No-Reference Image Quality Assessment Using Dynamic Complex-Valued Neural Model

Zihan Zhou
South China University of Technology
Guangzhou, China
cszzh@mail.scut.edu.cn

Yong Xu*
South China University of Technology
Guangzhou, China
yxu@scut.edu.cn

Ruotao Xu†‡
South China University of Technology
Guangzhou, China
xrt@scut.edu.cn

Yuhui Quan§¶
South China University of Technology
Guangzhou, China
csyhquan@scut.edu.cn

ABSTRACT

Deep convolutional neural networks (CNNs) have become a promising approach to no-reference image quality assessment (NR-IQA). This paper aims at improving the power of CNNs for NR-IQA in two aspects. Firstly, motivated by the deep connection between complex-valued transforms and human visual perception, we introduce complex-valued convolutions and phase-aware activations beyond traditional real-valued CNNs, which improves the accuracy of NR-IQA without bringing noticeable additional computational costs. Secondly, considering the content-awareness of visual quality perception, we include a dynamic filtering module for better extracting content-aware features, which predicts features based on both local content and global semantics. These two improvements lead to a complex-valued content-aware neural NR-IQA model with good generalization. Extensive experiments on both synthetically and authentically distorted data have demonstrated the state-of-the-art performance of the proposed approach.

CCS CONCEPTS

• **Computing methodologies → Computer vision**.

KEYWORDS

Image quality assessment, Complex-valued convolutional networks, Dynamic filtering, Deep learning

ACM Reference Format:
Zihan Zhou, Yong Xu, Ruotao Xu, and Yuhui Quan. 2022. No-Reference Image Quality Assessment Using Dynamic Complex-Valued Neural Model. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisbon, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3547982

*Yong Xu is also with Pazhou Laboratory of Guangzhou, Peng Cheng Laboratory of Shenzhen, and Communication and Computer Network Laboratory of Guangdong.
†Ruotao Xu is also with Pazhou Laboratory of Guangzhou, China.
‡Corresponding Author: Ruotao Xu.
§Yuhui Quan is also with Pazhou Laboratory of Guangzhou, China.
¶This project was initiated and led by Yuhui Quan.

MM '22, October 10–14, 2022, Lisbon, Portugal.

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3547982

1 INTRODUCTION

Image quality assessment (IQA) aims at quantifying human perception of image quality. In the past decades, huge efforts have been devoted to no-reference IQA (NR-IQA) which is about estimating the perceived quality of a distorted image without accessing any information of its pristine counterpart. Such a topic has received extensive attention from both industry and academic, due to its great potentials and values in many practical applications [16, 34].

The key of NR-IQA is how to extract quality-related features from only a single distorted image. Traditional NR-IQA methods extract such features using hand-crafted statistical models; see *e.g.* [11, 21, 53]. In recent years, the great success of convolutional neural networks (CNNs) achieved in image recognition has spawned a series of CNN-based NR-IQA approaches; see *e.g.* [42, 50, 60]. Following this line of research, this paper investigates the further potentials of CNN-based methods for NR-IQA. We aim to improve the power of CNNs for NR-IQA in the following two aspects.

Introducing complex-valued representations to NR-IQA CNNs. Most existing CNN models for IQA are built upon real-valued operations, while the power of complex-valued transformations has not been exploited yet. Indeed, complex transforms are not strangers to IQA. They have been widely adopted for quality-aware feature extraction, such as Gabor transform [46], discrete Fourier transform [31], and complex wavelet [58]. Another example is phase-related features (*e.g.* phase congruency), introduced by complex-valued representation, and they have been widely exploited in IQA; see *e.g.* [23, 52]. In the view of biological perception in human visual system, a visual signal is transmitted to the primary visual cortex (also called V1) through the lateral geniculate nucleus for visual abstraction processing, and responses of the cells in V1 are characterized by the selectivity, orientation and frequency, which can be modeled by complex-valued transforms [8, 36].

The merits of complex-valued transforms and representations inspired us to introduce complex-valued convolutions and phase-aware activations to exploit such merits in IQA-oriented CNNs. Complex-valued convolutional layers also benefit the model compactness. The information of a real-valued feature tensor can be

perfectly represented by its complex-valued counterpart while using one-half channels, as a complex number has its real and imaginary parts. Then, for a real-valued convolutional layer, its complex-valued counterpart may halve both the input and output channels. Together with that a 2D complex-valued kernel doubles the weights over its real-valued counterpart, we have that a complex-valued convolutional layer can halve the number of parameters over its real-valued counterpart.

Dynamic filtering for improving content-awareness in feature extraction. A traditional CNN model is typically content-agnostic since the spatially-invariant convolutional filters are shared across different pixels and different input images for a trained model. However, an ideal IQA model usually demands content-aware feature extraction. As shown in [20], even for the same pattern, human perceive different visual qualities when the content of its neighborhood and the semantics of the image vary. In general, different image contents may be sensitive to different distortions, *e.g.*, a flattened area is usually sensitive to noise while a textured area is sensitive to blur. Consider a large flattened area. It may be considered a high-quality pattern when it is located in an image of a clear blue sky, while it is regarded as a serious blurring in a textured image. Then, the extraction of quality-related features should be content-aware and the filters should be related to the image patterns to be analyzed. Due to the diversity of distortion and the variation of image contents, it is difficult to construct a fixed set of filters to cover all possible patterns, particularly for authentically distorted images. To improve the content-awareness, we introduce the dynamic filtering technique to the complex-valued NN for NR-IQA, where the filters are conditioned on both the local contents and global semantics for feature extraction.

Based on the aforementioned two improvements, we develop a dynamic complex-valued neural model for NR-IQA. Our contributions in this work are three-fold:

- We introduce complex-valued deep learning for NR-IQA, which utilizes the merits of complex-valued representations and transformations for improving the effectiveness of CNN-based NR-IQA. To the best of our knowledge, this is the first work to study complex-valued neural models for IQA.
- We introduce dynamic filtering to the complex-valued CNN for both locally and globally content-aware feature extraction, which can better mimic human perception of visual quality in handling diverse image contents and distortions in authentic cases. This is also the first work to apply dynamic filtering to NR-IQA.
- A dynamic complex-valued CNN is proposed for NR-IQA. Benefiting from its compactness and effectiveness, the proposed model can be trained fast without the need for backbone fine-tuning. Its effectiveness is demonstrated by the extensive experiments on IQA benchmark datasets.

2 RELATED WORK

2.1 NR-IQA Models

Traditional NR-IQA approaches involve two steps: quality-related feature extraction done by a hand-crafted process, *e.g.*, using natural scene statistics (NSS) [11, 29] or local binary patterns [21, 53], and quality score regression done by a learning-based model, *e.g.*, support vector regressor, random forest, and Gaussian process regression. Complex-valued representations are often exploited in the hand-crafted feature extraction process; see *e.g.* [31, 46, 58].

In recent years, an increasing number of CNN-based NR-IQA methods have been proposed with significant improvement achieved. As IQA is typically a small sample problem due to the expensive costs of collecting subjective scores of IQA, some studies (*e.g.* [2, 42]) focused on addressing the insufficiency of labeled data for training NR-IQA CNNs by exploring patch-wise pseudo labels. More studies introduce auxiliary tasks to exploit additional labeled data from a different source, *e.g.* distortion type classification [26, 54], image quality ranking [24, 55], and quality map estimation [33].

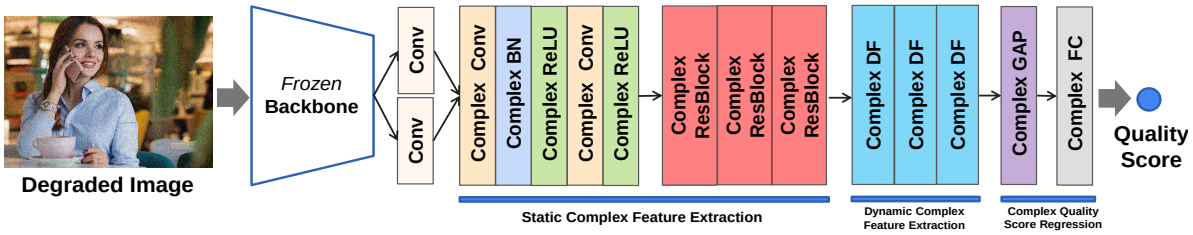
The network architecture design is another focus for deep NR-IQA. Most methods utilize the former layers of recognition backbones (*e.g.* VGG16 [39] and ResNet [13]) for feature extraction, where the backbones pre-trained on recognition tasks are fine-tuned on the IQA tasks. Note that the fine-tuning procedure is usually computationally intensive owing to a large number of trainable parameters in the backbone. Furthermore, a backbone pre-trained for recognition often lacks the capability in extracting quality-related features, particularly in a content-aware manner. Hence, great efforts have been made to build content-aware schemes into pre-trained backbones for improvement. Su *et al.* [40] proposed a content-aware regressor whose weights are predicted with image features. Gu *et al.* [12] introduced attention to the weight different image areas with predicted perceptual importance. For improvement, Chen *et al.* [3] used reinforcement learning to better capture the attention information in the input image. Note that while these attention mechanisms generate content-aware weights on extracted features, the feature extraction process remains content-agnostic. Ke [17] introduced the transformer architecture to NR-IQA, where self-attention builds a similarity-based transform for image-adaptive feature extraction. The transformer architecture usually requires much more learnable parameters and computation burden, which limits its applications.

2.2 Complex-Valued CNNs

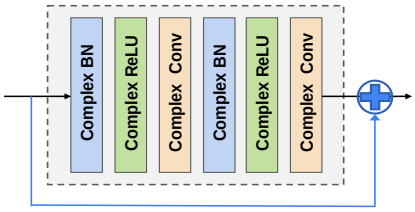
Complex-valued CNNs have been investigated for a long time. Many studies focus on theoretical aspects, such as optimization [1, 32], generalization [14], representational capacity [43], and invariance [5]. Most of these studies were conducted in the context of image classification. Recently, complex-valued CNNs have been applied to image recovery [36, 37], with great potentials demonstrated. However, the potentials of complex-valued CNNs for IQA have not been investigated yet.

2.3 Dynamic Filtering

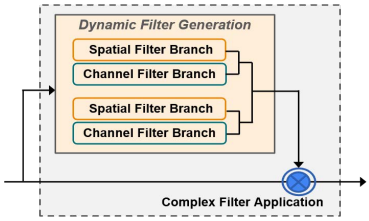
Dynamic filtering aims at overcoming the content-agnostic property of a standard convolution, so as to improve the capability and adaptivity of a CNN in handling spatially-varying image structures or effects. There are a series of works, such as conditionally-parameterized convolutions [47] and dynamic convolutions [57], which predict coefficients to combine several expert filters. These convolutions are still applied in a spatially-shared way. The kernel prediction network [28] addressed this by generating an individual



(a) Framework of DCNet



(b) Diagram of complex ResBlock



(c) Diagram of complex DF

Figure 1: Proposed DCNet for NR-IQA. Conv for Convolution, ReLU for Rectified Linear Unit, BN for Batch Normalization, ResBlock for Residual Block, DF for Dynamic Filtering, GAP for Global Average Pooling, and FC for Fully-Connected.

filter for every pixel location. To adjust the filter neighborhoods dynamically, the deformable convolution [7] adds offsets to the convolutional neighborhoods. To reduce the computational cost of kernel prediction, the decoupled dynamic filtering [59] uses a composite strategy with decoupled spatial and channel dynamic filters. Aiming at the deblurring problem, Quan *et al.* propose kernel mixture learning [38] by decomposing the predicted kernels into a group of spatially-shared bases and some spatially-varying coefficients, which considerably reduces the model size. All these works are for image recognition or recovery, not IQA.

3 PROPOSED APPROACH

The proposed NR-IQA model called DCNet (Dynamic Complex-valued Network) consists of two parts: a static complex-valued CNN for the pre-stage feature extraction, and a complex-valued dynamic filtering module for content-aware feature extraction. Finally, a complex-valued regressor is applied to predict the quality score with extracted complex features. Similar to existing works, the DCNet is combined with a pre-trained backbone for improvement.

3.1 Pre-Stage Static Feature Extraction

Given the real-valued feature tensor $\boldsymbol{X}_0$ formed by the pre-trained backbone on the input image, we first transform it to a complex-valued feature tensor by two convolutional layers as follows:

$$\boldsymbol{X} = \mathrm{Conv}_{\mathrm{R}}(\boldsymbol{X}_0) + i \cdot \mathrm{Conv}_{\mathrm{I}}(\boldsymbol{X}_0), \tag{1}$$

where $\mathrm{Conv}_{\mathrm{R}}(\cdot)$, $\mathrm{Conv}_{\mathrm{I}}(\cdot)$ are implemented as a pair of real-valued convolutional layers. Here we do not adopt a pre-defined complex-valued transform but the learnable ones for better adaptivity. Then, a series of complex-valued blocks shown in Figure 1(a) is applied to $\boldsymbol{X}$. The involved basic layers are detailed as follows.

Complex-valued convolutional layer. This layer is constructed by replacing the real-valued kernel with the complex one in the convolution process. A standard complex-valued convolution can be expressed as

$$(\boldsymbol{K} * \boldsymbol{X})(p) = \sum_{p' \in \Omega(p)} \boldsymbol{X}(p')\boldsymbol{K}(p' - p), \tag{2}$$

where $\boldsymbol{X}$, $\boldsymbol{K}$ denote a complex-valued feature map and a complex-valued kernel respectively, p denotes the pixel index, and $\Omega(\cdot)$ denotes convolution window around the p-th pixel. Considering the complex-valued product rule, Equation (2) can be rewritten as

$$\begin{aligned}(\boldsymbol{K} * \boldsymbol{X})(p) = &\sum_{p' \in \Omega(p)} [\boldsymbol{X}^{\mathrm{R}}(p')\boldsymbol{K}^{\mathrm{R}}(p' - p) - \boldsymbol{X}^{\mathrm{I}}(p')\boldsymbol{K}^{\mathrm{I}}(p' - p)] \\ &+ i \cdot \sum_{p' \in \Omega(p)} [\boldsymbol{X}^{\mathrm{R}}(p')\boldsymbol{K}^{\mathrm{I}}(p' - p) + \boldsymbol{X}^{\mathrm{I}}(p')\boldsymbol{K}^{\mathrm{R}}(p' - p)],\end{aligned}$$

where the superscript R and I denote the real and imaginary part of a complex variable, *i.e.*, $X = X^{\mathrm{R}} + i \cdot X^{\mathrm{I}}$. By using real-valued convolutions, we have a more concise form:

$$\boldsymbol{K} * \boldsymbol{X} = (\boldsymbol{K}^{\mathrm{R}} * \boldsymbol{X}^{\mathrm{R}} - \boldsymbol{K}^{\mathrm{I}} * \boldsymbol{X}^{\mathrm{I}}) + i \cdot (\boldsymbol{K}^{\mathrm{R}} * \boldsymbol{X}^{\mathrm{I}} + \boldsymbol{K}^{\mathrm{I}} * \boldsymbol{X}^{\mathrm{R}}). \tag{3}$$

A convolutional layer performs convolution operations in a multi-channel manner. Given the input $\boldsymbol{X} \in \mathbb{C}^{h \times w \times c}$, the output of a complex-valued convolution, denoted by $\boldsymbol{Y} \in \mathbb{C}^{h \times w \times c'}$, is calculated as

$$\boldsymbol{Y}_j = \sum_{i=1}^{c} \boldsymbol{K}_{i,j} * \boldsymbol{X}_i + \boldsymbol{B}_j, \tag{4}$$

for $j = 1, \cdots, c'$, and $\boldsymbol{X}_i$, $\boldsymbol{Y}_j$ denote the i-th channel of input, j-th channel of output, respectively. $\boldsymbol{B}_j$ denotes the bias.

Introducing complex-valued representations to convolutional layers can benefit the model compactness. Consider a real-valued convolutional layer with kernel size of $k \times k$, c_1 input channels and c_2 output channels, which involves $k^2 c_1 c_2$ parameters. To have the same amount of information, a complex-valued convolutional layer with $c_1/2$ input and $c_2/2$ output channels can be used, which has only $k^2 c_1 c_2/2$ parameters. In addition, complex-valued convolutions introduce interactions between real and imaginary parts.

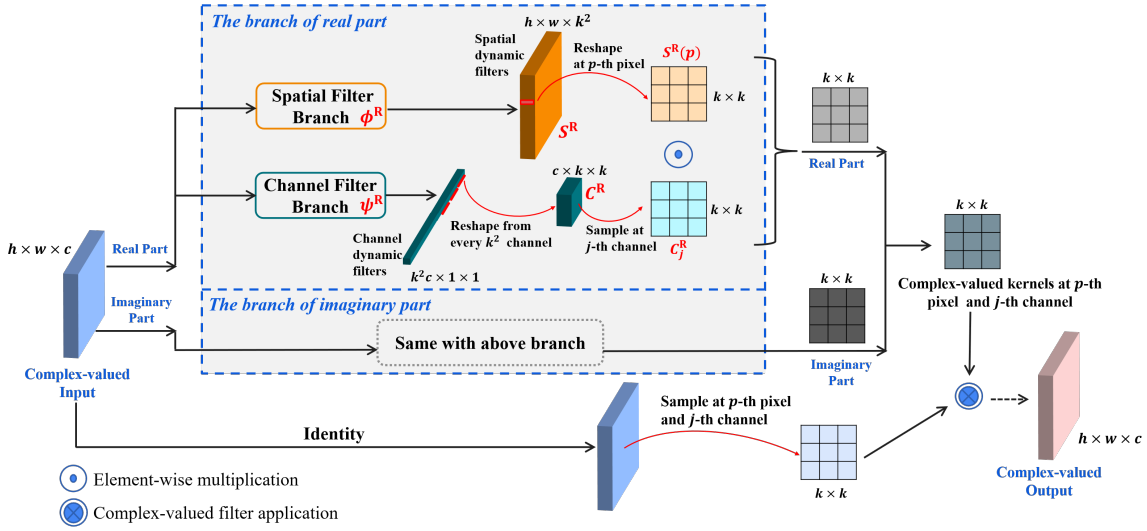


Figure 2: Complex DF module. The complex filter application means applying complex-valued convolution at a single position.

Though such interactions can be implemented in a real-valued one with additional connections, there is no motivation for such connections without the interpretation from complex-valued operations.

Complex-valued ReLU. For the activation function, we adopt complex-valued ReLU throughout our complex-valued CNN. There are many choices for complex-valued ReLU. Among these, we proposed to use the CReLU [43], which applies separate ReLUs on both the real and imaginary parts of a complex number:

$$\mathrm{CReLU}(\boldsymbol{X}) = \mathrm{ReLU}(\boldsymbol{X}^{\mathrm{R}}) + i \cdot \mathrm{ReLU}(\boldsymbol{X}^{\mathrm{I}}). \tag{5}$$

The CReLU defined in (5), though simple, performs a phase-aware activation. Consider the coordinate geometry of a complex number. CReLU keeps the complex number with phase falling in 1st quadrant and clips that with phase in the 3rd quadrant. For a complex number located in the 2nd or 4th quadrant, CReLU only keeps either part.

Complex-valued batch normalization The batch normalization in complex field is done by separately running batch normalization on the real and imaginary parts of complex features respectively:

$$\mathrm{CBN}(\boldsymbol{X}) = \mathrm{BN}(\boldsymbol{X}^{\mathrm{R}}) + i \cdot \mathrm{BN}(\boldsymbol{X}^{\mathrm{I}}), \tag{6}$$

where $\mathrm{BN}(\cdot)$ is the standard batch normalization.

Complex-valued residual block. The complex-valued residual block is a combination of the three modules described above, whose architecture is illustrated in Figure 1(b). Similar to the real-valued ones, it adds a skip connection, which addresses the vanishing gradients in the back-propagation during training, as well as benefits the preservation of image details and feature representations for IQA by feeding the features of the previous residual block to the subsequent blocks.

3.2 Dynamic Feature Extraction

To refine the pre-stage features in a content-aware manner, we follow [6] to construct a complex-valued dynamic filtering module and stack three to construct a dynamic feature extraction process. The module is illustrated in Figure 1(c) and Figure 2. For each channel, we calculate

$$Y_j(p) = \sum_{p' \in \Omega(p)} X_j(p') D_j(p', p). \tag{7}$$

where $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{C}^{h \times w \times c}$ denote the input and output tensors, and $\boldsymbol{D}$ denotes the dynamic kernel tensor. Compared to the static convolution defined in (2), the value $\boldsymbol{D}_j(p', p)$ in the dynamic kernel not only depends on the relative position $(p - p')$, but also involves the absolute position p, leading to a spatially-varying processing. Furthermore, unlike the standard convolution where kernels are shared across different inputs, the kernels in (7) are predicted based on the input. For the j-th channel, the desired complex-valued kernel tensor $\mathcal{W} = \mathcal{W}^{\mathrm{R}} + i \cdot \mathcal{W}^{\mathrm{I}}$ in the size of $h \times w \times c \times k \times k$ is first calculated by

$$\mathcal{W}^{\mathrm{R}} = g_{\mathrm{R}}(\boldsymbol{X}^{\mathrm{R}}) \quad \text{and} \quad \mathcal{W}^{\mathrm{I}} = g_{\mathrm{I}}(\boldsymbol{X}^{\mathrm{I}}), \tag{8}$$

where $g_{\mathrm{R}}(\cdot)$ and $g_{\mathrm{I}}(\cdot)$ are learnable functions to predict the real and imaginary parts separately. Each pixel-wise filter $\boldsymbol{D}_j(\cdot, p) \in \mathbb{C}^{k \times k}$ is then extracted from the kernel $\mathcal{W}_j(p) \in \mathbb{C}^{k \times k}$ located at p-th pixel and j-th channel of $\mathcal{W}$. Note that the outputs of $g_{\mathrm{R}}(\cdot)$ and $g_{\mathrm{I}}(\cdot)$ are both of a huge size, which unavoidably introduces a large number of learnable parameters. Even with a 1×1 convolution implementation, k^2c^2 parameters are needed for $g_{\mathrm{R}}(\cdot)$ or $g_{\mathrm{I}}(\cdot)$. To reduce the number of parameters, following [59], each prediction function ($g_{\mathrm{R}}(\cdot)$ or $g_{\mathrm{I}}(\cdot)$) is separated into two parts: a spatially-varying channel-shared part, and a spatially-shared channel-varying part. Regarding the real part $\mathcal{W}^{\mathrm{R}}$, we have

$$\mathcal{W}_j^{\mathrm{R}}(p) = \boldsymbol{S}^{\mathrm{R}}(p) \odot \boldsymbol{C}_j^{\mathrm{R}}, \tag{9}$$

for each pixel index p and channel index j, where $\boldsymbol{S}^{\mathrm{R}}(p) \in \mathbb{R}^{k \times k}$, $\boldsymbol{C}_j^{\mathrm{R}} \in \mathbb{R}^{k \times k}$ are extracted and reshaped from $\boldsymbol{S}^{\mathrm{R}} = \phi^{\mathrm{R}}(\boldsymbol{X}^{\mathrm{R}}) \in \mathbb{R}^{h \times w \times k^2}$, $\boldsymbol{C}^{\mathrm{R}} = \psi^{\mathrm{R}}(\boldsymbol{X}^{\mathrm{R}}) \in \mathbb{R}^{c \times k \times k}$, and $\odot$ denotes the element-wise product. The same composition strategy is applied to $\mathcal{W}^{\mathrm{I}}$.

The spatial filter branch $\phi^{\mathrm{R}}(\cdot)$ takes a local receptive field and predicts local-content-aware filters, while the channel filter branch

$\psi^{\mathrm{R}}(\cdot)$ takes the whole image as receptive field and predicts global-semantic-aware filters. The composition strategy leads to a locally and globally content-aware feature extraction process, with greatly reduced learnable parameters. Specifically, in our model, the spatial filter branch $\phi^{\mathrm{R}}(\cdot)$ contains one 1×1 convolution layer, and the channel filter branch $\psi^{\mathrm{R}}(\cdot)$ applies the global average pooling (GAP) to aggregate input features, then involves the stack of two fully-connected layers and an in-between ReLU, whose number of parameters are reduced to $k^2c + 2c^2$.

3.3 Score Regression

To predict a quality score from the extracted features, GAP is separately applied to the real and imaginary parts of each feature map denoted by $\boldsymbol{X}_k$:

$$z_k = \mathrm{GAP}(\boldsymbol{X}_k^{\mathrm{R}}) + i \cdot \mathrm{GAP}(\boldsymbol{X}_k^{\mathrm{I}}), \quad \forall k. \tag{10}$$

Then, a complex fully-connected layer is applied as a regressor:

$$h = \boldsymbol{w}\boldsymbol{z} = (\boldsymbol{w}^{\mathrm{R}}\boldsymbol{z}^{\mathrm{R}} - \boldsymbol{w}^{\mathrm{I}}\boldsymbol{z}^{\mathrm{I}}) + i \cdot (\boldsymbol{w}^{\mathrm{R}}\boldsymbol{z}^{\mathrm{I}} + \boldsymbol{w}^{\mathrm{I}}\boldsymbol{z}^{\mathrm{R}}), \tag{11}$$

where $\boldsymbol{w}^{\mathrm{R}}, \boldsymbol{w}^{\mathrm{I}} \in \mathbb{R}^{1\times C}$. The quality score is defined as

$$s = \sqrt{(h^R)^2 + (h^I)^2}. \tag{12}$$

3.4 Training

Due to its stronger robustness to outliers, we use Huber loss for training. Let s^* denote the subjective score of a training image measured by human and s the corresponding score predicted by the model. The training loss is defined by

$$\ell_\delta(s, s^*) = \begin{cases} \frac{1}{2}(s - s^*)^2 & \text{for } |s - s^*| \le \delta \\ \delta\left(|s - s^*| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases}. \tag{13}$$

and δ is a parameter to choose the way to penalty outliers. In implementation, we set $\delta = 1/9$ as suggested in [41]. During training, the pre-trained backbone is frozen and without fine-tuning. This leads to much faster training with less GPU memory consumed than existing fine-tuning-based methods.

4 EXPERIMENTS

Following [4, 19, 41], early layers of ResNet-101 [13] are adopted as the backbone, where the first 3 building blocks (conv1, conv2_x, conv3_x) are employed to extract low/middle-level semantic features. In DCNet, all convolution kernels are set to 3×3. In training, stochastic gradient descent is used as the optimizer, with the momentum of 0.9 and the weight decay of $1e{-}4$. The learning rate is initially set to 0.1 and decreased after error plateaus.

4.1 Experimental Setups

Six publicly available natural image quality databases are used for experimental evaluation, including (i) three artificially-distorted sets: CSIQ [18], TID2013 [35] and Kadid-10k [22]; and (ii) three realistically-distorted sets: LIVE-C [10], KonIQ-10k [15] and SPAQ [9]. See below for their details.

- CSIQ: 30 pristine images and 866 distorted images with 6 distortion types at 4 to 5 distortion levels.
- TID2013: 25 pristine images and a total of 3,000 distorted images with 17 distortion types at 4 degradation levels.
- Kadid-10k: 81 pristine images each of which is degraded by 25 distortions in 5 levels, resulting in 10,125 distorted images.
- LIVE-C: 1,162 realistically natural pictures with resized resolution of 500×500 pixels.
- KonIQ-10k: 10,073 realistically and complexly distorted images with resolution of 1024×768 pixels.
- SPAQ: 11,125 images captured by 66 smartphones with diverse resolutions.

Images in their original resolutions are used as input to test generalization to different image sizes. Following [41, 42], we randomly sample 80% of the images in each database for training and leave the rest for testing. Specifically, for synthetically-distorted datasets, we split the training and test sets according to the pristine images such that the content is not intersected between the two sets. Following [41], data augmentation comprising horizontal flip, vertical flip, and rotation of $\pm 3^\circ$ is randomly applied to the training images, which brings a small performance improvement in most cases. The excess area is removed by cropping when rotation produces extra borders. For performance comparison, the median values of evaluation metrics across ten sessions on the test sets are reported.

Two commonly-used evaluation metrics are adopted for performance comparison, including Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC). The former measures the prediction monotonicity, while the latrer measures the linear correlation. An effective IQA metric is expected to yield high values of PLCC and SROCC.

Two well-known traditional handcrafted NR-IQA models NIQE [30] and ILNIQE [51] are selected for performance comparison. Moreover, recent CNN-based methods are also selected for comparison, including PQR [50], deepIQA [2], DBCNN [54], SGDNet [48], CaHDC [45], HyperNet [40], MetaIQA [60], SiamIQA [56], AIGQA [25], UNIQUE [55] and OLNet [49]. Their experimental results are quoted from the original papers whenever possible or otherwise obtained by their codes from the authors.

4.2 Performance Evaluation

4.2.1 Evaluation on individual databases. We first conduct evaluation on individual databases including both synthetic and authentic distortions. The results on synthetically-distorted datasets (*i.e.* CSIQ, TID2013 and Kadid-10k) are reported in Table 1. It can be seen that our DCNet outperforms all the state-of-the-art methods by a large margin on all those three databases, in terms of both SROCC and PLCC metrics. The larger performance gain is observed on the TID2013 and Kadid-10k datasets which are two challenging datasets with diverse contents and complex distortions. Note that the compared methods perform well on CSIQ but show a large performance decrease on TID2013 and Kadid-10k. The performance gain of DCNet is probably because it is capable of learning more-accurate content-adaptive filters for quality-related feature extraction with a compact model, leading to better generalization on complex distortions.

The results on authentically-distorted databases (*i.e.* LIVE-C, Koniq-10k and SPAQ) are reported in Table 2. As the results indicate, our DCNet outperforms all compared methods significantly on the large-scale dataset KonIQ-10k while achieving comparative performance to UNIQUE* on LIVE-C. The images in LIVE-C are

Table 1: Performance comparison on synthetically-distorted databases. The best result for each metric is boldfaced.

Method	CSIQ		TID2013		Kadid-10k	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [30]	0.632	0.726	0.343	0.378	0.374	0.428
ILNIQE [51]	0.832	0.873	0.570	0.598	0.531	0.563
PQR [50]	0.872	0.901	0.745	0.798	-	-
deepIQA [2]	0.955	0.973	0.761	0.787	0.628	0.647
DBCNN [54]	0.946	0.959	0.816	0.865	-	-
SGDNet [48]	0.883	0.903	0.843	0.861	-	-
MetaIQA [60]	-	-	-	-	0.767	0.774
CaHDC [45]	0.874	0.915	0.862	0.878	-	-
SiamIQA [56]	0.962	-	0.855	-	0.913	-
AIGQA [25]	0.927	0.952	0.871	0.893	0.864	0.863
UNIQUE [55]	0.902	0.927	0.855	0.879	0.876	0.878
OLNet [49]	0.966	0.975	0.863	0.889	-	-
DCNet [Ours]	**0.974**	**0.978**	**0.952**	**0.954**	**0.922**	**0.931**

Table 2: Performance comparison on authentically-distorted databases. The best result for each metric is boldfaced.

Method	LIVE-C		KonIQ-10k		SPAQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [30]	0.464	0.515	0.601	0.597	0.703	0.712
ILNIQE [51]	0.469	0.536	0.552	0.573	0.714	0.721
PQR [50]	0.857	0.882	0.881	0.884	-	-
deepIQA [2]	0.671	0.686	0.797	0.805	-	-
DBCNN [54]	0.851	0.869	0.875	0.884	0.911	0.915
SGDNet [48]	0.851	0.872	0.897	0.917	-	-
MetaIQA [60]	0.802	0.835	0.851	0.887	0.875	0.877
CaHDC [45]	0.738	0.744	-	-	0.827	0.834
HyperNet [40]	0.859	0.882	0.906	0.917	0.916	0.918
SiamIQA [56]	0.851	-	0.894	-	-	-
AIGQA [25]	0.751	0.761	-	-	-	-
UNIQUE* [55]	0.854	**0.890**	0.896	0.901	-	-
OLNet [49]	0.849	0.858	0.877	0.882	-	-
DCNet [Ours]	**0.860**	0.881	**0.910**	**0.924**	**0.919**	**0.920**

originally resized to a uniform square size, which is inconsistent with the sizes when capturing them. Thus, the pixel distribution of different objects is disrupted differently for each image, probably resulting in the generalization problem. Fortunately, content-adaptive models such as HyperNet [40] and ours still perform well on it. Note that UNIQUE* [55] is marked because it uses all datasets for training, while ours use individual ones respectively. Even that, DCNet still competes against it. The good performance of DCNet on authentic data probably comes from the complex-valued content-aware filtering performed by the complex dynamic convolution module.

4.2.2 Cross-dataset evaluation. To further evaluate the generalization ability of DCNet, we also conduct a cross-dataset experiment by using the KonIQ-10k as the training set and LIVE-C/SPAQ as the test sets. Only compared methods with available results are selected for comparison. See Table 3 for the results, which indicate that our DCNet has good generalization for predicting the quality score for the images that have an arbitrary resolution, varying image contents, and real-world complicated distortions.

Table 3: Performance comparison on models trained on KonIQ-10k and tested on LIVE-C/SPAQ without fine-tuning. The best result under each setting is boldfaced.

LIVE-C	PQR	DBCNN	HyperNet	UNIQUE	**DCNet**
SROCC	0.772	0.755	0.785	0.786	**0.805**
PLCC	0.817	-	0.818	-	**0.827**
SPAQ	DBCNN	CaHDC	HyperNet	**DCNet**	
SROCC	0.783	0.730	0.807	**0.811**	
PLCC	0.792	0.778	0.818	**0.819**	

Following [44], a gMAD competition [27] is conducted on the SPAQ [9] database for direct visualization. The gMAD aims to efficiently select image pairs with maximum quality difference predicted by an attacking IQA model to challenge another defending model which partitions them to the same level of quality. The selected pairs are shown to the observer to determine whether the attacker or the defender is robust. In the gMAD competition, DCNet competes with the best CNN-based competitor (*i.e.* HyperNet [9]) and the best traditional competitor (*i.e.* ILNIQE [51]) in cross-dataset evaluation. The representative gMAD pairs in SPAQ are shown in Figure 3 and 4, where three models are trained on Koniq-10k. From Figures 3(a) and 3(b), it can be observed that the perceptual quality of images in the first row is slightly better on clear structures than those in the second row, indicating that the DCNet correctly attacked HyperNet. When we fix DCNet as the defender, it successfully survived the attack of HyperNet at the low and good quality levels. See Figure 3(c) and 3(d) for illustration. In addition, it is worth mentioning that both DCNet and hyperNet recognize the obvious low-quality images successfully. The similar observation can be found in Figures 4.

4.2.3 Comparison on diverse distortion types. In order to evaluate the performance on diverse distortion types, we collect the results of CNN-based models on diverse categories of TID2013 and CSIQ in Table 4 and Table 5. As can be seen, DCNet shows competitive performances on individual distortion types to other methods. It achieved the best evaluation performance for 17 out of 24 types on TID2013 and all on CSIQ.

4.3 Ablation Studies

To verify the effectiveness of the complex-valued architecture and the dynamic filtering module of DCNet, we conduct ablation studies on the CSIQ, LIVE-C and Koniq-10k databases in terms of SROCCs under several settings, using the following several baselines of DCNet. (i) Baseline-1: using only the pre-trained backbone without fine-tuning. (ii) Baseline-2: using the pre-trained model with fine-tuning on IQA datasets. (iii) RealNet: Replacing complex-valued operations in DCNet with real-valued ones and doubling the channels. (iv) StaticNet: Removing the complex dynamic filter module in DCNet. Note that the FC layer is used and trained for the first two baselines for score regression. The backbones in both RealNet and StaticNet are not fine-tuned.

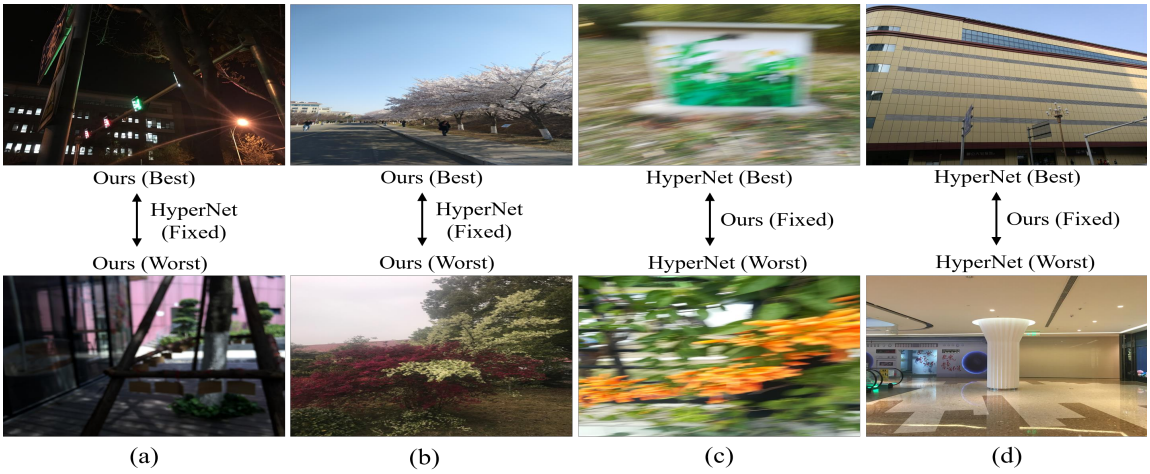


Figure 3: Representative gMAD pairs between DCNet and HyperNet on the SPAQ database. (a) Fixing HyperNet at low quality level. (b) Fixing HyperNet at high quality level. (c) Fixing DCNet at low quality level. (d) Fixing DCNet at high quality level.

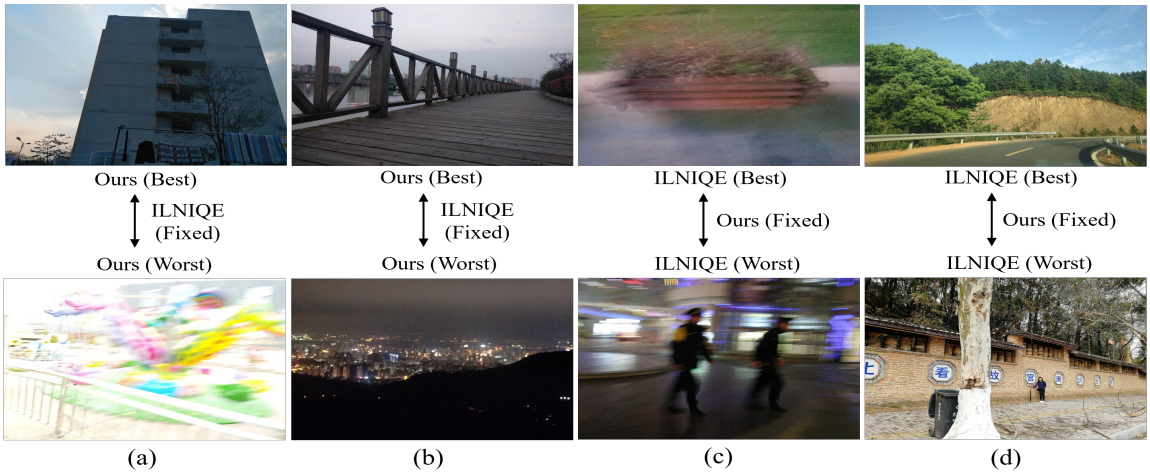


Figure 4: Representative gMAD pairs between DCNet and ILNIQE on the SPAQ database. (a) Fixing ILNIQE at low quality level. (b) Fixing ILNIQE at high quality level. (c) Fixing DCNet at low quality level. (d) Fixing DCNet at high quality level.

The results are given in Table 6. As can be seen, each component has a noticeable contribution to the performance of the DCNet: (i) Compared to Baseline-1, the SROCC gains of RealNet and StaticNet are over 0.02 and 0.03 respectively, through the datasets. Such improvements demonstrated the effectiveness of the complex-valued architecture and dynamic filtering of DCNet, respectively. (ii) Compared to Baseline-1, the original DCNet shows 0.055 SROCC gain on average, which demonstrates its effectiveness. Compared with RealNet and StaticNet, DCNet still achieved approximately 0.02 SROCC additional gain on average. Such improvement comes from the combination of the complex-valued architecture and dynamic filtering over the individual ones. (iii) Compared to Baseline-2, DCNet achieved better results through all databases, which again demonstrated its effectiveness. (iv) DCNet with backbone fine-tuning achieved even further improvement on two datasets, but showed a certain performance decrease on the LIVE-C database which is probably due to overfitting. Indeed, DCNet already performs very well without the time-consuming backbone fine-tuning.

4.4 More Analysis

4.4.1 Model complexity. The model complexity is evaluated in terms of floating-point operations per second (FLOPS) in processing a color image of size 1024×768. The recent methods including SGDNet, CaHDC and HyperNet are selected for comparison, using their released codes implemented by Python. See Table 7 for the results. Our DCNet is comparable with HyperNet, one of the best competitors in most experiments, in terms of FLOPS.

4.4.2 Performance with different backbones. Using different backbone models may influence the performance of the DCNet. Table 8 lists the results of DCNet using ResNet152, ResNet101, ResNet50 and VGG16 as the backbone, respectively. Comparing these results with those of previous experiments in Table 1 and 2, we can see that DCNet using different ResNet backbones still shows state-of-the-art

Table 4: SROCC results on diverse distortion types of TID2013. The best result for each distortion type is boldfaced.

Distortion	DBCNN	MetaIQA	HyperNet	AIGQA	**DCNet**
AGN	0.790	0.947	0.769	0.932	**0.956**
ANC	0.700	0.924	0.613	0.916	**0.929**
SCN	0.826	0.955	0.918	0.944	**0.961**
MN	0.646	0.728	0.448	0.662	**0.784**
HFN	0.879	0.952	0.839	0.953	**0.968**
IN	0.708	0.866	0.758	**0.911**	0.906
QN	0.825	0.745	0.828	0.908	**0.909**
GB	0.859	**0.977**	0.873	0.917	0.974
DEN	0.865	0.938	0.804	0.914	**0.955**
JPEG	0.894	0.934	0.860	0.945	**0.948**
JP2K	0.916	0.957	0.888	0.932	**0.960**
JGTE	0.772	**0.931**	0.723	0.858	0.922
J2TE	0.822	0.903	0.846	0.898	**0.919**
NEPN	0.270	0.729	0.369	0.130	**0.747**
Block	0.444	0.391	0.428	**0.723**	0.711
MS	-0.009	0.402	0.424	0.554	**0.629**
CTC	0.548	0.764	0.740	0.830	**0.850**
CCS	0.631	0.829	0.710	0.689	**0.841**
MGN	0.711	0.939	0.767	0.948	**0.954**
CN	0.752	**0.952**	0.786	0.886	0.951
LCNI	0.860	0.978	0.879	0.897	**0.983**
ICQD	0.833	0.859	0.785	**0.908**	0.854
CHA	0.732	0.927	0.739	0.889	**0.931**
SSR	0.902	**0.974**	0.910	0.908	0.967

Table 5: SROCC results on diverse distortion types of CSIQ. Bold on digits denote the best result for each distortion type.

Distortion	PQR	DBCNN	HyperNet	OLNet	**DCNet**
GB	0.921	0.947	0.915	0.965	**0.968**
AWGN	0.915	0.948	0.927	0.945	**0.964**
JPEG	0.934	0.940	0.934	0.968	**0.972**
JP2K	0.955	0.953	0.960	0.945	**0.966**
APN	0.926	0.941	0.931	0.953	**0.958**
CTD	0.837	0.872	0.874	0.925	**0.931**

Table 6: Ablation study on three IQA databases in terms of SROCC. The best SROCC result for each dataset is boldfaced.

Model	TID2013	LIVE-C	KonIQ-10k
Baseline-1	0.881	0.803	0.872
Baseline-2	0.937	0.832	0.909
RealNet	0.924	0.823	0.907
StaticNet	0.932	0.840	0.902
DCNet	0.952	**0.860**	0.910
DCNet + Finetune backbone	**0.961**	0.835	**0.929**

Table 7: Comparison of model complexity in terms of FLOPS

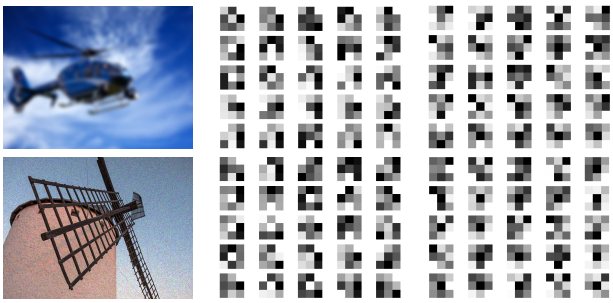
SGDNet	CaHDC	HyperNet	**DCNet**
1.23×10^{11}	0.37×10^{11}	1.98×10^{11}	1.73×10^{11}

performance. Using a larger backbone such as ResNet152 may lead to further improvement in some cases.

Table 8: Performance on different backbones

Backbones	TID2013		KonIQ-10k	
	SROCC	PLCC	SROCC	PLCC
ResNet152	0.954	0.946	0.901	0.914
ResNet101 (Ours)	0.952	0.954	0.910	0.924
ResNet50	0.942	0.931	0.908	0.919
VGG16	0.901	0.891	0.809	0.839

4.4.3 Visualization of learned filters. We show the 3×3 filters output by the complex dynamic filtering module on two distorted images in Figure 5.



(a) Image (b) Real part (c) Imaginary part

Figure 5: Examples of filters output by the complex dynamic filtering module.

5 CONCLUSION

This paper improved the effectiveness of CNNs for NR-IQA by introducing a complex-valued architecture and a dynamic filtering module, based on which a complex-valued dynamic neural NR-IQA model is developed. Such a model has its key operations defined in a complex field, so as to exploit the merits of complex-valued transforms in the IQA realm. The dynamic filtering module further improves the content-awareness during the feature extraction process in the CNN. Extensive experiments on six publicly datasets have demonstrated the superior performance of the proposed model over the state-of-the-art ones in handling both synthetically-distorted images and authentically-distorted images. In future, we would like to further investigate other merits of complex-valued CNN modules for the improvement of NR-IQA, as well extend the proposed approach to other IQA tasks, such as full-reference or reduced-reference IQA. In addition, the connection between complex-valued convolution and dynamic filtering will be further studied.

ACKNOWLEDGMENTS

The authors would like to acknowledge the partial support from National Natural Science Foundation of China under Grants 62072188, 61872151 and 62106077, Science and Technology Program of Guangdong Province under Grants 2022A1515011087, 2022A1515011755 and 2020A1515011128, and Postdoctoral Foundation of China under Grant 2020M682705.

REFERENCES

[1] Martin Arjovsky, Amar Shah, and Yoshua Bengio. 2016. Unitary evolution recurrent neural networks. In *Proc. Intel. Conf. Machine Learning*. 1120–1128.

[2] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. 2018. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* 27, 1 (2018), 206–219.

[3] Diqi Chen, Yizhou Wang, and Wen Gao. 2020. No-reference image quality assessment: An attention driven approach. *IEEE Trans. Image Process.* 29 (2020), 6496–6506.

[4] Aladine Chetouani. 2020. Image quality assessment without reference by mixing deep learning-based features. In *Proc. IEEE Conf. Multimedia Expo.* 1–6.

[5] Soumith Chintala, Marc'Aurelio Ranzato, Arthur Szlam, Yuandong Tian, Mark Tygert, and Wojciech Zaremba. 2017. Scale-invariant learning and convolutional networks. *Appl. Comput. Harmonic Anal.* 42, 1 (2017), 154–166.

[6] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 1251–1258.

[7] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. 2017. Deformable convolutional networks. In *Proc. IEEE Conf. Comput. Vision*. 764–773.

[8] B M Dow. 1974. Functional Classes of Cells and Their Laminar Distribution in Monkey Visual Cortex. *J. Neurophysiology* 37, 5 (1974), 927–946.

[9] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. 2020. Perceptual quality assessment of smartphone photography. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 3677–3686.

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

[11] Deepti Ghadiyaram and Alan C Bovik. 2017. Perceptual quality prediction on authentically distorted images using a bag of features approach. *J. vision* 17, 1 (2017), 32–32.

[12] Jie Gu, Gaofeng Meng, Shiming Xiang, and Chunhong Pan. 2019. Blind image quality assessment via Learnable Attention-Based Pooling. *Pattern Recognition* 91 (2019), 332–344.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 770–778.

[14] Akira Hirose and Shotaro Yoshida. 2012. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Trans. Neural Netw. Learning Syst.* 23, 4 (2012), 541–551.

[15] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. 2020. KonIQ-10k: An ecologically Valid database for deep learning of blind image quality assessment. *IEEE Trans. Image Process.* 29 (2020), 4041–4056.

[16] Parimala Kancharla and Sumohana S Channappayya. 2021. Completely blind quality assessment of user generated video content. *IEEE Trans. Image Process.* 31 (2021), 263–274.

[17] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. 2021. Musiq: Multi-scale image quality transformer. In *Proc. IEEE Conf. Comput. Vision*. 5148–5157.

[18] Eric Cooper Larson and Damon Michael Chandler. 2010. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Elec. Imaging* 19, 1 (2010), 011006.

[19] Dingquan Li, Tingting Jiang, and Ming Jiang. 2020. Norm-in-norm loss with faster convergence and better performance for image quality assessment. In *Proc. ACM Intel. Conf. Multimedia*. 789–797.

[20] Dingquan Li, Tingting Jiang, Weisi Lin, and Ming Jiang. 2019. Which has better visual quality: The clear blue sky or a blurry animal? *IEEE Trans. Multimedia* 21, 5 (2019), 1221–1234.

[21] Qiaohong Li, Weisi Lin, and Yuming Fang. 2016. No-Reference Quality assessment for Multiply-Distorted Images in Gradient Domain. *IEEE Signal Process. Letters* 23, 4 (2016), 541–545.

[22] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. 2019. KADID-10k: A large-scale artificially distorted IQA database. In *Proc. IEEE Conf. Quality Multimedia Exp.* 1–3.

[23] Delei Liu, Yong Xu, Yuhui Quan, and Patrick Le Callet. 2014. Reduced reference image quality assessment using regularity of phase congruency. *Signal Processing: Image Communication* 29, 8 (2014), 844–855.

[24] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. 2017. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 1040–1049.

[25] Jupo Ma, Jinjian Wu, Leida Li, Weisheng Dong, Xuemei Xie, Guangming Shi, and Weisi Lin. 2021. Blind image quality assessment with active inference. *IEEE Trans. Image Process.* 30 (2021), 3650–3663.

[26] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. 2017. End-to-end blind image quality assessment using deep neural networks. *IEEE Trans. Image Process.* 27, 3 (2017), 1202–1213.

[27] Kede Ma, Qingbo Wu, Zhou Wang, Zhengfang Duanmu, Hongwei Yong, Hongliang Li, and Lei Zhang. 2016. Group mad competition-a new methodology to compare objective image quality models. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 1664–1673.

[28] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. 2018. Burst denoising with kernel prediction networks. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 2502–2510.

[29] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-Reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* 21, 12 (2012), 4695–4708.

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

[31] Manish Narwaria, Weisi Lin, Ian Vince McLoughlin, Sabu Emmanuel, and Liang-Tien Chia. 2012. Fourier transform-based scalable image quality measure. *IEEE Trans. Image Process.* 21, 8 (2012), 3364–3377.

[32] Tohru Nitta. 2002. On the critical points of the complex-valued neural network. In *Proc. Int. Conf. Neural Info. Process.*, Vol. 3. IEEE, 1099–1103.

[33] Da Pan, Ping Shi, Ming Hou, Zefeng Ying, Sizhe Fu, and Yuan Zhang. 2018. Blind predicting similar quality map for image quality assessment. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 6373–6382.

[34] Fidel Alejandro Guerrero Peña, Pedro Diamel Marrero Fernández, Tsang Ing Ren, Jorge de Jesus Gomes Leandro, and Ricardo Massahiro Nishihara. 2019. Burst ranking for blind multi-image deblurring. *IEEE Trans. Image Process.* 29 (2019), 947–958.

[35] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Comm.* 30 (2015), 57–77.

[36] Yuhui Quan, Yixin Chen, Yizhen Shao, Huan Teng, Yong Xu, and Hui Ji. 2021. Image denoising using complex-valued deep CNN. *Pattern Recognition* 111 (2021).

[37] Yuhui Quan, Peikang Lin, Yong Xu, Yuesong Nan, and Hui Ji. 2021. Nonblind image deblurring via deep learning in complex field. *IEEE Trans. Neural Netw. Learning Syst.* pp, 99 (2021), 1–14.

[38] Yuhui Quan, Zicong Wu, and Hui Ji. 2021. Gaussian Kernel Mixture Network for Single Image Defocus Deblurring. *Proc. Adv. Neural Inf. Process. Syst.* 34 (2021), 20812–20824.

[39] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint: 1409.1556* (2014).

[40] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. 2020. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. 3664–3673.

[41] Yicheng Su and Jari Korhonen. 2020. Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling. In *Proc. Int. Conf. Image Process.* 191–195.

[42] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE Trans. Image Process.* 27, 8 (2018), 3998–4011.

[43] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. 2017. Deep complex networks. *arXiv preprint:1705.09792* (2017).

[44] Zhihua Wang, Zhi-Ri Tang, Jianguo Zhang, and Yuming Fang. 2021. Learning from synthetic data for opinion-free blind image quality assessment in the wild. *arXiv preprint: 2106.14076* (2021).

[45] Jinjian Wu, Jupo Ma, Fuhu Liang, Weisheng Dong, Guangming Shi, and Weisi Lin. 2020. End-to-end blind image quality prediction with cascaded deep neural network. *IEEE Trans. Image Process.* 29 (2020), 7414–7426.

[46] Yong Xu, Delei Liu, Yuhui Quan, and Patrick Le Callet. 2015. Fractal analysis for reduced reference image quality assessment. *IEEE Trans. Image Process.* 24, 7 (2015), 2098–2109.

[47] Brandon Yang, Gabriel Bender, Quoc V Le, and Jiquan Ngiam. 2019. Condconv: Conditionally parameterized convolutions for efficient inference. In *Proc. Adv. Neural Inf. Process. Syst.*, Vol. 32.

[48] Sheng Yang, Qiuping Jiang, Weisi Lin, and Yongtao Wang. 2019. Sgdnet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment. In *Proc. ACM Intel. Conf. Multimedia*. 1383–1391.

[49] Xiwen Yao, Qinglong Cao, Xiaoxu Feng, Gong Cheng, and Junwei Han. 2022. Learning to assess image quality like an observer. *IEEE Trans. Neural Netw. Learning Syst.* (2022).

[50] Hui Zeng, Lei Zhang, and Alan C Bovik. 2017. A probabilistic quality representation approach to deep blind image quality prediction. *arXiv preprint: 1708.08190* (2017).

[51] Lin Zhang, Lei Zhang, and Alan C Bovik. 2015. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* 24, 8 (2015), 2579–2591.

[52] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. 2011. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 20, 8 (2011), 2378–2386.

[53] Min Zhang, Chisako Muramatsu, Xiangrong Zhou, Takeshi Hara, and Hiroshi Fujita. 2014. Blind image quality assessment Using the Joint Statistics of Generalized

Local Binary Pattern. *IEEE Signal Process. Letters* 22, 2 (2014), 207–210.

[54] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang. 2018. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Trans. Circuits Syst. Video Technol.* 30, 1 (2018), 36–47.

[55] Weixia Zhang, Kede Ma, Guangtao Zhai, and Xiaokang Yang. 2021. Uncertainty-aware blind image quality assessment in the laboratory and wild. *IEEE Trans. Image Process.* 30 (2021).

[56] Weixia Zhang, Kede Zhai, Guangtao Zhai, and Xiaokang Yang. 2020. Learning to blindly assess image quality in the laboratory and wild. In *Proc. IEEE Intel. Conf. Image Process.* 111–115.

[57] Yikang Zhang, Jian Zhang, Qiang Wang, and Zhao Zhong. 2020. Dynet: Dynamic convolution for accelerating convolutional neural networks. *arXiv preprint: 2004.10694* (2020).

[58] Xiaochun Zhong, Chaofeng Li, Wei Zhang, and Yiwen Ju. 2014. No-reference image quality assessment using dual-tree complex wavelet transform. In *Proc. Intel. Congress Image Signal Process.* IEEE, 596–601.

[59] Jingkai Zhou, Varun Jampani, Zhixiong Pi, Qiong Liu, and Ming-Hsuan Yang. 2021. Decoupled dynamic filter networks. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition.* 6647–6656.

[60] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. 2020. MetaIQA: Deep meta-learning for no-reference image quality assessment. In *Proc. IEEE Conf. Comput. Vision Pattern Recognition.* 14143–14152.