Domain-Aware No-Reference Image Quality Assessment
Wei Hao Xia, Yuju Yang, Jing-Hao Xue, Jing Xiao

Abstract—No-reference image quality assessment (NR-IQA) is a fundamental yet challenging task in low-level computer vision. It is to predict the perceptual quality of an image with unknown distortion. Its difficulty is particularly pronounced as the corresponding reference for assessment is typically absent. Various mechanisms to extract features ranging from natural scene statistics to deep features have been leveraged to boost the NR-IQA performance. However, these methods treat images of different degradations the same and the representations of distortions are under-exploited. Furthermore, identifying the distortion type should be an important part for NR-IQA, which is rarely addressed in the previous methods. In this work, we propose the domain-aware no-reference image quality assessment (DA-NR-IQA), which for the first time exploits and disentangles the distinct representation of different degradations to access image quality. Benefiting from the design of domain-aware architecture, our method can simultaneously identify the distortion type of an image. With both the by-product distortion type and quality score determined, the distortion in an image can be better characterized and the image quality can be more precisely assessed. Extensive experiments show that the proposed DA-NR-IQA performs better than almost all the other state-of-the-art methods.

Index Terms—No-reference image quality assessment, distortion type, disentangled representations, neural networks, generative adversarial network

I. INTRODUCTION

No-reference image quality assessment (NR-IQA) is a fundamental yet challenging task that automatically predicts the perceptual quality of a degraded image without the corresponding reference for comparison. It serves as a key component in low-level computer vision community. In practice, many tasks like image restoration [1], [2] and image generation [3] adopt full-reference SSIM and PSNR as metrics to evaluate the quality of generated images. However, it is difficult or even impossible to acquire a non-distorted image as reference in some applications. The ill-posed nature of NR-IQA is particularly pronounced for the absence of distortion form and the corresponding non-distorted reference. Although mechanisms to extract features ranging from natural scene statistics to deep features have been proposed, the performance bottleneck still exists.

Numerous efforts [4], [5] have been made to design powerful feature representation models. Traditional methods usually leverage natural scene statistics and lead to the lack of flexibility and diversity for complex diverse degradations. In recent years, deep learning methods achieve promising results in NR-IQA. Due to the extremely limited datasets, many methods utilize various data augmentation or multi-task strategies to better leverage the power of Deep Neural Networks (DNNs).

Some methods [6], [7], [8] focus on simulating the behavior of Human Visual System (HVS) by Generative Adversarial Networks (GANs). Specifically, [6] propose a discrepancy-guided quality regression network to leverage incorporated discrepancy information which encodes the difference between distorted image and hallucinated reference to make precise prediction. [7] propose the restorative adversarial net, i.e., the restorator reconstructs the input distorted patches while the discriminator tries to distinguish the restored patches from the pristine distortion-free ones, and the evaluator takes the distorted and the restored patches as inputs and predicts a perceptually quantified score.

However, these methods treat images of different distortions the same and representations of distortions are under-exploited. On the one hand, different distorted types transubstantiate distortion-free images into different distorted versions, and lead to significantly different visual perception. The discrepancy in appearance means necessity and obligation to characterize features for different distortions. On the other hand, feature extraction mechanisms adopted by the previous methods, from natural scene statistics to deep features, are all acting on the images. It leads to the results that the obtained feature may cover more other unrelated or indirect factors (e.g., color, texture) rather than distortion information.

Some previous NR-IQA methods [9] try to evaluate the image quality considering the information of the distortion type, but the results are not comparable with state-of-the-art NR-IQA methods, and most of them simply leverage the distortion information by adding a classification to identify the distortion. Recently, [10] propose an image restoration method by disentangling content and blur features from
blurred images. It would be significative if there is a way to disentangle the distortion in images with different distortions simultaneously, and directly use these disentangled distinct distortion representation for IQA instead of other indirect or unrelated features.

Furthermore, distortion type is also another significant part of NR-IQA. The quality of an image is closely related to its distortion type. As shown in Figure 1, the degradation arises from the specific distortion and the image quality depends on the distortion type and level. With both the determined distortion type and quality score, the distortion in an image can be better characterized. Unfortunately, almost all existing IQA methods only evaluate the image quality ignoring the distortion type. A recent method [11] encode distortions for IQA and is capable of evaluating the quality score and identifying distortions simultaneously. But it is a full reference framework and still needs a distort-free reference.

To address the above drawbacks, we propose the domain-aware no-reference image quality assessment (DA-NR-IQA), which exploits the distorted representation of specific degradation for image quality assessment. Domain here is defined as images with the same type and level of distortion. We exploit how to disentangle latent representations for each domain to represent particular semantic or degradation information. Furthermore, instead of extending IQA module for multi-task by adding a classification layer as in [9], [11], benefiting from the design of domain-aware network architecture, DA-NR-IQA can also identify the distortion type of an image simultaneously.

Our contributions are summarized as follows:

- We propose the domain-aware no-reference image quality assessment (DA-NR-IQA), which for the first time exploits the discriminative disentangled features of different distortions for image quality assessment.
- Our method can identify the distortion type of an image, and use the determined distortion type and quality score to characterize the image quality.
- Our method achieves superior performance on popular IQA datasets compared with state-of-the-art methods.

II. RELATED WORK

A. No-Reference Image Quality Assessment

The existing studies on NR-IQA can be broadly classified into two categories: designing better hand-crafted features [12], [13] and learning discriminant visual features automatically. The first category of methods typically use a two-stage framework, which performs distortion identification and the distortion-specific quality estimation accordingly. However, experiments in [13] has shown that such two-stage methods are not superior to the distortion-blind approaches.

The second category of work attempts to learn discriminant visual features automatically without using hand-crafted features. [14] constructs a small yet accurate codebook to look up the proper features. [4], [9], [5] adopt deep neural network to extract features from the raw input and perform regression to estimate perceptual scores.

The above NR-IQA methods can be summarized as feature extraction and regression only based on distorted images. However, according to the free-energy theory [15], HVS tends to restore the distorted image before quality assessment. Despite building NR-IQA models based on the free-energy theory, [16], [17] restore the distorted image with a linear autoregressive model, which is not capable of producing a satisfactory result when the input suffers high-level distortion and therefore may not be consistent with HVS.

[6], [7] simulate the behavior of HVS by using Generative Adversarial Networks (GANs) to generate the corresponding restored counterparts as reference. However, their methods don’t exhibit ability in disentangling and characterising discriminative latent representations for different degradations, which is one of our key contributions.

B. Representation Disentanglement

Many recent works on disentangled representation aim to learn an interpretable and transferability representation. For example, [19] separates time-independent and time-varying components for long-term video prediction. Some works [19], [20], [21] focus on disentanglement of content and style to achieve multi-domain image translation. It is difficult to define content and style explicitly, and different works may adopt different definitions due to their specific tasks. [22] proposes a unified model that learns disentangled representation for describing and manipulating data across multiple domains. For image restoration, [10] disentangles content and blur features from blurred images. Similarly, we disentangle content and discriminative distortion representations.

In this work, we exploit the disentangled representation of different degradations, and leverage domain-specific information for image quality assessment task.

C. Domain-Aware Applications

The word Domain-aware mostly occurs in NLP applications and its meaning varies from case to case. For example, [24] propose a domain-aware dialog system, which aims to maintain a fluent and natural conversation within the domain as well as during switching of domains. The domain in this case refers to the topic or the theme of the conversation. Slightly different, [24] refer domain to a specific field such as retail, travel and entertainment, and introduce the task of multimodal domain-aware conversations.

Domain in our work refers to the collections of images with different degradation. The process of image restoration for multiple degradations can be formulated as the translation from degraded domain to pristine domain. Thus our work is mostly related to multi-domain image translation. Specifically, [25] recently propose a unified model to achieve multi-domain image-to-image translation. [22] propose a model that is able to perform continuous cross-domain image translation and exhibits ability in learning and disentangling desirable latent representations.

III. OUR APPROACH

A. Problem Formulation

Given a distorted image $I_d$, our goal is to learn a mapping $f : I_d \rightarrow s$, in which $s \in \mathbb{R}^+$ denoted the predicted quality.
with different types and levels of degradations, referred to (HVS), we first restore degraded contents as reference. Assume processes are implemented by two corresponding components, I
aware distorted representation and obtained restored images further train a regression network using the high-level domain-
for different degradations. After convergence of restoration, we reacts on generator to further disentangle discriminative features D
indistinguishable with the real pristine image by the image G
from one domain, the generator I
as
in Figure 2. As shown, given several collections of images B. Overview of the Proposed Approach

The framework of our proposed DA-NR-IQA is demonstrated in Figure 2. The framework for simplicity) tries to distinguish if the input image is real or fake and the domain discriminator D_i recognizes the domain category. The original distorted image I_d and its discrepancy map |I_d - G(I_d)| are fed into Discrepancy-Guided Quality Regression Network (c). The two images are first through feature extractors to get their corresponding features f_d and f_r. The two feature extractors share the same weights. Then f_d and f_r are concatenated with high-level domain-aware distortion representation f(H(I_d)), which is extracted from the restoration network. The fused feature is regressed to a patchwise quality and weight estimation. The score of an image is calculated by weighted averaging all scores of its local patches.

score of I_d and should be consistent with the result of Mean Opinion Score (MOS). To simulate Human Visual System (HVS), we first restore degraded contents as reference. Assume in a dataset, there are n domains {D_1, D_2, · · ·, D_n} of images. Each domain represents a collection of images with certain degradation or distortion. For each image I_d ∈ D_i, it can be represented as the combination of pristine image I_g and a certain distortion d, i.e., I_d = I_g ⊗ d. Our goal is to disentangle the representation d of a distorted image I_d to get the restored image I_r consistent with I_g, and use this domain-aware distortion pattern to assess image quality score s under the supervision of ground truth human visual quality score s_gt.

B. Overview of the Proposed Approach

The framework of our proposed DA-NR-IQA is demonstrated in Figure 2. As shown, given several collections of images with different types and levels of degradations, referred to as n domains, an image I_d ∈ D_i is randomly selected from one domain, the generator G tries to produce an image indistinguishable with the real pristine image by the image discriminator D. At the same time, the domain discriminator D_i tries to categorize the labels c_i of representation and also reacts on generator to further disentangle discriminative features for different degradations. After convergence of restoration, we further train a regression network using the high-level domain-aware distorted representation and obtained restored images I_r to get the desired quality score s. The aforementioned processes are implemented by two corresponding components, domain-aware image restoration network and discrepancy-guided quality regression network, respectively.

C. Domain-Aware Image Restoration Network

The overview of proposed Domain-Aware Image Restoration Network (DA-Restore) are shown in Figure 3. This DA-Restore module takes in the distorted image and aims to produce the corresponding restoration. Meanwhile, the model tries to disentangle the distorted representation of specific degradation and content. The restored image could act as a hallucinated reference for the distorted image, which compensates the absence of true reference information and simulates the behavior of the human visual system. Furthermore, due to the design of domain awareness, it declares the determined distortion type of the input distorted image.

1) Latent Feature Loss: The latent feature loss aims to penalize the latent representations from two aspects. To learn disentangled representation across domains, we use the first term L_{kl} = KL(q(z|x)||p(z)) to calculate Kullback-Leibler divergence, which makes the latent code z close to a prior Gaussian distribution p(z) ( as z ≈ N(0, I)). However, this term alone cannot guarantee the disentanglement of domain-specific information from the latent space, since the generator recovers the distorted-free images simply from the representation z without any domain information.

To achieve simultaneous training of multiple domains with a single model, we extend the loss by adding a domain classification loss term. We assign a unique class label for each domain, and the auxiliary domain classifier tries to distinguish images from different domains.
Different from [25], which aims to translate facial attributes among domains, image restoration transfers several domains of different distortions into one single distortion-free domain. Thus, we assign labels on representations instead of images. More precisely, we introduce a domain discriminator \( D_c \), which takes the latent representation \( z \) in domain category \( c \) as input and aims to distinguish the predicted domain vector from its real domain vector. In contrast, the encoder \( E \) tries to confuse \( D_c \) from predicting the correct domain category. The second term of latent feature loss, named domain classification loss, can be defined as:

\[
\mathcal{L}_{E} = -\mathcal{L}_{D}^{cls} = -E_{x,c}[- \log D_c(v_c|E(x,c))],
\]

where \( z = E(x,c) \), and \( x \) here donates the input image. Ideally, the latent representation \( z \) should contain distortion-free image content information together with disentangled domain-specific degradation information.

2) Perceptual Loss: From our preliminary experiments, we observe some unpleasing artifacts in the restored images. Motivated by [6], we add a perceptual loss [26] between the original distorted images and the corresponding restored ones:

\[
\mathcal{L}_P = E[\frac{1}{N} \|\phi_l(I_r) - \phi_l(I_d)\|_1],
\]

where \( \phi_l \) donates \( l \)th layer of the feature maps extracted from the pre-trained VGG-19 network.

3) Adversarial Loss: The VAE architecture tends to generate blurry samples [27], which would not be desirable for practical uses. To get satisfactory image restoration from latent representation, we additionally introduce an image discriminator \( D \), it also enhances the ability of representation disentanglement from the latent space. We define the objective functions \( \mathcal{L}^{adv} \) and \( \mathcal{L}^{adv}_G \) for adversarial learning between image discriminator \( D \) and generator \( G \) as:

\[
\mathcal{L}^{adv}_D = E[\log(D(\hat{x}))] + E[\log(1 - D(x))]
\]

\[
\mathcal{L}^{adv}_G = -E[\log(D(\hat{x}))]
\]

where \( x \) and \( \hat{x} \) donate the input image and its restoration, respectively.

4) Overall Training Loss: For stable training, high image quality and considerable diversity, we use the least-squares GAN [28] in our experiment. Thus, the total training loss functions of the encoder \( E \), decoder \( G \) and image discriminator \( D \) are defined as follows:

\[
\mathcal{L}_E^C = \lambda_1 \mathcal{L}_{E} + \lambda_2 \mathcal{L}_P + \mathcal{L}_{E}^{cls}, \quad \mathcal{L}_G^C = \mathcal{L}_{E}^{cls}, \quad \mathcal{L}_D = \mathcal{L}_D^{adv},
\]

where \( \lambda_1 \) and \( \lambda_2 \) are regularization parameters controlling the importance of losses.

D. Discrepancy-Guided Quality Regression Network

As shown in Figure 2, the DA-Restore module generates the residual image for restoration. This residual image is different from the concept of error map in FR-IQA, which represents pixel-wise error between the distorted image and the reference. To emphasize the difference, we refer to the residual image as discrepancy map and use the original distorted image together with the discrepancy map as input to acquire the desired quality score. The NR-IQA problem is now formulated as a regression by solving

\[
\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(R(I_d, |I_d - G(I_d)|), s_{gt}),
\]

where \( R(\cdot) \) denotes the regression network for predicting the quality score, \(|I_d - G(I_d)|\) is the discrepancy map generated by the DA-Restore module \( G \), \( s_{gt} \) represents the ground truth score.

Previous studies have shown that the quality results obtained by methods based on HVS are greatly affected by the eligibility of the restoration [6, 7]. An unqualified hallucination as the reference would introduce large bias by deteriorating the gap between the distorted image and the restored one. To alleviate this drawback and stabilize the quality regression, we fuse discrepancy information together with high-level fusion from the generative network as in [6]. Different from their work, ours contains domain-specific distortion information.

Assume \( G \) has been trained, as shown in Figure 2, features of original distorted image and its discrepancy map, \( f_r \) and \( f_d \), are extracted by a CNN and are concatenated with high-level domain-aware distortion representation \( f(\mathcal{H}(I_d)) \), which is extracted from the restoration network. \( \mathcal{H}(I_d) \) donates the feature maps of the distorted image, \( f(\cdot) \) is a linear projection to make sure the dimensions of \( \mathcal{H} \) and \( \mathcal{K} \) are equal, \( \mathcal{K} \) represents the feature maps that concatenate with \( \mathcal{H} \). Then the fused feature is regressed to a patchwise quality and weight estimation, i.e., \( s_i \), \( w_i \), \( w_i \) are treated as relative importance for each patch \( i \). The loss function of the patch quality estimation would be formulated as:

\[
\mathcal{L}_R = \frac{1}{T} \sum_{i=1}^{T} \|s_i - s_{gt}\|_1,
\]

where \( s_i = \mathcal{R}_2(f(\mathcal{H}(I_d)) \oplus \mathcal{K}(I_d, |I_d - G(I_d)|)), \mathcal{R}_2 \) is the fully connected layers of \( \mathcal{R} \), \( \oplus \) denotes the concatenation operation and \( T \) is the number of patches.

We assign different weights for the respective patches and use these normalized weight to estimate quality score of the whole image. The reason of integrating a weight estimation...
branch is that simply averaging all local quality scores does not consider the effect of spatially variance of relative image quality and perceptual relevance of local quality. The two branches are parallel and share the same dimensionality. The weight \( w_i \) of each input patch \( s_i \) is calculated by activating the output of weight estimation branch \( w' \) through a ReLU and adding a small stability term:

\[
w_i = \max(0, w') + \epsilon.
\]  

The final image quality score \( \hat{s} \) is calculated using the following two kinds of weighted sum:

\[
\hat{s} = \sum_i w_i s_i.
\]

1) Network Architecture: For domain-aware image restoration network, inspired by recent image translation studies, our generators follow an encoder-decoder architecture similar to the method in \[22\].

Given depth \( d = 5 \), the \( i \)th layer of Encoder \( E \) operates on \( 4 \times 4 \) spatial regions with a stride 2 and produces a feature map with size of \( \{64 \times 2^{d-1}\}_{j=1}^{d} \), i.e., 16, 128, 256, 512, 1024, respectively. Each convolutional layer is followed by Instance Normalization \[33\] and Leaky ReLU are utilized.

Generator \( G \) follows a reversed symmetrical architecture of \( E \). Finally, the idea of residual learning is adopted because it has been shown effective for image processing tasks and helpful on convergence. That is, the generator only learns the difference between the input image and the ground truth image. The domain discriminator \( D_c \) (discriminator with an auxiliary domain classifier) is built on top of the discriminator \( D \).

For discrepancy-guided quality regression network, the domain-aware distortion representation is extracted from well-trained restoration network and acts as compensation for IQA. The discrepancy-guided quality regression network takes the distorted patch and the corresponding restored residual patch as input. Feature representations \( f_r, f_d \) are extracted by the same layers and fused with obtained feature map \( \hat{z} \). The fused feature vector is then fed into two branches for perceptual score \( s_i \) and weight \( w_i \), respectively.

### IV. Experiments

In this section, we conduct several experiments to test the performance of our proposed method on various IQA benchmarks. We pretrain DA-NR-IQA on Waterloo Exploration \[34\], perform cross validation on TID2013 \[35\] and LIVE \[36\].

#### A. Datasets and Evaluation Protocols

1) TID2013: TID2013 includes 25 distortion-free reference images and 3000 distorted images. These images are created from references with 24 types and 5 levels of distortions, ranging from additive Gaussian noise to Chromatic aberrations. Every image is annotated with Mean Opinion Scores (MOS), which is produced by several observers in subjective tests. The obtained MOS has to vary from 0 to 9 and its larger values correspond to better perceptual quality. Its wide range makes it one of the most comprehensive IQA databases.

2) LIVE: LIVE consists of 29 reference images and 779 distorted samples with 5 distortion types including Fast Fading, Gaussian Blur, White Noise, JPEG Compression and JP2K Compression. Each image is provided with Differential Mean Opinion Scores (DMOS), ranging from 0 to 100. Lower DMOS means higher perceptual quality. DMOS value of zero indicates the image is distortion-free.

3) Waterloo Exploration: Waterloo Exploration contains 4744 pristine natural images and 94880 distorted images. The distorted images are generated by MATLAB with four distortion types and five levels. Compared to TID2013 and LIVE, Waterloo Exploration has much larger amounts of images, thus it also has a great diversity of image content. The four types, i.e., JPEG Compression, JP2K Compression, Gaussian Blur and White Noise, are also considered the most common distortion types and are covered both in TID2013 and LIVE. Instead of annotating distorted data with subjective mean opinion score (MOS), which is impractical for such a large database, Waterloo Exploration claims to preset MATLAB parameters which cover a wide range of subjective quality scale.

4) Evaluation Metrics: Following most previous works \[4\], \[5\], \[7\], the performances on above datasets are evaluated

| Category | Method       | TID2013 SROCC | TID2013 PLCC | LIVE SROCC | LIVE PLCC |
|----------|--------------|---------------|--------------|------------|-----------|
| FR-IQA   | PSNR         | 0.889         | 0.847        | 0.880      | 0.805     |
|          | SSIM [29]    | 0.856         | 0.867        | 0.918      | 0.780     |
|          | FSIM [30]    | **0.963**     | **0.932**    | **0.952**  | **0.822** |
|          | VSI [31]     | 0.947         | **0.939**    | 0.936      | **0.853** |
| NR-IQA   | DIVINE [13]  | 0.855         | 0.831        | 0.885      | 0.853     |
|          | BLINDS-II [12]| 0.877        | 0.841        | 0.931      | 0.930     |
|          | BRISQUE [3]  | 0.922         | 0.917        | 0.940      | 0.911     |
|          | CORNIA [14]  | 0.903         | 0.917        | 0.913      | 0.888     |
|          | CNN [4]      | 0.903         | 0.917        | 0.913      | 0.888     |
|          | CNN++ [9]    | 0.843         | 0.804        | 0.928      | 0.897     |
|          | DIQaM-NR [5] | 0.933         | 0.909        | 0.960      | 0.972     |
|          | RAN [7]      | 0.948         | **0.937**    | 0.972      | 0.968     |
|          | DA-NR-IQA    | **0.952**     | 0.929        | **0.977**  | **0.975** |

TABLE I: CROSS VALIDATION ON TID2013 AND LIVE. OUR PROPOSED METHOD PERFORMS BETTER THAN ALMOST ALL THE OTHER STATE-OF-THE-ART NR-IQA METHODS IN TERMS OF BOTH PLCC AND SROCC ON LIVE DATASET.
by two common metrics for model evaluation: the Linear Correlation Coefficient (LCC) and the Spearman’s Rank Order Correlation Coefficient (SROCC). LCC is a measure of the linear correlation between the ground-truth and model prediction, which is defined as

\[
LCC = \frac{\sum_{i=1}^{N} (y_i - \bar{y}_i) (\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^{N} (\hat{y}_i - \bar{\hat{y}}_i)^2}},
\]

(9)

where \(\bar{y}_i\) and \(\bar{\hat{y}}_i\) denote the means of the ground truth and predicted score, respectively. SROCC is a measure of the monotonic relationship between the ground-truth and model prediction, which could be formulated as:

\[
SROCC = 1 - \frac{6 \sum_{i=1}^{N} (u_i - p_i)^2}{N (N^2 - 1)},
\]

(10)

where \(N\) represents the number of distorted images, and \(r\) is the difference of ranking.

B. Implementation Details

All the training images are randomly sampled from the original images of size 256 × 256 with stride 96. We train our model with Pytorch on the GeForce GTX 1080Ti with a batch size of 32. We apply Adam [37] solver with parameters of learning rate 0.0002, \(\beta_1 = 0.5, \beta_2 = 0.999\). For domain-aware image restoration, the weights \(\lambda_1, \lambda_2\) are set as 5 and 100, respectively. For discrepancy-guided quality regression, we split the TID2013 and LIVE datasets into 6:2:2 for training, validation, and test, respectively. During the testing of image quality assessment, we extract overlapped patches from each test image at a fixed stride and calculate the weighted scores of all patches as the final quality score.

C. Cross Validation on TID2013 and LIVE

We firstly train the domain-aware image restoration module on Waterloo Exploration, then train the discrepancy-guided quality regression on TID2013 and LIVE to perform cross validation, respectively. Since Waterloo Exploration contains only four distortion types, we train and test DA-NR-IQA on these following types: Gaussian Blur, White Noise, JPEG and JP2K, following the same setting as in [7]. As shown in Table [I] the proposed DA-NR-IQA performs better than almost all the other state-of-the-art NR-IQA methods in terms of both PLCC and SROCC on LIVE dataset.

Furthermore, our model outputs simultaneously quality estimation and distortion identification. As shown in Table [II] our method also wins the distortion identification task compared with other distortion-identification IQA methods [38], [9]. Table [III] shows the confusion matrices produced by our method on LIVE Dataset. The column and the raw contain ground truths and predicted distortion types, respectively.

For visualization, we also provide the distribution diagrams of subjective MOS values with respect to objective values on LIVE database in Figure [4] in which we denote the distorted images with blue “+” and the black curves are obtained in the curve fitting process as in [39]. One can see the blue “+” of our method gather evenly and close to the black curve and the curve also exhibits almost a straight line, which manifests the better correlation between the scores given by our method and the subjective judgements for the image quality.

### Table II

| Method          | Accuracy          |
|-----------------|-------------------|
| LIVE TID2013    |                   |
| CNN++ [9]       | 0.925             | 0.819             |
| MEON [38]       | 0.912             | 0.859             |
| DA-NR-IQA       | 0.942             | 0.937             |

### D. Results on More Distortion Types

We show results on four distortions in Section [V.C] to demonstrate the scalability of our domain-aware framework when handling more distortion types, we further train and test on the full TID2013 dataset. For pre-training, as in the previous studies [40], [38], we reproduce 17 out of a total of 24 distortion types in TID2013 and apply them to the 840 high-quality images. For the distortions we did not reproduce (i.e., #3, #4, #12, #13, #20, #21, #24), we fine-tune from other distortions. From Table [IV] we can observe that the proposed method outperforms previous models by a clear margin.

### E. Feature Disentanglement Visualization

To demonstrate the ability for disentanglement and transferability of learned features, many works [41], [22] conduct feature visualization with t-SNE [42]. Similarly, we simultaneously perform feature disentanglement of distortions at different domains and show the results in Figure [5] Each color indicates a different domain, i.e., noisy images with \(\sigma = 15/25/30/50/70\), JPEG compressed images with quality factor 10/20 and low-resolution images with factor 2/3/4. As shown, features of images from different domains are discriminated significantly well.

### F. Ablation study

To demonstrate the efficacy of the key components of our method for the performance, we conduct several ablation experiments on TID2013, in which we remove perceptual loss, adversarial learning, high-level semantic fusion, or domain classification and test the performance of the remaining framework by comparing both SROCC and LCC results, as shown in Table [V].

1) Domain Classification: To show how “domain classification” contributes to the performance, we remove the domain classification mechanism. Domain classification mechanism is a crucial component to make our approach “domain-aware”. As illustrated in Figure [5] without domain classification, the representation disentanglement would be only separation of content and degradation. The representations of different degradations are still entangled and no discriminative features of certain degradation are learned. The results in Table [V] show that the discrimination of different degradations can actually boost the performance of the IQA task.
Fig. 4. The distribution diagrams of MOS values with respect to objective values on LIVE dataset.

### TABLE III

| Category | JP2K | JPEG | WN | BLUR | Pristine |
|----------|------|------|----|------|----------|
| JP2K     | 0.915| 0.010| 0.000| 0.023| 0.032    |
| JPEG     | 0.048| 0.919| 0.000| 0.022| 0.011    |
| WN       | 0.000| 0.000| 1.000| 0.000| 0.000    |
| BLUR     | 0.059| 0.007| 0.000| 0.926| 0.008    |
| Pristine | 0.007| 0.013| 0.000| 0.034| 0.950    |

Fig. 5. Feature visualization (a) without and (b) with domain-aware mechanism. Better view in color.

2) **High-Level Semantic Fusion.**: The two aforementioned mechanisms contribute to the IQA performance by indirectly improving the quality of restored reference. When the restoration is unqualified, a large bias would be introduced and lead to deterioration of the gap between the distorted image and the restored one. Thus we design high-level semantic fusion to alleviate this drawback and stabilize the quality regression. To show its impact, we remove the fusion module, and the ablation results shown in Table [X] demonstrate the validity.

3) **Perceptual Loss.**: We first evaluate the effect of perceptual loss. The ablated model obtains an obvious performance decline by removing the perceptual loss since such a loss helps to restoration at the training process.

4) **Adversarial Learning.**: To explore how adversarial learning contributes to the restoration performance, we further evaluate the model without image discriminator and adversarial loss. Removal of adversarial learning leads to significant performance decline since the discriminator no longer propelled the generator.

V. CONCLUSION AND DISCUSSION

In this paper, we propose the domain-aware no-reference image quality assessment (DA-NR-IQA). The proposed DA-NR-IQA exploits and leverages the discriminative disentangled feature representations of specific degradation for image quality assessment. Benefiting from the design of domain-aware network architecture, our method is able to identify the distortion type of an image, and use the determined distortion type and quality score to characterize the image quality. Experiments on various standard IQA dataset have shown its superiority over state-of-the-art IQA methods.

REFERENCES

[1] W. Yang, X. Zhang, Y. Tian, W. Wang, and J. Xue, “Deep learning for single image super-resolution: A brief review,” *IEEE Trans. Multimedia*, 2019.
[2] B. Yan, B. Bare, C. Ma, K. Li, and W. Tan, “Deep objective quality assessment driven single image super-resolution,” *IEEE Trans. Multimedia*, 2019.
[3] J. Liu, S. Yang, Y. Fang, and Z. Guo, “Structure-guided image inpainting using homography transformation,” *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3252–3265, 2018.
[4] L. Kang, P. Ye, Y. Li, and D. S. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1733–1740.
[5] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek, “Deep neural networks for no-reference and full-reference image quality assessment,” *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, 2018.
[6] K. Lin and G. Wang, “Hallucinated-iqa: No-reference image quality assessment via adversarial learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 732–741.
[7] H. Ren, D. Chen, and Y. Wang, “RAN4IQA: restorative adversarial nets for no-reference image quality assessment,” in *Proc. Conf. Arti. Intell.*, 2018, pp. 7308–7314.
| Method       | #01 | #02 | #03 | #04 | #05 | #06 | #07 | #08 | #09 | #10 | #11 | #12 | #13 |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| TID2013      | 0.714 | 0.728 | 0.825 | 0.358 | 0.852 | 0.664 | 0.780 | 0.852 | 0.754 | 0.808 | 0.862 | 0.251 | 0.755 |
| BRISQUE [13] | 0.630 | 0.424 | 0.727 | 0.125 | 0.032 | 0.560 | 0.282 | 0.680 | 0.804 | 0.715 | 0.800 | 0.562 | 0.175 |
| CORNIA [14]  | 0.341 | -0.196 | 0.689 | 0.184 | 0.155 | 0.125 | 0.032 | 0.560 | 0.282 | 0.680 | 0.804 | 0.715 | 0.800 |
| RankIQA      | 0.597 | 0.622 | 0.268 | 0.613 | 0.662 | 0.619 | 0.644 | 0.800 | 0.779 | 0.629 | 0.859 | 0.780 |
| DA-NR-IQA    | 0.477 | 0.695 | 0.438 | 0.674 | 0.709 | 0.709 | 0.709 | 0.852 | 0.713 | 0.897 | 0.808 | 0.784 | 0.882 |

**TABLE IV**

**Performance Evaluation (SROCC) on the Entire TID2013 Dataset.**

| Method       | #14 | #15 | #16 | #17 | #18 | #19 | #20 | #21 | #22 | #23 | #24 | ALL |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| TID2013      | 0.081 | 0.371 | 0.159 | -0.082 | 0.109 | 0.699 | 0.222 | 0.451 | 0.815 | 0.568 | 0.856 | 0.550 |
| BRISQUE [13] | 0.175 | 0.184 | 0.155 | 0.125 | 0.032 | 0.560 | 0.282 | 0.680 | 0.804 | 0.715 | 0.800 | 0.562 |
| CORNIA [14]  | 0.161 | 0.949 | 0.727 | 0.321 | 0.032 | 0.560 | 0.282 | 0.680 | 0.804 | 0.715 | 0.800 | 0.562 |
| RankIQA      | 0.597 | 0.622 | 0.268 | 0.613 | 0.662 | 0.619 | 0.644 | 0.800 | 0.779 | 0.629 | 0.859 | 0.780 |
| DA-NR-IQA    | 0.477 | 0.695 | 0.438 | 0.674 | 0.709 | 0.709 | 0.709 | 0.852 | 0.713 | 0.897 | 0.808 | 0.784 |

**TABLE V**

**Ablation Experiment on TID2013. “w/o M” means our model without component M.**

| Ablation                      | TID2013 PLCC | TID2013 SROCC |
|-------------------------------|-------------|--------------|
| Proposed (Ours)               | 0.929 | 0.952 |
| w/o Domain Classification     | 0.913 | 0.934 |
| w/o Semantic Fusion           | 0.917 | 0.902 |
| w/o Perceptual Loss           | 0.883 | 0.876 |
| w/o Adversarial Learning      | 0.864 | 0.859 |

[8] H. Lim, H. G. Kim, and Y. M. Ra, “VR IQA NET: deep virtual reality image quality assessment using adversarial learning,” in Proc. Conf. Acoustics, Speech and Signal Process., 2018, pp. 6377–6381.

[9] L. Kang, P. Ye, Y. Li, and D. S. Doermann, “Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks,” in Proc. Conf. Image Process., 2015, pp. 2791–2795.

[10] B. Lu, J.-C. Chen, and R. Chellappa, “Unsupervised domain-specific deblurring via disentangled representations,” CoRR, vol. abs/1903.01594, 2019.

[11] C. Huang, T. Jiang, and M. Jiang, “Encoding distortions for multi-task full-reference image quality assessment,” in Proc. IEEE Int. Conf. Multimedia Expo, 2019, pp. 1864–1869.

[12] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” IEEE Trans. Image Process., vol. 20, no. 12, pp. 3350–3364, 2011.

[13] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, 2012.

[14] P. Ye, J. Kumar, L. Kang, and D. S. Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 1098–1105.

[15] K. Friston, J. Kilner, and L. Harrison, “A free energy principle for the brain,” Journal of Physiology-Paris, vol. 100, pp. 70–87, 2006.

[16] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, “A psychovisual quality metric in free-energy principle,” IEEE Trans. Image Process., vol. 21, no. 1, pp. 41–52, 2012.

[17] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Using free energy principle for blind image quality assessment,” IEEE Trans. Multimedia, vol. 17, no. 1, pp. 50–63, 2015.

[18] E. L. Denton et al., “Unsupervised learning of disentangled representations from video,” in Proc. Adv. Neural Inf. Process. Sys., 2017, pp. 4414–4423.

[19] T. Xiao, J. Hong, and J. Mu, “Elegant: Exchanging latent encodings with gan for transferring multiple face attributes,” in Proc. Eur. Conf. Comput. Vis., September 2018, pp. 172–187.

[20] H. Lee, H. Tseng, J. Huang, M. Singh, and M. Yang, “Diverse image-to-image translation via disentangled representations,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 36–52.

[21] X. Huang, M. Liu, S. J. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 179–196.
[41] J. Hoffman, S. Guadarrama, E. Tzeng, R. Hu, J. Donahue, R. B. Girshick, T. Darrell, and K. Saenko, “LSDA: large scale detection through adaptation,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 3536–3544.
[42] P. E. Rauber, A. X. Falcão, and A. C. Telea, “Visualizing time-dependent data using dynamic t-sne,” in Euro. Conf. Vis., 2016, pp. 73–77.