and use the smooth $\mathcal{L}_1$ loss as the loss function. In the initial phase, we train for 9 epochs on the dataset D^A, with 3 warmup epochs and a learning rate of 2×10^{-4}, which decreases by 0.1 every 3 epochs. In the incremental phase, we train for 6 epochs on blocks of the dataset D^B, with an initial learning rate of 2×10^{-5}, which decreases by 0.1 every 2 epochs. The dataset D^B is divided into 3 blocks. For the datasets D^A and D^B, we select one representative sample

Method \ Dataset D^A	LIVE		CSIQ		TID2013		KADID		LIVEC		KonIQ		LIVEFB		SPAQ	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
ILNIQE	0.906	0.902	0.865	0.822	0.648	0.521	0.558	0.534	0.508	0.508	0.537	0.523	0.332	0.294	0.712	0.713
MEON	0.955	0.951	0.864	0.852	0.824	0.808	0.691	0.604	0.710	0.697	0.628	0.611	0.394	0.365	-	-
WaDIQaM	0.955	0.960	0.844	0.852	0.855	0.835	0.752	0.739	0.671	0.682	0.807	0.804	0.467	0.455	-	-
DBCNN	0.971	0.968	0.959	0.946	0.865	0.816	0.856	0.851	0.869	0.851	0.884	0.875	0.551	0.545	0.915	0.911
TIQA	0.965	0.949	0.838	0.825	0.858	0.846	0.855	0.850	0.861	0.845	0.903	0.892	0.581	0.541	-	-
MetaIQA	0.959	0.960	0.908	0.899	0.868	0.856	0.775	0.762	0.835	0.802	0.887	0.850	0.507	0.540	-	-
P2P-BM	0.958	0.959	0.902	0.899	0.856	0.862	0.849	0.840	0.842	0.844	0.885	0.872	0.598	0.526	-	-
HyperIQA	0.966	0.962	0.942	0.923	0.858	0.840	0.845	0.852	0.882	0.859	0.917	0.906	0.602	0.544	0.915	0.911
TReS	0.968	0.969	0.942	0.922	0.883	0.863	0.858	0.859	0.877	0.846	0.928	0.915	0.625	0.554	-	-
MUSIQ	0.911	0.940	0.893	0.871	0.815	0.773	0.872	0.875	0.746	0.702	0.928	**0.916**	**0.661**	**0.566**	**0.921**	**0.918**
DACNN	0.980	0.978	0.957	0.943	0.889	0.871	**0.905**	**0.905**	**0.884**	**0.866**	0.912	0.901	-	-	**0.921**	0.915
SS-IQA(ours)	**0.980**	**0.978**	**0.969**	**0.960**	**0.910**	**0.891**	0.895	0.896	0.869	0.835	**0.932**	0.913	0.582	0.542	0.824	0.826
Method \ Dataset D^B	CSIQ*		LIVE		KADID*		TID2013*		KonIQ		LIVEC		SPAQ		LIVEFB	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
LPSI	0.866	0.771	0.828	0.818	0.767	0.527	0.811	0.705	0.206	0.039	0.299	0.083	0.146	0.000	-	-
PIQUE	0.862	0.837	0.836	0.840	0.789	0.762	0.841	0.825	0.300	0.145	0.171	0.108	0.249	0.156	-	-
NIQE	0.877	0.871	0.906	0.908	0.830	0.832	0.808	0.797	0.398	0.371	0.495	0.450	0.503	0.501	-	-
ILNIQE	0.814	0.879	0.902	0.897	0.869	0.855	0.860	0.842	0.563	0.536	0.504	0.439	**0.707**	**0.696**	-	-
Q-NIQE	0.913	0.905	0.908	0.911	-	-	0.858	0.859	-	-	0.562	0.520	-	-	-	-
CCF	0.410	0.263	0.442	0.387	0.804	0.285	0.679	0.693	0.552	0.552	0.047	0.110	0.589	0.207	**0.629**	0.183
UCIQE	0.148	0.275	0.431	0.561	0.447	0.322	0.643	0.654	0.673	0.576	0.200	0.136	0.573	0.192	0.500	0.176
FDUM	0.375	0.220	0.847	0.785	0.579	0.220	0.596	0.518	0.653	0.527	0.531	0.556	0.581	0.496	0.113	0.104
SS-IQA(ours)	**0.964**	**0.949**	**0.948**	**0.952**	**0.892**	**0.866**	**0.929**	**0.940**	**0.741**	**0.724**	**0.806**	**0.780**	0.420	0.426	0.429	**0.386**

Table 1: Performance comparison is measured by averages of SRCC and PLCC, where bold entries indicate the best results. *: For the CSIQ, TID2013, and KADID datasets, previous unsupervised methods only used partial data on common distortion types such as JPEG, JPEG2000, white noise, and Gaussian blur. For a fair comparison, we follow them.

for every 100 samples, respectively.

We conduct our experiments on 8 different settings: LIVE $\rightarrow$ CSIQ, CSIQ $\rightarrow$ LIVE, TID2013 $\rightarrow$ KADID, KADID $\rightarrow$ TID2013, LIVEC $\rightarrow$ KonIQ, KonIQ $\rightarrow$ LIVEC, SPAQ $\rightarrow$ LIVEFB, and LIVEFB $\rightarrow$ SPAQ. The batch sizes are set differently for each dataset. For all datasets, we use PLCC and SRCC as the evaluation metric. For each dataset, we use 80% of the data for training and 20% for testing. We repeat each experiment 10 times and calculate the average PLCC and SRCC to mitigate the performance bias.

Overall Prediction Performance Comparison

We compare our approach to state-of-the-art supervised methods on dataset D^A, including both hand-crafted and deep-learning-based BIQA methods, as well as current state-of-the-art unsupervised methods on dataset D^B, including some underwater unsupervised methods, CCF (Wang et al. 2018) and UCIQE (Yang and Sowmya 2015).

As shown in Table 1, our model performs comparably to state-of-the-art methods on synthetic datasets, outperforming the current SOTA on the CSIQ $\rightarrow$ LIVE task. It significantly surpasses the performance of unsupervised state-of-the-art methods on D^B. However, on authentic datasets like LIVEC $\rightarrow$ KonIQ and KonIQ $\rightarrow$ LIVEC, our performance slightly decreases on D^A, but our model still outperforms unsupervised state-of-the-art methods on D^B. Our performance is comparatively poor on the LIVEFB $\rightarrow$ SPAQ and SPAQ $\rightarrow$ LIVEFB tasks. This could be due to specific image characteristics or distortions in these tasks that might have been less represented or rare in the labeled data, making it challenging for the model to generalize well in these scenarios. Overall, achieving leading performance across datasets with diverse image content and distortion types is challenging. Nevertheless, these observations effectively confirm the effectiveness and superiority of SS-IQA and demonstrate its potential to tackle the challenges in real-world production processes.

Method	LIVEC		KonIQ	
	PLCC	SRCC	PLCC	SRCC
SSLIQA	0.706	0.695	0.867	0.841
ours	**0.776**	**0.745**	**0.891**	**0.881**
	KonIQ		KADID	
	PLCC	SRCC	PLCC	SRCC
SSL	0.900	0.890	**0.910**	**0.900**
ours	**0.932**	**0.913**	0.895	0.896

Table 2: Comparison with semi-supervised methods, SSLIQA and SSL. Due to the significant differences in experimental setups between SSLIQA and our SS-IQA, we followed the experimental settings outlined in the original paper to conduct our experiments.

Method	LIVE->CSIQ			
	LIVE		CSIQ	
	PLCC	SRCC	PLCC	SRCC
Baseline	**0.980**	**0.978**	0.957	0.940
std	±0.005	±0.005	±0.030	±0.032
+ IL	0.979	**0.978**	0.963	0.948
std	±0.007	±0.005	±0.015	±0.015
+ KRR	0.979	0.977	0.961	0.947
std	±0.007	±0.009	±0.012	±0.019
SS-IQA(ours)	**0.980**	**0.978**	**0.964**	**0.949**
std	±0.005	±0.005	±0.023	±0.022

Table 3: Ablation experiments on LIVE and CSIQ datasets. Bold entries indicate the best performance.

Method	KADID->TID2013			
	KADID		TID2013	
	PLCC	SRCC	PLCC	SRCC
Baseline	0.891	0.894	0.918	0.927
std	±0.026	±0.022	±0.048	±0.047
+ IL	0.895	**0.896**	0.927	0.938
std	±0.017	±0.015	±0.060	±0.054
+ KRR	**0.896**	**0.896**	0.921	0.928
std	±0.018	±0.020	±0.037	±0.042
SS-IQA(ours)	0.895	**0.896**	**0.929**	**0.940**
std	±0.022	±0.019	±0.036	±0.030

Table 4: Ablation experiments on KADID and TID2013 datasets. Bold entries indicate the best performance.

	CSIQ->LIVE				KADID->TID2013			
	CSIQ		LIVE		KADID		TID2013	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
Q=1	0.968	0.958	0.938	0.946	**0.896**	**0.897**	0.925	0.936
Q=2	0.965	0.954	0.943	0.949	0.895	0.896	0.928	0.935
Q=3	**0.969**	**0.960**	0.948	0.952	0.895	0.896	0.929	0.940
Q=4	0.965	0.959	0.937	0.946	**0.896**	**0.897**	0.929	0.938
Q=6	**0.969**	**0.960**	**0.952**	**0.958**	0.895	0.896	**0.931**	**0.942**

Table 5: Analysis of Block Quantity in Incremental Dataset D^B. Bold entries indicate the best performance.

LIVE->CSIQ		LIVE		CSIQ	
		PLCC	SRCC	PLCC	SRCC
K_1=20	K_2=0	0.978	0.976	0.961	0.945
	K_2=10	0.979	0.977	0.965	**0.954**
	K_2=20	0.979	0.977	0.962	0.949
	K_2=50	0.977	0.976	0.964	0.950
	K_2=100	0.978	0.976	0.962	0.947
K_2=20	K_1=0	0.977	0.975	0.964	0.951
	K_1=10	0.979	0.977	0.962	0.949
	K_1=50	0.979	0.977	0.963	0.949
	K_1=100	0.979	0.977	**0.967**	0.952
K_1=100	K_2=100	**0.980**	**0.978**	0.964	0.949

Table 6: Analysis of K_1 and K_2. Bold entries indicate the best performance.

Performance Comparison with Semi-Supervised BIQA

We compared our method with SSLIQA (Yue et al. 2022) and SSL (Prabhakaran and Swamy 2023) on three datasets: LIVEC, KADID, and KonIQ. Due to the significant differences in experimental setups between SSLIQA and our SS-IQA, we followed the experimental settings outlined in the original paper to conduct our experiments. Both SSL and we have utilized unlabeled datasets, a direct comparison with them is appropriate. As shown in Table 2, our approach outperformed SSLIQA and demonstrated strengths distinct from SSL. Specifically, our method exhibited superior performance across multiple evaluation metrics, showcasing its efficacy in the context of semi-supervised image quality assessment. It's plausible that our method employs a more effective semi-supervised learning strategy and captures more discriminative features related to image quality.

Ablation Study

The Impact of Each Component. SS-IQA is a novel model that integrates knowledge distillation and incremental learning, consisting of two essential components: pseudo-label generation using Kernel Ridge Regression (KRR) based knowledge distillation and Incremental Learning with Experience Replay. Each component plays a crucial role in accurately characterizing image quality and improving the overall performance of the model. To better understand the importance of each component, we conduct ablation experiments on the LIVE → CSIQ and KADID → TID2013. The results, as shown in Table 3 and 4, indicate that all components of our proposed method made significant contributions to image quality characterization. Our proposed sampling method provides significant improvements in both accuracy and stability, especially when KRR is used for teacher distillation. This demonstrates the effectiveness of the distillation method, which allows our student model to improve its ability to distinguish quality-related features.

Analysis of Block Quantity in Incremental Dataset D_B. In this analysis, we examine the impact of the number of unlabeled datasets D^B blocks. We conduct experiments on CSIQ → LIVE and KADID → TID2013, with Q values of 1, 2, 3, 4, and 6, respectively. Q represents the number of blocks in the unlabeled dataset. As shown in Table 5, the results show that increasing the value of Q indicates a slight improvement in the model's performance on unlabeled datasets, LIVE and TID2013. Considering accuracy and training duration, we set Q=3.

Analysis of Representative Sample Size. During each incremental learning phase, we denote the frequency of selecting samples from datasets D^A and D^B as K_1 and K_2, respectively. The number of samples selected from D^A and D^B are $\frac{|D^A|}{K_1}$ and $\frac{|D^B|}{K_2}$, respectively, where $|D^A|$ and $|D^B|$ are the sizes of corresponding datasets, respectively. The re-

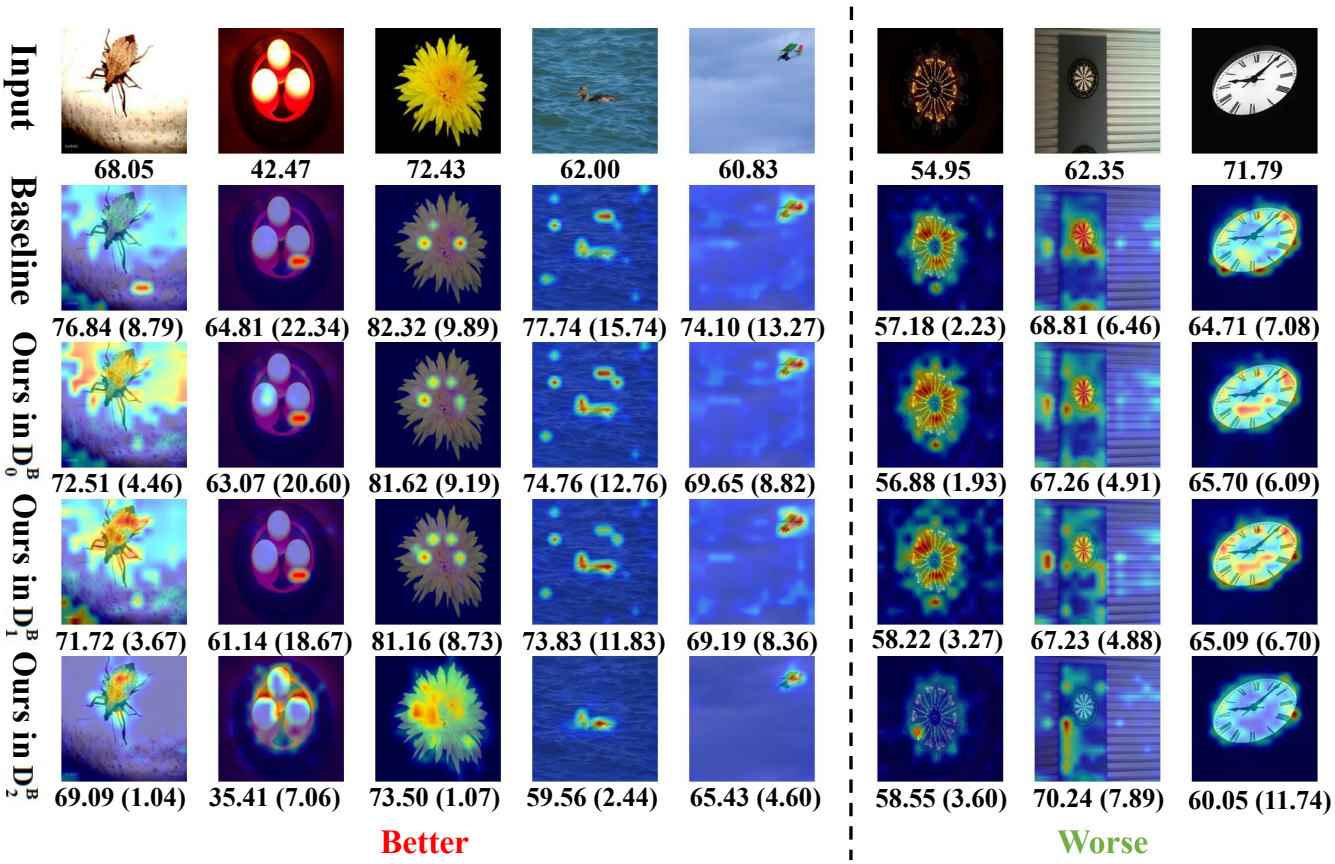

Figure 3: Comparison of activation maps between our baseline model and SS-IQA using Grad-CAM (Selvaraju et al. 2017). Rows 1-5 show input images, CAMs from the baseline, and CAMs from three training phases of SS-IQA on dataset D^B. The first line of numbers represents the ground truth of the input image. The numbers below each image in lines 2-5 represent the predicted scores, and the number in parentheses represents the distance between the predicted score and the ground truth.

sults in Table 6 demonstrate that changing the values of K_1 and K_2 does not significantly impact the performance, with differences not exceeding 1%. To minimize memory consumption, we set K_1 and K_2 to 100 in our experiments, and this setting remains consistent throughout our study.

Visualization of Class Activation Map

We utilize GradCAM (Selvaraju et al. 2017) to visualize the feature attention maps of the input images in both our baseline model and SS-IQA, as shown in Fig. 3. The number below each figure represents the predicted quality score of the model, while the number in parentheses shows the distance between the predicted score and the ground truth in the first row. D_0^B, D_1^B, and D_2^B illustrate the different phases during training on dataset D^B. The left results demonstrate that in most cases, our incremental learning approach gradually leads to better-predicted results within each iteration, as compared to the previous phases. Meanwhile, our model gradually narrows its focus on the salient area during the iteration process. After the final phase, SS-IQA accurately and comprehensively focuses on the salient area, while the baseline model loses its attention. However, our method does not always evolve toward better performance. When the unlabeled data is not covered by the original data space, our model may lose focus and result in poor performance. In summary, our model transfers knowledge from labeled data to unlabeled data and effectively prevents catastrophic forgetting during the incremental learning process.

Conclusion

Blind Image Quality Assessment (BIQA) has a great demand for labeled data, which is often insufficient in practice. In this paper, we propose a unified framework for BIQA that combines knowledge distillation and incremental learning to solve the insufficient data problem. Knowledge distillation is employed to generate pseudo-labels for unlabeled data, resulting in expanding datasets and implementing semi-supervised learning. Experience Replay is applied to prevent catastrophic forgetting during multiple semi-supervised learning. Experimental results demonstrate the effectiveness of our proposed approach across multiple IQA datasets. In conclusion, our proposed method demonstrates its potential to tackle the challenges in real-world production processes.

Acknowledgements

This work was supported by National Key R&D Program of China (No.2022ZD0118202), the National Science Fund for Distinguished Young Scholars (No.62025603), the National Natural Science Foundation of China (No. U21B2037, No. U22B2051, No. 62176222, No. 62176223, No. 62176226, No. 62072386, No. 62072387, No. 62072389, No. 62002305 and No. 62272401), and the Natural Science Foundation of Fujian Province of China (No.2021J01002, No.2022J06001).

References

Bosse, S.; Maniry, D.; Müller, K.-R.; Wiegand, T.; and Samek, W. 2017. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on image processing*, 27(1): 206–219.

Buhmann, M. D. 2000. Radial basis functions. *Acta numerica*, 9: 1–38.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

Fang, Y.; Zhu, H.; Zeng, Y.; Ma, K.; and Wang, Z. 2020. Perceptual quality assessment of smartphone photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3677–3686.

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

Golestaneh, S. A.; Dadsetan, S.; and Kitani, K. M. 2022. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1220–1230.

Gu, K.; Zhai, G.; Yang, X.; and Zhang, W. 2014. Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*, 17(1): 50–63.

Hosu, V.; Lin, H.; Sziranyi, T.; and Saupe, D. 2020. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29: 4041–4056.

Hu, R.; Yang, R.; Liu, Y.; and Li, X. 2021. Simulation and mitigation of the wrap-around artifact in the MRI image. *Frontiers in Computational Neuroscience*, 15: 746549.

Huynh-Thu, Q.; and Ghanbari, M. 2008. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13): 800–801.

Kang, L.; Ye, P.; Li, Y.; and Doermann, D. 2014. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1733–1740.

Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; and Yang, F. 2021. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5148–5157.

Keerthi, S. S.; and Lin, C.-J. 2003. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural computation*, 15(7): 1667–1689.

Larson, E. C.; and Chandler, D. M. 2010. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1): 011006–011006.

Lin, H.; Hosu, V.; and Saupe, D. 2019. KADID-10k: A large-scale artificially distorted IQA database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 1–3. IEEE.

Liu, J.; Zhou, W.; Li, X.; Xu, J.; and Chen, Z. 2022. LIQA: Lifelong blind image quality assessment. *IEEE Transactions on Multimedia*.

Liu, X.; Tanaka, M.; and Okutomi, M. 2013. Single-image noise level estimation for blind denoising. *IEEE Trans. Image Process.*, 22(12): 5226–5237.

Liu, X.; van de Weijer, J.; and Bagdanov, A. D. 2017. RankIQA: Learning From Rankings for No-Reference Image Quality Assessment. In *The IEEE International Conference on Computer Vision (ICCV)*.

Liu, Y.; Gu, K.; Zhang, Y.; Li, X.; Zhai, G.; Zhao, D.; and Gao, W. 2019. Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4): 929–943.

Ma, K.; Liu, W.; Zhang, K.; Duanmu, Z.; Wang, Z.; and Zuo, W. 2017. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3): 1202–1213.

Ma, R.; Luo, H.; Wu, Q.; Ngan, K. N.; Li, H.; Meng, F.; and Xu, L. 2021. Remember and reuse: Cross-task blind image quality assessment via relevance-aware incremental learning. In *Proceedings of the 29th ACM International Conference on Multimedia*, 5248–5256.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.

Moorthy, A. K.; and Bovik, A. C. 2011. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. Image Process.*, 20(12): 3350–3364.

Pan, Z.; Zhang, H.; Lei, J.; Fang, Y.; Shao, X.; Ling, N.; and Kwong, S. 2022. Dacnn: Blind image quality assessment via a distortion-aware convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11): 7518–7531.

Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30: 57–77.

Prabhakaran, V.; and Swamy, G. 2023. Image Quality Assessment using Semi-Supervised Representation Learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 538–547.

Saad, M. A.; Bovik, A. C.; and Charrier, C. 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE transactions on Image Processing*, 21(8): 3339–3352.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.

Sheikh, H. R.; Sabir, M. F.; and Bovik, A. C. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11): 3440–3451.

Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; and Zhang, Y. 2020. 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*, 3667–3676.

Touvron, H.; Cord, M.; and Jégou, H. 2022. Deit iii: Revenge of the vit. *arXiv preprint arXiv:2204.07118*.

Venkatanath, N.; Praneeth, D.; Bh, M. C.; Channappayya, S. S.; and Medasani, S. S. 2015. Blind image quality evaluation using perception based features. In *2015 twenty first national conference on communications (NCC)*, 1–6. IEEE.

Wang, Y.; Li, N.; Li, Z.; Gu, Z.; Zheng, H.; Zheng, B.; and Sun, M. 2018. An imaging-inspired no-reference underwater color image quality assessment metric. *Computers & Electrical Engineering*, 70: 904–913.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wang, Z.; Li, D.; and Ma, K. 2021. Semi-supervised deep ensembles for blind image quality assessment. *arXiv preprint arXiv:2106.14008*.

Wang, Z.; Sheikh, H. R.; and Bovik, A. C. 2002. No-reference perceptual quality assessment of JPEG compressed images. In *Proceedings. International conference on image processing*, volume 1, I–I. IEEE.

Welling, M. 2013. Kernel ridge regression. *Max Welling's classnotes in machine learning*, 1–3.

Wu, L.; Zhang, X.; Chen, H.; and Zhou, Y. 2020. Unsupervised quaternion model for blind colour image quality assessment. *Signal Processing*, 176: 107708.

Wu, Q.; Wang, Z.; and Li, H. 2015. A highly efficient method for blind image quality assessment. In *2015 IEEE International Conference on Image Processing (ICIP)*, 339–343. IEEE.

Yang, M.; and Sowmya, A. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12): 6062–6071.

Yang, N.; Zhong, Q.; Li, K.; Cong, R.; Zhao, Y.; and Kwong, S. 2021. A reference-free underwater image quality assessment metric in frequency domain. *Signal Processing: Image Communication*, 94: 116218.

Ying, Z.; Niu, H.; Gupta, P.; Mahajan, D.; Ghadiyaram, D.; and Bovik, A. 2020a. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3575–3585.

Ying, Z.; Niu, H.; Gupta, P.; Mahajan, D.; Ghadiyaram, D.; and Bovik, A. 2020b. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3575–3585.

You, J.; and Korhonen, J. 2021. Transformer for image quality assessment. In *2021 IEEE International Conference on Image Processing (ICIP)*, 1389–1393. IEEE.

Yue, G.; Cheng, D.; Li, L.; Zhou, T.; Liu, H.; and Wang, T. 2022. Semi-supervised authentically distorted image quality assessment with consistency-preserving dual-branch convolutional neural network. *IEEE Transactions on Multimedia*.

Zhai, G.; Wu, X.; Yang, X.; Lin, W.; and Zhang, W. 2011. A psychovisual quality metric in free-energy principle. *IEEE Trans. Image Process.*, 21(1): 41–52.

Zhang, L.; Zhang, L.; and Bovik, A. C. 2015. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8): 2579–2591.

Zhang, W.; Ma, K.; Yan, J.; Deng, D.; and Wang, Z. 2018. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1): 36–47.

Zhou, H.; Yang, R.; Hu, R.; Shu, C.; Tang, X.; and Li, X. 2023. ETDNet: Efficient Transformer-based Detection Network for Surface Defect Detection. *IEEE Transactions on Instrumentation and Measurement*.

Zhu, H.; Li, L.; Wu, J.; Dong, W.; and Shi, G. 2020. MetaIQA: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14143–14152.