Blind visual quality assessment of light field images based on distortion maps

Sana Alamgeer* and Mylène C. Q. Farias*

[1]Department of Electrical Engineering, University of Brasília, Brasília, Brazil

Light Field (LF) cameras capture spatial and angular information of a scene, generating a high-dimensional data that brings several challenges to compression, transmission, and reconstruction algorithms. One research area that has been attracting a lot of attention is the design of Light Field images quality assessment (LF-IQA) methods. In this paper, we propose a No-Reference (NR) LF-IQA method that is based on reference-free distortion maps. With this goal, we first generate a synthetically distorted dataset of 2D images. Then, we compute SSIM distortion maps of these images and use these maps as ground error maps. We train a GAN architecture using these SSIM distortion maps as quality labels. This trained model is used to generate reference-free distortion maps of sub-aperture images of LF contents. Finally, the quality prediction is obtained performing the following steps: *1)* perform a non-linear dimensionality reduction with a isometric mapping of the generated distortion maps to obtain the LFI feature vectors and *2)* perform a regression using a Random Forest Regressor (RFR) algorithm to obtain the LF quality estimates. Results show that the proposed method is robust and accurate, outperforming several state-of-the-art LF-IQA methods.

KEYWORDS

image quality assessment, epipolar planes, canny edge detector, two-stream convolution neural network, 4D light field images

1 Introduction

Unlike conventional cameras, Light Field (LF) camera captures spatial and angular information of scene, which is represented by a scalar function $L(u, v, s, t)$, where (u, v) and (s, t) depict the angular and spatial domains, respectively. The 4D light field can be described as a 2D projection of sub-aperture images (SAIs). Figure 1 illustrates a grid of 10×10 sub-aperture images of LFI (ArtGallery3) from the MPI dataset (Adhikarla et al., 2017). SAIs are generated from micro-lens images, with an operation known as raw data decoding. LF images (LFI) carry rich information that is widely used for refocusing (Hahne et al., 2018) and 3-Dimensional (3D) reconstruction. However, the high-dimensionality of LFIs creates several challenges to the area of communications, requiring the development of specific compression (Hou et al., 2019), transmission, and reconstruction techniques. Unfortunately, these techniques inevitably distort the perceived quality of LFIs (Paudyal et al., 2017). In order to monitor the visual quality and
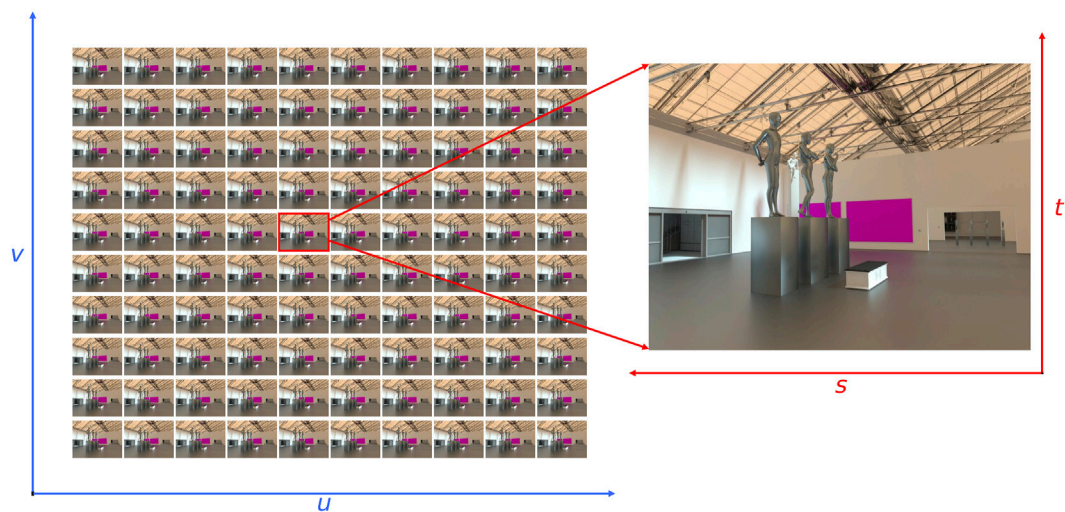
FIGURE 1
A grid of 10 × 10 sub-aperture images of a Light Field image (ArtGallery3) from MPI dataset (Adhikarla et al., 2017).

quantify the amount of visual distortion in LF contents, we must develop efficient LFI Quality Assessment (LF-IQA) methods.

The visual quality of images can be assessed (IQA) subjectively or objectively. Subjective quality assessment methods are experimental methodologies in which human observers are asked to estimate one or more features of the stimuli (e.g., the overall quality of a video). The data collected in these experiments are statistically analyzed, often generating Mean Opinion Scores (MOS) for each of the test stimuli. Although subjective quality assessment methods are considered as the most reliable ways of estimating quality, these methods are time-consuming and cannot be implemented in real-time systems. Objective quality assessment methods (also known as quality metrics) are algorithms that automatically assess the quality of a content by measuring the physical signal. Based on the available reference information, objective quality assessment methods are divided into the full-reference (FR), reduced-reference (RR) and blind/no-reference (NR) methods. To estimate the performance of objective quality methods, we compare their output scores with MOS values obtained using subjective methods.

As mentioned before, LFIs contain not only spatial information, but also angular information. Therefore, classical 2D image quality assessment methods cannot be directly used for LFI quality assessment. In the past few years, efforts have been devoted to develop LF-IQA methods. For example, Tian *et al.* (Tian et al., 2018) proposed a FR LF-IQA metric that uses a multi-order derivative feature-based method (MDFM), which extracts detailed features with different degrees with a discrete derivative filter. Paudyal *et al.* (Paudyal et al., 2019) proposed a RR LF-IQA that uses two IQA methods - SSIM (Wang et al., 2004) and PSNR - to process the depth maps. Fang *et al.* (Fang

et al., 2018) presented a FR LF-IQA method that uses local and global features to predict quality. The local features are extracted from reference and test LFIs using a gradient magnitude operator, while the global features are extracted from the epipolar-plane images (EPIs) of reference and test LFIs using the same operator. Tian *et al.* (Tian et al., 2020b) proposed a FR LF-IQA method that uses symmetry information, which are extracted using Gabor filters (Field, 1987), and depth features computed from EPIs of the reference and test LFIs. Meng *et al.* (Meng et al., 2020) presented a FR LF-IQA method that computes the spatial LFI quality using the structural similarity index metric (SSIM) of the Difference of Gaussian (DoG) features of central SAIs and computes the angular LFI quality using the SSIM of refocused images.

Tian *et al.* (Tian et al., 2020a) presented a FR LF-IQA method in which the salient features are extracted from reference and test EPIs and SAIs using single-scale and multi-scale log-Gabor operators. The NR LFQA method proposed by Shi *et al.* (Shi et al., 2019a) predicts quality using EPI information and natural statistics. The NR LF-IQA method proposed by Luo *et al.* (Luo et al., 2019) employs the spatial information from SAIs and the angular information from the micro-lens images. Jiang *et al.* (Jiang et al., 2018) proposed a FR LF-IQA method that uses the entropy information and the gradient magnitude features to extract spacial features from the SAIs. To extract the angular features from SAIs, dense distortion curves are generated and the best fitting features are chosen. Wei Zhou *et al.* (Shi et al., 2019b; Zhou et al., 2019) proposed NR LF-IQA methods (BELIF and Tensor-NLFQ, respectively) that are based on tensor theory, employing SAIs view stacks along horizontal, vertical, left diagonal, and right diagonal orientations. The local spatial quality features are extracted using local frequency
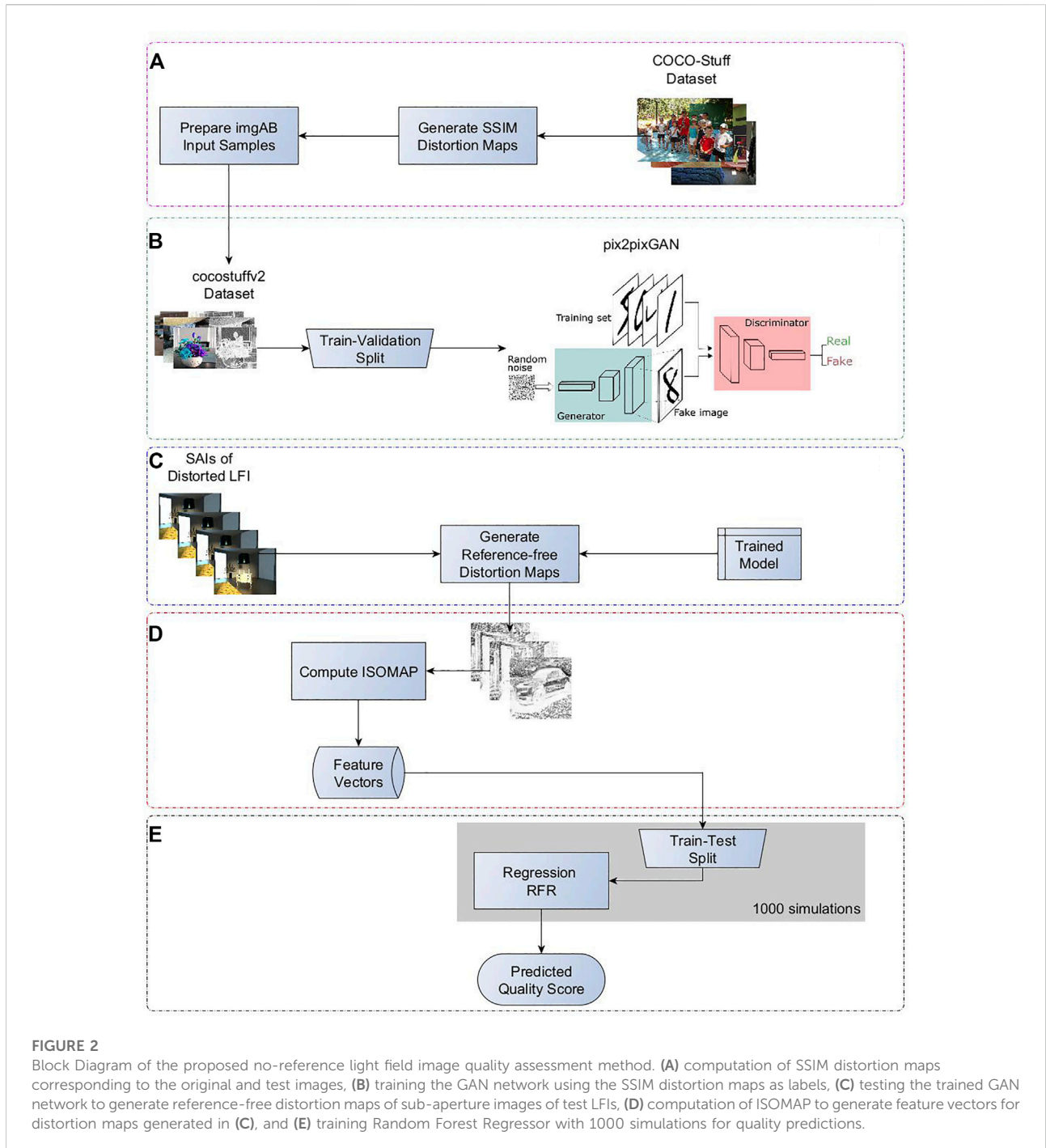
FIGURE 2
Block Diagram of the proposed no-reference light field image quality assessment method. **(A)** computation of SSIM distortion maps corresponding to the original and test images, **(B)** training the GAN network using the SSIM distortion maps as labels, **(C)** testing the trained GAN network to generate reference-free distortion maps of sub-aperture images of test LFIs, **(D)** computation of ISOMAP to generate feature vectors for distortion maps generated in **(C)**, and **(E)** training Random Forest Regressor with 1000 simulations for quality predictions.

distribution, while the global spatial quality features are extracted using the naturalness distribution of individual color channels. Ak et al. (Ak et al., 2020) proposed a NR LF-IQA method based on the structural representations of EPIs, by training a convolutional sparse coding codebook and a Bag Of World dictionary on EPIs. Shan et al. (Shan et al., 2019) designed a NR LF-IQA method that is based on 2D (from SAIs) and 3D (from EPIs) LFI features. Xiang et al. (Xiang et al., 2020)

presented a NR LF-IQA method (VBLFI) based on the mean difference image and on curvelet-transform characteristics of LFIs.

Despite of the work mentioned above, there is a lot of room for improvement in terms of prediction accuracy, robustness, computational complexities, and generality of NR LF-IQA methods. In this paper, we propose a NR LF-IQA method that is based on reference-free distortion maps. Considering

FIGURE 3
Sample images taken from COCO-Stuff dataset (Caesar et al., 2018).

that pixel distortions are affected by neighboring pixels, we have focused on designing a blind deep learning quality model using pixel-by-pixel distortion maps. In summary, we present two main contributions: *1)* generation of reference-free distortion maps, and *2)* NR LF-IQA method that is derived from the generated distortion maps.

To generate LF reference-free distortion maps, we have used a deep-learning architecture called Generative Adversarial Network (GAN) network (Goodfellow et al., 2014) that can learn from synthetically generated distorted images and their corresponding ground truth distortion maps. Since ground truth distortion maps are not available in any of the existing IQA datasets (Sheikh et al., 2006), we use distortion maps generated by SSIM (Wang et al., 2004) as ground-truth distortion maps to train the GAN. Specifically, first we generate a synthetically distorted dataset of 2D images (because the sub-aperture images are 2D representation of LFIs) and, then, we compute SSIM distortion maps corresponding to the original and test images, as shown in Figure 2A. Then, we train the GAN network using the SSIM distortion maps as labels, as shown in Figure 2B. The trained model is used to generate reference-free distortion maps of sub-aperture images of test LFIs, as shown in Figure 2C. The generated distortion maps (GDMs) are used as measurement maps for describing the test LFIs. Results show that the proposed method outperforms other state-of-the-art LF-IQA methods.

The rest of the paper is organized as follows. Section 2 describes the proposed LF-IQA method. Section 3 describes

the experimental results. Finally, Section 5 presents our conclusion.

2 Proposed methodology

Generative Adversarial Networks (GAN) consists of a pair of competing network structures called generator (G) and discriminator (D) respectively, which can learn deep features with sufficient labeled training data. In this work, to learn the features of distorted images, we use the Pix2PixGAN architecture (Isola et al., 2017) for the GAN architecture because of its strong fitting capability. The Pix2PixGAN is composed of promising approach for many image-to-image translation tasks, especially those involving highly structured graphical outputs. Most importantly, the Pix2PixGAN is general-purpose, i.e., it learns a loss adapted to the task and data at hand, which makes it feasible in a wide variety of settings.

Since a GAN architecture requires a large number of training samples and LF-IQA datasets do not have a large number of samples, we use the COCO-Stuff dataset (Caesar et al., 2018) to generate a synthetically distorted dataset. The COCO-Stuff dataset is derived from the COCO dataset (Lin et al., 2014). This dataset has 1.2 million images captured from diverse scenes, with a total of 182 semantic classes. Sample images of COCO-Stuff dataset are shown in Figure 3.

To generate synthetic distorted versions of 1.2 million images, we used the Albumentation library (Buslaev et al., 2020). In total,

TABLE 1 Distortions used from Albumentation Library.

Number	Function	Description
1	VerticalFlip	Flip the input vertically around the x-axis
2	HorizontalFlip	Flip the input horizontally around the y-axis
3	IAAPerspective	Apply random four point perspective transformations to images
4	RandomRotate90	Randomly rotate the input by 90°
5	Transpose	Transpose the input by swapping rows and columns
6	ShiftScaleRotate	Randomly apply affine transforms: translate, scale and rotate the input
7	Blur	Blur the input image using a random-sized kernel
8	OpticalDistortion	Image magnification decreases with distance from the optical axis. Straight lines appear to bend outwards from the center of the image
9	GridDistortion	Grid-distortion is an image warping technique which is driven by the mapping between equivalent families of curves, arranged in a grid structure Arad (1998)
10	HueSaturationValue	Randomly change hue, saturation and value of the input image
11	IAAAdditiveGaussianNoise	Apply additive gaussian noise to the input image
12	GaussNoise	Apply gaussian noise to the input image
13	MotionBlur	Apply motion blur to the input image using a random-sized kernel
14	MedianBlur	Blur the input image using a median filter with a random aperture linear size
15	IAAPiecewiseAffine	Place a regular grid of points on each image and then randomly move each point around by 1–5 percent with respect to the image height and width
16	IAASharpen	Sharpen the input image and overlays the result with the original image
17	IAAEmboss	Emboss the input image and overlays the result with the original image
18	RandomContrast	Adjust the contrast of an image or images by a random factor
19	RandomBrightness	Randomly change brightness of the input image
20	Flip	Flip the input vertically around the x-axis
21	strong_aug_oneOfs	Custom function combined of distortions IAAAdditiveGaussianNoise, GaussNoise, MotionBlur, Blur, OpticalDistortion, GridDistortion, IAAPiecewiseAffine, CLAHE, IAASharpen, IAAEmboss, RandomContrast, RandomBrightness and HueSaturationValue
22	augment_flips_color	Custom function combined of distortions CLAHE, RandomRotate90, Transpose, ShiftScaleRotate, Blur, OpticalDistortion, GridDistortion and HueSaturationValue
23	RGBShift	Randomly shift values for each channel of the input RGB image
24	JpegCompression	Decrease Jpeg compression of an image
25	ToGray	Convert the input RGB image to grayscale. If the mean pixel value for the resulting image is greater than 127, invert the resulting grayscale image
26	RandomGamma	Draw samples from a Gamma distribution
27	InvertImg	Invert the input image by subtracting pixel values from 255
28	ChannelShuffle	Randomly rearrange channels of the input RGB image
29	CLAHE	Apply Contrast Limited Adaptive Histogram Equalization to the input image

we used 29 augmentation functions with pre-set parameters, which are depicted in Table 1. In total, we generated a dataset with 3.57 million synthetically distorted images, which we named cocostuffv2. Then, we computed SSIM distortion maps between each reference and test images in the cocostuffv2 dataset. For training the Pix2PixGAN architecture, we prepared an input tuple *imgAB* of size 512 × 256, which consists of a concatenation of the test image *A* and the SSIM distortion map *B*. Figure 4 shows some examples of input tuples *imgAB*.

For training the Pix2PixGAN network, we divided the cocstuffv2 dataset into two content-independent training and validation subsets, i.e. distorted images generated from one reference in the test subset are not present in the training

subset and vice-versa. We define a group of scenes as a group containing the reference LFI and its corresponding test versions. Then, 80% of the groups were randomly selected for training and the remaining 20% were used for validation. It is worth mentioning that we trained the Pix2PixGAN network from scratch (instead of using pre-trained model) with 50 epochs.

Next, the trained Pix2PixGAN network is used to generate reference-free distortion maps of the sub-aperture images of corresponding test LFIs. Figure 5 illustrates examples of generated distortion maps of central LF SAIs taken from the MPI dataset (Adhikarla et al., 2017). Even though we have not used any of LFIs in the training process, the Pix2PixGAN network is able to localize distortions in test SAIs. As
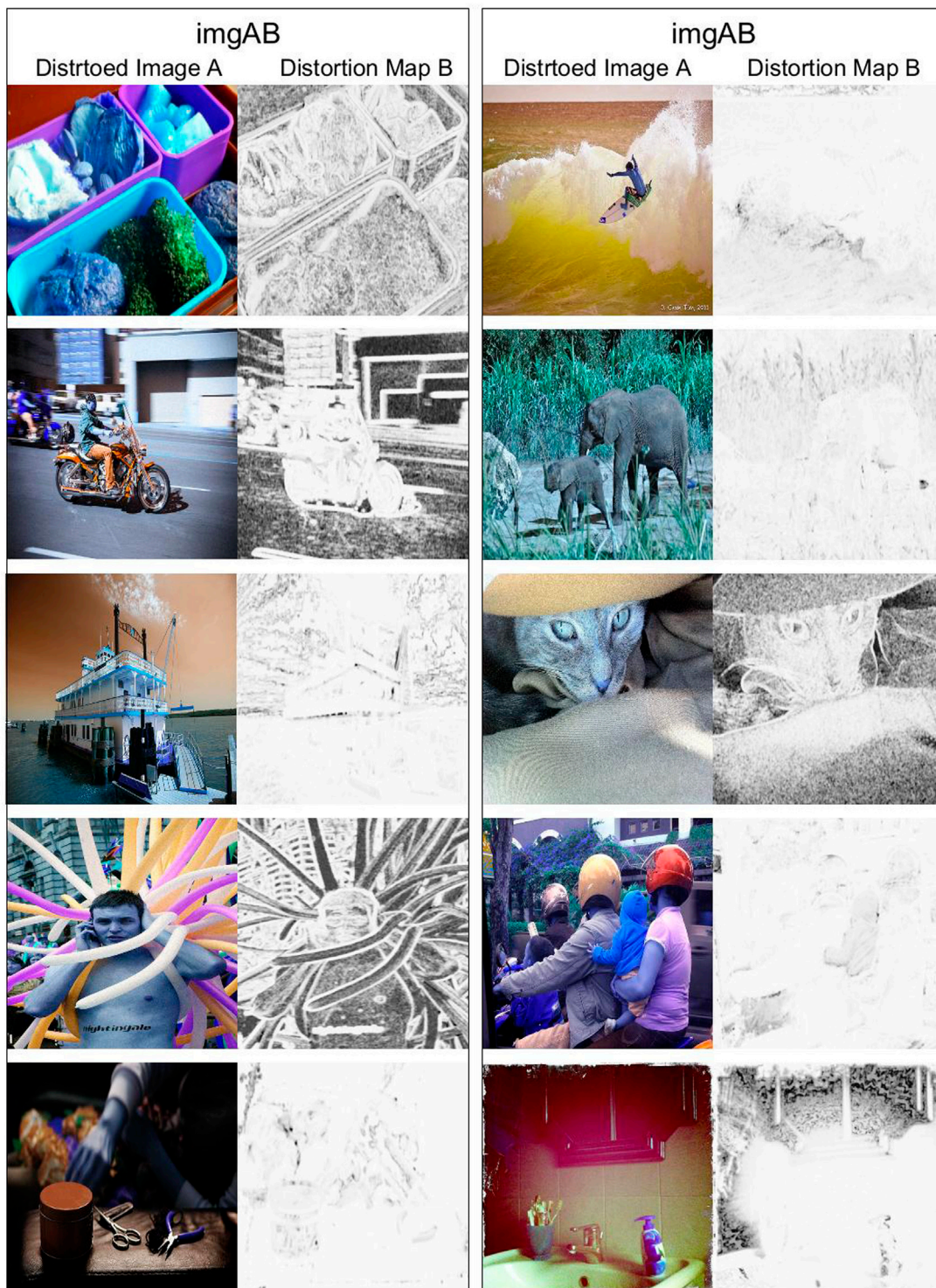
FIGURE 4
Random input samples from cocstuffv2 dataset. Distorted image A is obtained by Albumentation library of augmentations, where Distortion Map B is obtained by SSIM index method.
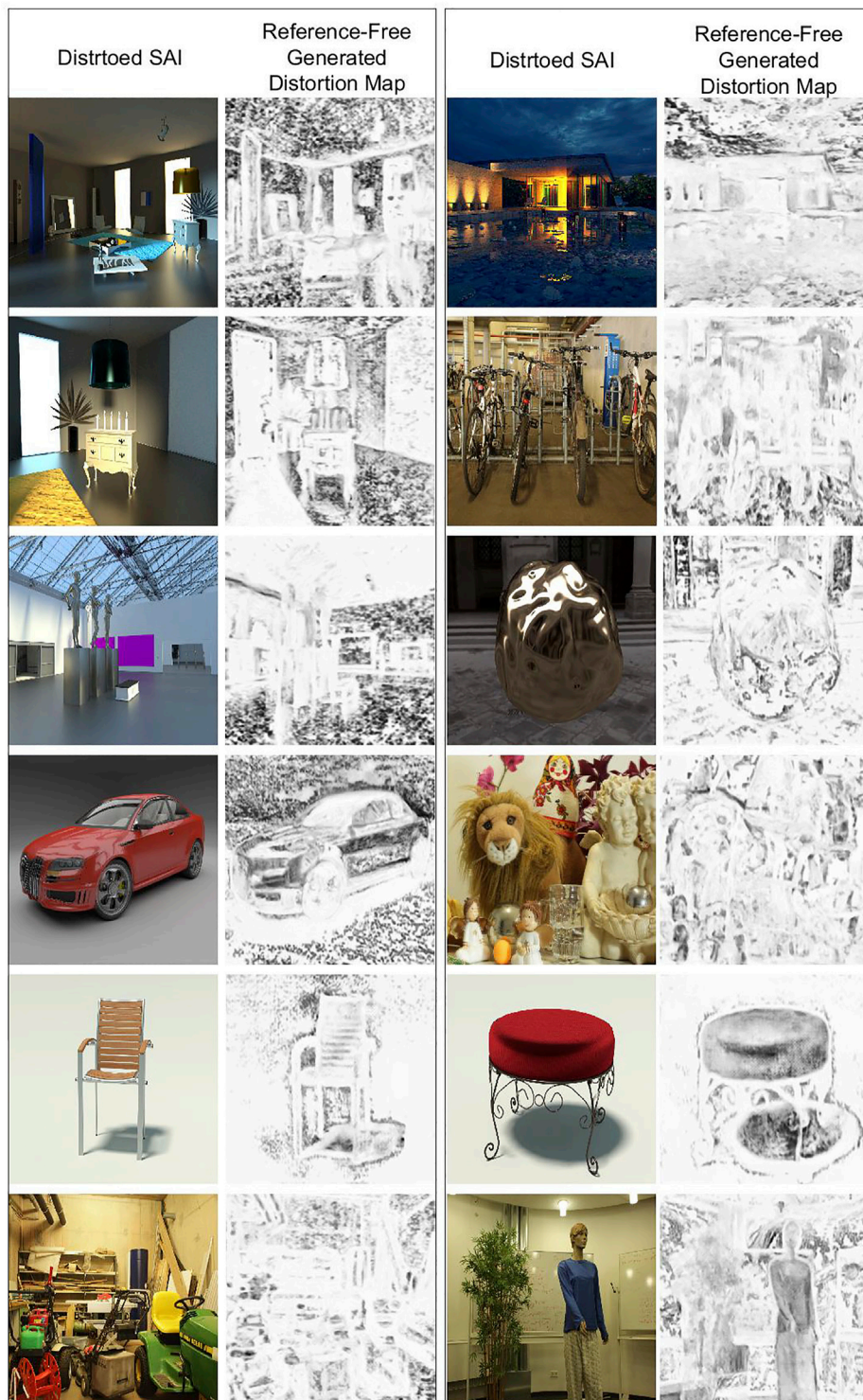
FIGURE 5
Examples of generated distortion maps of central SAIs of different test LFIs from MPI dataset (Adhikarla et al., 2017)

illustrated in Figure 2D, to prepare the feature vectors for test LFIs, we perform a non-linear dimensionality reduction using an Isometric Mapping (Tenenbaum et al., 2000) (ISOMAP) on the generated distortion maps. The ISOMAP algorithm contains three stages. First, it computes k nearest neighbors. Then, it searches for and establishes the shortest path graph. Finally, the Eigen-vectors are computed over the largest Eigen values. The algorithm outputs feature vectors in a multi-dimensional Euclidean space that best represent the intrinsic geometry of the data. As illustrated in Figure 2E, the prepared features vectors of LFIs are fed to a Random Forest Regressor, which performs a regression to predict the LF quality. We chose the RFR because in previous studies it has shown a robust performance (Fern ández-Delgado et al., 2014; Freitas et al., 2018), when compared to other machinelearning algorithms (e.g. neural networks, support vector machines, generalized linear models, etc.).

3 Experimental setup

To train and test the proposed method, we have used the following four LF image quality datasets. We have chosen these datasets because of their content diversity, the types of distortions, and the availability of the corresponding subjective quality scores.

- The MPI Light Field image quality dataset (Adhikarla et al., 2017) contains 13 different scenes with references, 336 test LFIs, and the corresponding subjective quality scores (Mean Observer Scores - MOS). This dataset has typical light field distortions that are specific to transmission, reconstruction, and display.
- The VALID Light Field image quality dataset (Viola and Ebrahimi, 2018) contains five contents, taken from EPFL (Rerabek and Ebrahimi, 2016) light field image dataset, and 140 test LFIs that are compressed using state-of-the-art compression algorithms.

The dataset contains both subjective (MOS) and objective quality scores (PSNR and SSIM).

- The SMART Light Field image quality dataset (Paudyal et al., 2016, 2017) has 16 original LFIs representing both indoor and outdoor scenes. The image content corresponds not only to the scenes with different levels of colorfulness, spatial information, and texture, but also LF specific characteristics such as reflection, transparency, and depth of field. The dataset also contains 256 distorted sequences obtained using four compression algorithms, with their corresponding MOS values.
- The Win5-LID Light Field image quality dataset (Shi et al., 2018) contains six real scenes (captured by a Lytro Illum) and four synthetic scenes, with a total of 220 test LFIs. The

TABLE 2 Mean SROCC and PLCC values for VALID, SMART, MPI, and Win5-LID datasets obtained by 1,000 simulations of RFR.

Dataset	Distortion	Proposed	
		SROCC	PLCC
MPI	QD	0.9290	0.9866
	Gaussian	0.9886	0.9698
	HEVC	0.9581	0.9876
	OPT	0.9347	0.9394
	Linear	0.9499	0.9960
	NN	0.9753	0.9955
	ALL	0.9743	0.9878
VALID	10bit_HEVC	0.9275	0.9871
	10bit_P3	0.9864	0.9931
	10bit_P5	0.9866	0.9843
	10bit_VP9	0.9258	0.9797
	8bit_HEVC	0.9758	0.9678
	8bit_VP9	0.9380	0.9781
	ALL	0.9650	0.9388
SMART	HEVC	0.9101	0.9463
	JPEG	0.9069	0.9501
	JPEG2000	0.9529	0.8947
	SSDC	0.9050	0.9713
	ALL	0.9307	0.9529
Win5-LID	HEVC	0.9690	0.9398
	JPEG2000	0.9367	0.9752
	LN	0.9550	0.9148
	NN	0.9324	0.9286
	EPICNN	0.9059	0.9383
	ALL	0.9441	0.9535

selected contents carry abundant semantic features, such as people, nature, and objects. The LFIs have an identical angular resolution of 9, ×, 9. The real scenes have spatial resolution equal to 434×625, while the synthetic scenes have a spatial resolution equal to 512×512. The distortions are obtained with compression and interpolation algorithms.

As performance evaluation methods, we used only the Spearman's Rank-Order Correlation Coefficient (SROCC) and the Pearson's Linear Correlation Coefficient (PLCC) for simplicity. We compared the proposed NR LF-IQA method with the following state-of-art LF-IQA methods: MDFM (Tian et al., 2018), LFIQM (Paudyal et al., 2019), Fang et al. (Fang et al., 2018), SDFM (Tian et al., 2020b), Meng et al. (Meng et al., 2020), LGF-LFC (Tian et al., 2020a), NR-LFQA (Shi et al., 2019a), LF-QMLI (Luo et al., 2019), Jiang et al. (Jiang et al., 2018), BELIF Shi

TABLE 3 SROCC and PLCC values obtained for state-of-the-art LF-IQA methods tested on VALID, SMART, MPI, and Win5-LID datasets.

Category	Type	Methods	Year	MPI		VALID		SMART		Win5-LID	
				SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
Pre-defined Functions	FR	UQI	2002	0.7400	0.8460	0.9310	0.9550	0.6480	0.7980	0.8252	0.8764
	FR	SSIM	2004	0.9120	0.9320	0.9500	0.9640	0.7550	0.8010	0.6812	0.7880
	FR	VIF	2006	0.8600	0.8960	0.9620	0.9790	0.7260	0.8370	0.9347	0.9555
	FR	NICE	2009	0.5821	0.5122	0.6211	0.6544	0.5214	0.5426	0.4892	0.5002
	FR	STMAD	2011	0.8650	0.8940	0.7940	0.8020	0.6640	0.8010	0.8489	0.9074
	FR	IW-SSIM	2011	0.9320	0.9440	0.9650	0.9780	0.8060	0.8850	0.8212	0.8736
	FR	IW-PSNR	2011	0.9300	0.9160	0.9470	0.9670	0.7840	0.8520	0.8842	0.9022
	FR	MJ3DFR	2013	0.8720	0.9300	0.9560	0.9700	0.8160	0.8480	0.8836	0.8998
	FR	GMSD	2014	0.7358	0.7410	0.6821	0.6948	0.7264	0.8000	0.4352	0.5041
	FR	FI-PSNR	2014	0.8740	0.8510	0.7060	0.7060	0.7730	0.8320	0.6951	0.7419
	FR	PSNR-YUV	2014	0.9342	0.9452	0.9230	0.9310	0.9102	0.9211	0.9007	0.9215
	FR	MW-PSNR	2016	0.7251	0.7698	0.6869	0.6904	0.5281	0.5869	0.7582	0.7758
	FR	MDFM Tian et al. (2018)	2018	0.8346	0.8123	0.7120	0.7198	0.7535	0.7683	0.8157	0.8591
	FR	Fang et al. (2018)	2018	0.8065	0.7942	—	—	—	—	—	—
	RR	LFIQM Paudyal et al. (2019)	2019	0.6815	0.7013	0.3934	0.5001	0.4503	0.4763	0.4503	0.4763
	FR	SDFM Tian et al. (2020b)	2020	0.8435	0.8423	0.824	0.8542	0.7514	0.7941	0.6742	0.7142
	FR	Meng et al. (2020)	2020			0.9579	0.9762	—	—	—	—
	FR	LGF-LFC Tian et al. (2020a)	2020	0.8543	0.8476	—	—	0.8246	0.8276	—	—
DL	FR	Jiang et al. (2018)	2018	-	0.8954	—	—	—	—	—	—
	NR	BELIF Shi et al. (2019a)	2019	0.8854	0.9096	0.8863	0.8950	0.8367	0.8833	0.8719	0.8910
	NR	NR-LFQA Shi et al. (2019b)	2019	0.9119	0.9155	0.9257	0.9658	0.8803	0.9105	0.9032	0.9206
	NR	LF-QMLI Luo et al. (2019)	2019	—	—	0.9286	0.9683	—	—	0.8802	0.9038
	NR	Shan et al. (2019)	2019	—	—	—	—	0.8917	0.9106	—	—
	NR	Tensor-NLFQ Zhou et al. (2019)	2019	0.9101	0.9225	0.9326	0.9746	0.8702	0.9028	0.9101	0.9217
	NR	Ak et al. (2020)	2020	0.8942	0.9005	—	—	—	—	—	—
	NR	VBLIF Xiang et al. (2020)	2020	0.9015	0.9158	—	—	—	—	0.9009	0.9232
DL + ML	NR	Proposed	2020	0.9743	0.9878	0.9650	0.9781	0.9307	0.9529	0.9441	0.9535

et al. (2019b), Tensor-NLFQ (Zhou et al., 2019), Ak *et al.* (Ak et al., 2020), Shan *et al.* (Shan et al., 2019) and VBLIF (Xiang et al., 2020). We also compared the proposed method with the following 2D image/video quality metrics: PSNR-YUV (Sze et al., 2014), IW-PSNR (Wang and Li, 2011), FI-PSNR (Lin and Wu, 2014), MW-PSNR (Sandić-Stanković et al., 2016), SSIM (Wang et al., 2004), IW-SSIM (Wang and Li, 2011), UQI (Zhou and Bovik, 2002), VIF (Sheikh and Bovik, 2006), MJ3DFR (Chen et al., 2013), GMSD (Xue et al., 2014), NICE (Rouse and Hemami, 2009) and STMAD (Vu et al., 2011).

For training and testing the RFR method, we divided each dataset into two content-independent training and testing subsets, i.e,. distorted images generated from one reference in the testing subset are not present in the training subset and vice-

versa. We define a group of scenes as a group containing the reference LFI and its corresponding distorted versions. Then, 80% of the groups were randomly selected for training and the remaining 20% were used for testing. The partition was repeated 1,000 times to eliminate the bias caused by data division. We reported the mean correlation values for the test set over 1,000 simulations.

4 Experimental results

Table 2 shows the correlation values obtained for the VALID, SMART, MPI, and Win-LID LFI quality datasets. The rows in this table show the results for each dataset and for each

distortion, with the "All" row corresponding to the results obtained for the complete datasets. The proposed method performs very well for the MPI dataset obtaining SROCC of 0.97 and PLCC of 0.98, and for the VALID dataset obtaining SROCC of 0.96 and PLCC of 0.93. For the SMART dataset, the method obtains SROCC of 0.93 and PLCC of 0.95, while for the Win5-LID dataset, the method achieves SROCC of 0.94 and PLCC of 0.95. Across the different distortions, the proposed method also performed very well, with only a few distortions showing slightly lower values (e.g. SROCC values of JPEG and SSDC in SMART dataset, and EPICNN in Win5-LID dataset).

Table 3 illustrates the comparison of the results with other state-of-the-art LFI-IQA methods. In this table, the NR and FR LF-IQA methods are classified into three categories, taking into consideration the models used to map the pooled features into quality estimates. The categories include methods that use *1*) a pre-defined function, *2*) a machine-learning (ML) algorithm, or *3*) a deep-learning (DL) approach to obtain the predicted quality score. Notice that, for simplicity, only the overall performance ("ALL") correlation values are reported for each dataset. Also, since the authors of these LF-IQA methods did not publish their results for all four datasets, our matrix is incomplete. For VALID dataset, NR-LFQA (Shi et al., 2019a) and Tensor-NLFQ (Zhou et al., 2019) methods have reported correlations separately for 8bit and 10bit compressed LF images. For comparison of results in Table 3, we have shown averaged SROCC and PLCC obtained by these methods for complete VALID dataset. Notice that the proposed method has achieved the highest correlation values among all LF-IQA methods for all of the four datasets. It is also worth pointing out that the pooling and mapping strategies in the proposed NR LF-IQA method has achieved significant improvement in quality predictions and shown higher SROCC and PLCC than the original SSIM method.

5 Conclusion

In this paper, we have proposed a blind LF-IQA method that is based on reference-free distortion maps. To generate reference-free distortion maps from test LFIs, we have used a GAN deep-learning architecture, the Pix2PixGAN, which learns from synthetically generated distorted images and their corresponding ground truth distortion maps. Since the ground truth distortion maps are not available in any of the existing LF or image quality datasets, we use distortion maps generated by SSIM as the ground truth distortion map. Next, we train the Pix2PixGAN using the synthetically generated dataset. The proposed LF-IQA method has following five stages: *1*) Generation of a synthetically distorted dataset of 2D images, *2*) Pix2PixGAN training to generate 2D distortion maps, using SSIM distortion maps as ground truth, *3*) generation of distortion maps of sub-aperture images using the trained Pix2PixGAN, *4*) non-linear reduction of dimensionality through Isometric Mapping on the generated distortion maps to obtain the LFI feature vectors, and *5*) perform regression using RFR algorithm to predict LFI quality. The correlation values of the proposed method computed on four different datasets are higher than what is obtained by other state-of-the-art LF-IQA methods. As future work, we plan to explore using different state-of-the-art FR-IQA metrics to generate the ground truth distortion maps, and train the Pix2PixGAN architecture. It is worth pointing out that the proposed method can work as a framework to train for other types of no-reference LF-IQA methods.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SA performed all the coding and tests. MF and PG worked on the idea, discussed the work. SA and MF wrote the paper.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adhikarla, V. K., Vinkler, M., Sumin, D., Mantiuk, R., Myszkowski, K., Seidel, H.-P., et al. (2017). "Towards a quality metric for dense light fields," in Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).

Ak, A., Ling, S., and Le Callet, P. (2020). "No-reference quality evaluation of light field content based on structural representation of the epipolar pane image," in The 1st ICME Workshop on Hyper-Realistic Multimedia for Enhanced Quality of Experience, London, UK.

Arad, N. (1998). Grid-distortion on nonrectangular grids. *Comput. Aided Geom. Des.* 15, 475–493. doi:10.1016/S0167-8396(98)00003-X

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information* 11. doi:10.3390/info11020125

Caesar, H., Uijlings, J., and Ferrari, V. (2018). "Coco-stuff: Thing and stuff classes in context," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1209–1218. doi:10.1109/CVPR.2018.00132

Chen, M.-J., Su, C.-C., Kwon, D.-K., Cormack, L. K., and Bovik, A. C. (2013). Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process. Image Commun.* 28, 1143–1155. doi:10.1016/j.image.2013.05.006

Fang, Y., Wei, K., Hou, J., Wen, W., and Imamoglu, N. (2018). "Light filed image quality assessment by local and global features of epipolar plane image," in 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), 1–6. doi:10.1109/BigMM.2018.8499086

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* 15, 3133–3181.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A, Opt. image Sci.* 4 (12), 2379–2394.

Freitas, P. G., Alamgeer, S., Akamine, W. Y. L., and Farias, M. C. Q. (2018). "Blind image quality assessment based on multiscale salient local binary patterns," in MMSys '18: Proceedings of the 9th ACM Multimedia Systems Conference (New York, NY, USA: Association for Computing Machinery), 52–63. doi:10.1145/3204949.3204960

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (Cambridge, MA, USA: MIT Press), 2672–2680.

Hahne, C., Lumsdaine, A., Aggoun, A., and Velisavljevic, V. (2018). Real-time refocusing using an fpga-based standard plenoptic camera. *IEEE Trans. Industrial Electron.* 65, 9757–9766. doi:10.1109/TIE.2018.2818644

Hou, J., Chen, J., and Chau, L. (2019). Light field image compression based on bi-level view compensation with rate-distortion optimization. *IEEE Trans. Circuits Syst. Video Technol.* 29, 517–530. doi:10.1109/TCSVT.2018.2802943

Isola, P., Zhu, J., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5967–5976. doi:10.1109/CVPR.2017.632

Jiang, G., Huang, Z., Yu, M., Xu, H., Song, Y., and Jiang, H. (2018). New quality assessment approach for dense light fields. *Proc. Volume 10817, Optoelectron. Imaging Multimedia Technol. V* 44, 1081717. doi:10.1117/12.2502277

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: Common objects in context," in *Computer vision – eccv 2014*. Editors D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer International Publishing), 740–755.

Lin, Y., and Wu, J. (2014). Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. *IEEE Trans. Image Process.* 23, 1527–1542.

Luo, Z., Zhou, W. R., Shi, L., and Chen, Z. (2019). No-reference light field image quality assessment based on micro-lens image. *ArXiv abs/1908.10087*.

Meng, C., An, P., Huang, X., Yang, C., and Liu, D. (2020). Full reference light field image quality evaluation based on angular-spatial characteristic. *IEEE Signal Process. Lett.* 27, 525–529.

Paudyal, P., Battisti, F., and Carli, M. (2019). Reduced reference quality assessment of light field images. *IEEE Trans. Broadcast.* 65, 152–165.

Paudyal, P., Battisti, F., Sjöström, M., Olsson, R., and Carli, M. (2017). Toward the perceptual quality evaluation of compressed light field images. *IEEE Trans. Broadcast.* 63, 1–16. doi:10.1109/TBC.2017.2704430

Paudyal, P., Olsson, R., Sjostrom, M., Battisti, F., and Carli, M. (2016). "Smart: A light field image quality dataset," in Procs. of the ACM Multimedia Systems 2016 Conference, (MMSYS).

Rerabek, M., and Ebrahimi, T. (2016). *New light field image dataset*.

Rouse, D. M., and Hemami, S. S. (2009). "Natural image utility assessment using image contours," in 2009 16th IEEE International Conference on Image Processing (ICIP), 2217–2220. doi:10.1109/ICIP.2009.5413882

Sandić-Stanković, D., Kukolj, D., and Le Callet, P. (2016). Dibr-synthesized image quality assessment based on morphological multi-scale approach. *EURASIP J. Image Video Process.* 2017, 4. doi:10.1186/s13640-016-0124-7

Shan, L., An, P., Meng, C., Huang, X., Yang, C., and Shen, L. (2019). A no-reference image quality assessment metric by multiple characteristics of light field images. *IEEE Access* 7, 127217–127229.

Sheikh, H. R., and Bovik, A. C. (2006). Image information and visual quality. *IEEE Trans. Image Process.* 15, 430–444. doi:10.1109/TIP.2005.859378

Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* 15, 3440–3451. doi:10.1109/TIP.2006.881959

Shi, L., Zhao, S., and Chen, Z. (2019a). "Belif: Blind quality evaluator of light field image with tensor structure variation index," in 2019 IEEE International Conference on Image Processing (ICIP), 3781–3785.

Shi, L., Zhao, S., Zhou, W., and Chen, Z. (2018). "Perceptual evaluation of light field image," in 2018 25th IEEE International Conference on Image Processing (ICIP), 41–45. doi:10.1109/ICIP.2018.8451077

Shi, L., Zhou, W. R., and Chen, Z. (2019b). No-reference light field image quality assessment based on spatial-angular measurement. *ArXiv abs/1908.06280*.

V. Sze, M. Budagavi, and G. Sullivan (Editors) (2014). *High efficiency video coding (HEVC): Algorithms and architectures* (Cham: Springer). doi:10.1007/978-3-319-06895-4

Tenenbaum, J. B., Silva, V. d., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi:10.1126/science.290.5500.2319

Tian, Y., Zeng, H., Hou, J., Chen, J., and Ma, K. (2020a). Light field image quality assessment via the light field coherence. *IEEE Trans. Image Process.* 29, 7945–7956.

Tian, Y., Zeng, H., Hou, J., Chen, J., Zhu, J., and Ma, K. (2020b). A light field image quality assessment model based on symmetry and depth features. *IEEE Trans. Circuits Syst. Video Technol.* 31, 2046–2050. doi:10.1109/TCSVT.2020.2971256

Tian, Y., Zeng, H., Xing, L., Chen, J., Zhu, J., and Ma, K.-K. (2018). A multi-order derivative feature-based quality assessment model for light field image. *J. Vis. Commun. Image Represent.* 57, 212–217.

Viola, I., and Ebrahimi, T. (2018). "Valid: Visual quality assessment for light field images dataset," in 10th International Conference on Quality of Multimedia Experience (QoMEX), Sardinia, Italy, 3. doi:10.1109/QoMEX.2018.8463388

Vu, P. V., Vu, C. T., and Chandler, D. M. (2011). "A spatiotemporal most-apparent-distortion model for video quality assessment," in 2011 18th IEEE International Conference on Image Processing, 2505–2508.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. image Process.* 13, 600–612.

Wang, Z., and Li, Q. (2011). Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* 20, 1185–1198.

Xiang, J., Yu, M., Chen, H., Xu, H., Song, Y., and Jiang, G. (2020). "Vblfi: Visualization-based blind light field image quality assessment," in 2020 IEEE International Conference on Multimedia and Expo (ICME), 1–6.

Xue, W., Zhang, L., Mou, X., and Bovik, A. C. (2014). Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* 23, 684–695. doi:10.1109/TIP.2013.2293423

Zhou, W, and Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. doi:10.1109/97.995823

[Dataset] Zhou, W., Likun, S., and Chen, Z. (2019). Tensor oriented no-reference light field image quality assessment. *IEEE Trans. Image Process.* 29, 4070. doi:10.1109/TIP.2020.2969777