FROM IMAGE QUALITY TO PATCH QUALITY: AN IMAGE-PATCH MODEL FOR NO-REFERENCE IMAGE QUALITY ASSESSMENT

Wen Heng and Tingting Jiang

National Engineering Laboratory for Video Technology, Cooperative Medianet Innovation Center,
School of Electronics Engineering and Computer Science, Peking University

ABSTRACT

Supervised learning is gradually used for image quality assessment (IQA). For the patch-based methods, the 'ground truth' quality of patches is essential for training, but in practice it's easy to obtain the ground truth quality of images rather than patches. So we propose an Image-Patch model (IPM) to estimate the 'ground truth' quality for patches with known ground truth quality of images. Combined with baseline image quality estimator e.g. convolutional neural network IQA (CNN-IQA), the IPM can reduce the noise in patches' labels and make training more efficiently. The experiments show that the IPM improves the performance of baseline estimator on most of the distortion types while make great progress in evaluating local quality.

Index Terms— image quality assessment, convolutional neural networks, patch quality, supervised learning

1. INTRODUCTION

Perceptual quality of images is a fundamental metric in many image processing tasks or image-related applications. Image quality assessment (IQA) methods fall into three categories: Full-Reference IQA (FR-IQA), Reduced-Reference IQA (RR-IQA) and No-Reference IQA (NR-IQA). In real-world scenarios, due to the limitation of getting access to non-distortion reference images, NR-IQA methods are preferable.

In recent years many learning-based IQA methods have been proposed, especially deep learning methods [1, 2, 3]. These methods usually require a large amount of data for model training and existing IQA datasets can't meet the demand. To solve this problem, patch-based methods are gradually used in IQA, e.g. CORNIA [4], CNN-IQA [3].

In patch-based IQA methods, there are bilateral relationships between image quality and patch quality: from patch to image and from image to patch. In the former scenario, the image quality estimation problem is transformed into the patch quality estimation problem. With estimated patch quality scores, a spatial pooling algorithm is taken to obtain the image quality score. In practice, spatial pooling has been treated superficially for convenience, e.g. using a simple spatial average. [5] evaluated the effect of different pooling strategies and concluded that the information content-weight pooling strategy performed best. In addition, some works focused on weighted pooling strategy based on visual attention, e.g. [6] proposed a visual importance pooling for IQA.

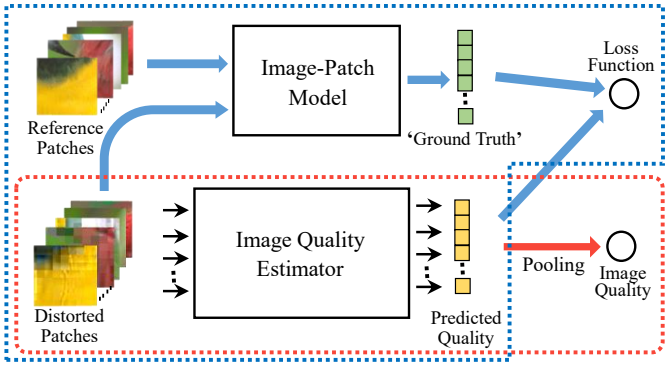


Fig. 1: The NR-IQA framework with IPM and learning-based image quality estimator. The chart in the blue dashed box represents the training process. And the chart in the red dashed box represents the inference process.

From image to patch is the other important direction which is often ignored. The key to the problem is how we can get patch quality score based on the known image quality score. Why is this important? The patch-based learning methods requires the 'ground truth' of patch quality for training but there are only the ground truth quality of images instead of patches in IQA datasets. To deal with this problem, existing works usually assign the image quality score to all patches in this image as their 'ground truth', e.g. CNN-IQA [3]. This approach might introduce much noise in patches' labels because in some distortion types the quality of patches in one image varies much and the patches' quality score can't be simply assigned as the image quality core. To avoid introducing much noise, a well designed method is expected to estimate quality score for each patch based on the ground truth quality score of one image.

We propose an Image-Patch model to help estimate the 'ground truth' quality of patches based on the known image quality score and the reference image. In a formal statement, the aim is to find a nonlinear function $f(\mathbf{x};\theta)$ to estimate the

patch quality where $\mathbf{x}$ is the feature of the patch and θ is the parameter of the model. For simplicity, we try patch's Mean Square Error (MSE) and Structural Similarity Index Measure (SSIM) [7] as the feature and assume the model is a curve model which is fitted by a cubic polynomial function. Then based on the following two observations (1) MSE=0 or SSIM=1 means theoretically best quality score (2) the mean quality score of the patches with lowest quality is linear to the ground truth quality of image, we use some prior points to fit the curve model. Then it's easy to get the 'ground truth' quality for all patches with the fitted model. Lastly, the 'ground truth' quality of patches can be used to train a learning-based image quality estimator. In this paper, we conducted experiments with the framework that combined our IPM with a baseline estimator: CNN-IQA.

It is worth noting that the IPM only works in the training stage of baseline estimator, and it helps the baseline estimator learn more representative features. However, in the inference stage, we only sample raw patches from test images and send them into the learned model to get the predicted quality scores. Therefore, the proposed framework is FR in training stage but NR in inference stage. In terms of the practical applications of IQA methods, we believe it is reasonable to classify the proposed framework as a kind of NR-IQA metric.

Our contributions are as follows. First, we propose a method to model the relationship from image quality score to patch quality score. To our best knowledge, it is the first work to solve this problem. Second, the IPM can increase the amount of training data through estimating credible 'ground truth' quality scores for patches. Finally, it improves the performance of existing learning-based IQA methods.

2. IMAGE-PATCH MODEL

2.1. Formulation

Given a distorted image $\mathbf{D}$, its reference image $\mathbf{R}$ and the ground truth quality score of distorted image Q, the objective is to estimate the 'ground truth' quality $\{q_1, \ldots, q_l\}$ for patches $\{\mathbf{d}_1, \ldots, \mathbf{d}_l\}$ of $\mathbf{D}$ whose corresponding patches of $\mathbf{R}$ are $\{\mathbf{r}_1, \ldots, \mathbf{r}_l\}$, where l is the number of patches sampled from one image. All the patches have the same size.

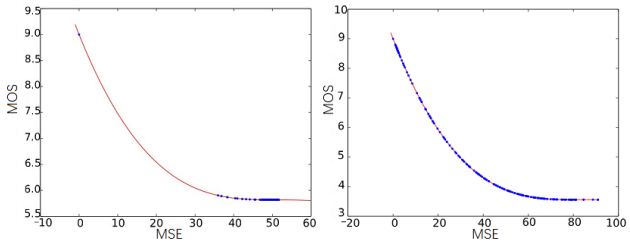
In the Image-Patch model, we use a non-linear function f to model the relationship between variables q_i and $\mathbf{d}_i$ in the following form,

$$q_i = f(\Phi(\mathbf{d}_i); \theta) \tag{1}$$

where $\Phi(\mathbf{d}_i)$ represents the feature of patch $\mathbf{d}_i$, and θ is the parameters of the non-linear function.

The design of features for patches is the key to the problem. It's feasible to take the raw patch as the model input, but it will be hard to solve the parameters in the model due to the high dimension of input. Therefore, the feature should be compact to represent the distortion degree of patches.

We found FR-IQA methods are good choice for the design of features because these methods use the information from both reference and distortion image. In our work, we tried the frequently used FR-IQA methods: MSE and SSIM [7], and the experiments results show that both can improve the performance of a baseline image quality estimator.



(a) (b)

Fig. 2: (a) and (b) are the fitted non-linear model of one GN and JPEGTE image from TID2008. The points represent the IPM fitted quality scores for all patches in one image. The number of patches are same for (a) and (b).

2.2. Priori Observations for IPM

Next, we introduce some assumptions on IPM properties based on some priori observations. In addition, we define a reasonable learning set with which we can solve the IPM through a optimization process under the constraints of above assumed properties.

One obvious property of the IPM is monotonicity, which means if the quality of input patch is low then the estimated 'ground truth' quality should also be low. For example, using s_i and s_j denote the SSIM for patch i and j. Then, if $s_i > s_j$, $f(s_i) > f(s_j)$ (higher SSIM means higher quality).

Another priori observation is if a patch with MSE=0 or SSIM=1 and its quality score should be the highest MOS value (e.g. MOS=9 in TID2008 [8]). This doesn't mean that the IPM only works for images with local distortion which contain distortion-free patches. It just guarantee the fitting curve in IPM will always cross the point that MSE=0 or SSIM=1 and MOS equals the highest value regardless of whether there are distortion-free patches in the images.

Moreover, previous work [5, 6] showed that the lowest quality region in one image emphatically influences the subjective judgement of its quality. [9] found that in one image the percentile pooling results of 10% patches with lowest FSIM [10] quality scores have much better linearity with the quality score of image than the 100% patches. Based on this observation, it would be reasonable to roughly assign the image quality score to the 10% patches with lowest MSE or highest SSIM.

With above observations, we get the learning set that can be used to learn the non-linear model. The learning set includes two parts: first, denoted as L_{fix}, the patch with MSE=0 or SSIM=1 and quality score equals the highest MOS, which guarantees the non-linear model crossing one fixed point; second, 10% patches with worst quality, denoted

as L_{worst}, that their quality scores are close to the image's quality score, which would push the non-linear model to be quite flat at the higher MSE or lower SSIM phase as shown in Fig. 2. Meanwhile considering that the model have the monotonicity, we apriori assume the non-linear curve model in IPM is a cubic polynomial function in the following form,

$$f(\Phi(\mathbf{d});\theta) = a\Phi(\mathbf{d})^3 + b\Phi(\mathbf{d})^2 + c\Phi(\mathbf{d}) + d \quad (2)$$

$$s.t. \quad \begin{cases} d = \mathrm{MOS}_{\max} & \mathbf{d} = L_{fix}, for\ MSE \\ a+b+c+d = \mathrm{MOS}_{\max} & \mathbf{d} = L_{fix}, for\ SSIM \\ f(\Phi(\mathbf{d}_i);\theta) = Q & \mathbf{d}_i \in L_{worst} \end{cases}$$

where $\theta = \{a, b, c, d\}$ are the parameters of the non-linear model, and $\mathrm{MOS}_{\max} = 9.0$ represents the theoretically maximum MOS in TID2008.

2.3. Model Fitting

We use the least square method to fit the above curve model and obtain the parameters that best fit the learning set. Then we take the MSE $\{m_1, \ldots, m_l\}$ or SSIM $\{s_1, \ldots, s_l\}$ of all patches as the input of the IPM and get their corresponding estimated quality score $\{q_1, \ldots, q_l\}$. Fig. 2 shows two fitting-curves of two example images. It's shown that the patch quality distribution of different distortion types vary greatly, which indicates the necessity of designing 'ground truth' quality for patches accodingly .

3. IQA FRAMEWORK WITH CNN-IQA

The proposed IPM aims to make the image quality estimator training more efficiently, as shown in Fig. 1. We choose currently one state-of-the-art learning-based NR-IQA method: CNN-IQA[3] as the basic image quality estimator.

The CNN architecture in [3] includes one convolution layer, one max-min pooling layer, two fully connected layers and an output node. The network directly takes raw patch as input and predicts the quality score for it. Different from the IPM, in [3] the author directly assigned the image quality score to the corresponding all extracted patches as their 'ground truth'.

3.1. Training and Inference

Training For each training image, we fit an IPM to estimate the quality scores $\{q_1, \ldots, q_l\}$ for all extracted patches. Then the fitted quality score will play the role of 'ground truth' to conduct loss function as shown in Fig. 1. Let $F(\mathbf{d_i}; \mathbf{w})$ be the predicted score of patch $\mathbf{d_i}$ with network parameter $\mathbf{w}$ in image quality estimator: CNN-IQA. The adopted objective function is similar to [3] as follows:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \|F(\mathbf{d}_i; \mathbf{w}) - q_i\|_1 \quad (3)$$

The above is a minimum optimization problem. Same to [3], We use Stochastic Gradient Decent (SGD) to solve this problem. A validation set is used to select parameters of the trained model and prevent over-fitting. In experiments we perform SGD until convergence when training and keep the model parameters that generate the highest Spearman Rank Order Correlation Coefficient (SROCC) on the validation set.

Inference For each test image, We feed all extracted patches into the network, and obtain the predicted patch quality scores. Then, different from mean pooling strategy in [3], we take a percentage pooling strategy that take the average of the lowest 10% patches' quality scores as the predicted quality score for image. This way corresponds to the priori assumption in subsection 2.2.

4. EXPERIMENTS

Dataset: We conducted experiments on TID2008 dataset [8]. There are total 1700 distorted images derived from 25 reference images with 17 different distortion types at 4 degradation levels. Mean Opinion Score (MOS) for this dataset were computed for each image, which is in the range 0 to 9. The 17 distortion types include: Additive Gaussian noise (AGN), Additive noise in color components (ANCC), Spatially correlated noise (SCN), Masked noise (MN), High frequency noise (HFN), Impulse noise (IN), Quantization noise (QN), Gaussian blur (GB), Image denoising(IDN), JPEG compression (JPEG), JPEG2000 compression (JP2K), JPEG transmission errors (JPEGTE), JPEG2000 transmission errors (JP2KTE), Non eccentricity pattern noise (NEPN), Local block-wise distortions (LBD), Intensity shift (IS) and Contrast change (CC).

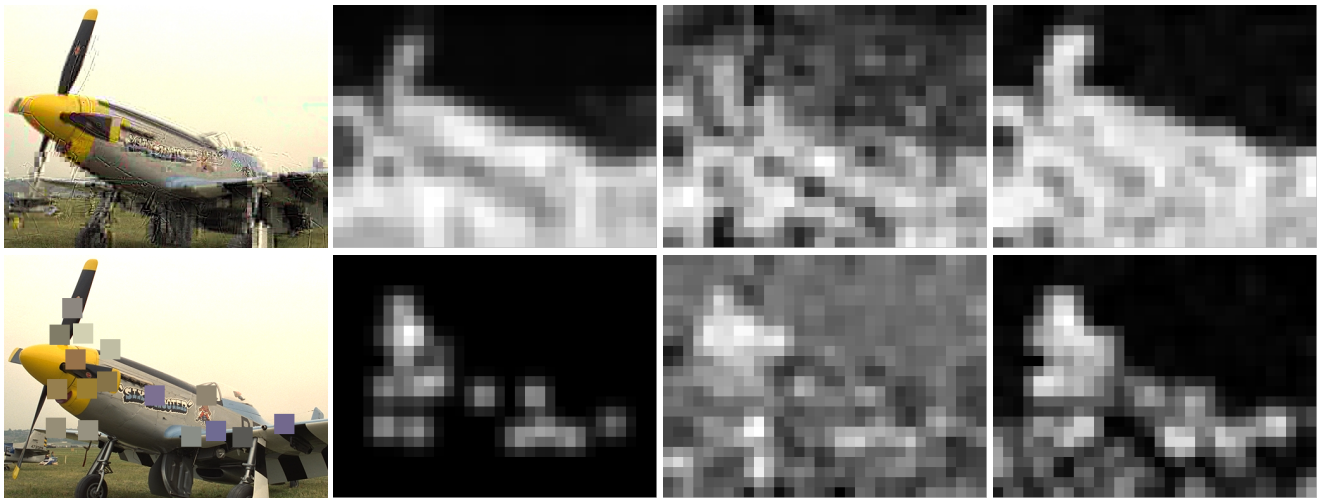
Evaluation Method: SROCC was used to measure the consistency strength between predicted quality score and the ground truth quality of image. To eliminate the bias due to division of the data, we performed a repeated random train-test split for 20 trials. Each trial selected 60% reference image and their distorted versions as training set, 20% as validation set and the rest 20% as test set.

Image Preprocessing: Before model training the local contrast normalization was performed on all images. Then patches with size 32×32 were sampled at stride 16 with overlap from each image.

4.1. Experimental results on TID2008

Table 1 shows the comparison results between our method and three FR-IQA methods: peak-signal-to-noise ratio (PSNR), SSIM [7], VIF [11], and four NR-IQA methods: CS [12], QAC [13], BRISQUE [14], CNN-IQA [3]. Ours(MSE) represents using the IPM with MSE, and Ours(SSIM) represents using the IPM with SSIM. All experiments are under the same experimental configuration. Except CNN-IQA all source codes were provided by the authors. Due to the fact that CNN-IQA neither published the source code nor reported their results on TID2008, we implemented it ourselves based on Theano toolbox [15].

Of all 17 distortion types in TID2008, we experimented on the first 15 because our method is not suitable for the last t-



(a) (b) (c) (d)

Fig. 3: Comparison of local quality evaluation results. (a) JPEGTE and LBD distortion images. (b) MSE maps. (c) Quality maps from CNN-IQA. (d) Quality maps from our method (MSE).

		AGN	ANCC	SCN	MN	HFN	IN	QN	GB	IDN	JPEG	JP2K	JPEGTE	JP2KTE	NEPN	LBD	NDS
FR	PSNR	**0.897**	**0.899**	**0.914**	**0.855**	**0.925**	0.883	0.872	0.917	0.948	0.887	0.820	0.769	0.853	0.641	0.673	0.747
	SSIM [7]	0.864	0.838	0.832	0.763	0.900	0.774	0.829	0.948	**0.955**	**0.920**	0.965	**0.847**	**0.895**	0.605	**0.885**	**0.816**
	VIF [11]	0.887	0.884	0.887	0.837	0.906	**0.886**	**0.876**	**0.950**	0.926	0.911	**0.972**	0.817	0.873	**0.800**	0.869	0.696
NR	CS [12]	0.856	0.470	0.132	0.554	**0.941**	0.904	0.365	0.853	0.621	0.930	0.859	-0.185	0.014	0.111	0.025	0.265
	QAC [13]	0.641	0.643	0.218	0.640	0.809	0.755	0.604	0.795	0.465	0.805	0.837	0.035	0.284	0.082	0.392	0.253
	BRISQUE [14]	0.886	**0.887**	0.819	0.794	0.932	**0.931**	0.799	0.783	0.677	0.842	0.832	0.440	**0.827**	-0.033	0.456	0.618
	CNN-IQA [3]	0.790	0.744	0.767	0.851	0.882	0.827	0.700	0.899	**0.919**	0.908	**0.930**	**0.839**	0.808	0.513	0.718	0.621
	Ours(MSE)	0.908	0.876	**0.915**	**0.867**	0.928	0.890	0.842	**0.945**	0.867	0.939	0.920	0.837	0.805	0.736	**0.752**	**0.861**
	Ours(SSIM)	**0.912**	0.877	0.914	0.862	0.928	0.882	**0.843**	0.941	0.865	**0.941**	0.922	0.816	0.818	**0.764**	0.703	0.858

Table 1: Median SROCC of 20 train-test splits on TID2008 dataset. The bold type indicates best results in FR-IQA and NR-IQA methods.

wo due to the preprocessing step. Both distortion-specific (DS) and non-distortion-specific (NDS) experiments were performed for all methods.

For distortion-specific experiments, our method outperformed other NR-IQA methods on more than half of the distortion types. Compared to CNN-IQA, our method with IPM improved a lot because the IPM can reduce the noise in patches' labels and make the CNN training more efficiently. For non-distortion-specific experiments, our method performed even better than FR-IQA methods. This indicates that IPM helps make the image quality estimator more general and work widely without the limitation of specific distortion type.

It should be pointed out that, for LBD distortion the proportion of distorted patches ranges from 5% to 30% in different images. Therefore, in DS experiment, we used 5% patches to fit the IPM in training process and took 5% percentile pooling strategy in inference process.

4.2. Local Quality Evaluation

To visually show the promotion of using IPM, we test the ability to predict local quality of the proposed method and CNN-IQA. With the trained model from NDS experiments, we predicted the quality score for patches which sampled in the size 32×32 at a stride 16, and then normalized the predicted quality score into range [0,255] to form a quality map.

The tested original images, MSE maps and corresponding quality maps are shown in Fig. 3. It's obvious that our method can locate the low quality region more accurately. It's important to note that our method can predict the quality map for any given image without the need of the reference image.

5. CONCLUSION

We propose an Image-Patch model to help estimate the 'ground truth' for patches. We combined the IPM with one basic image quality estimator: CNN-IQA, and find IPM can improve the performance of the basic estimator on most distortion types. In addition, the IPM also makes the estimator more general and gains great progress in evaluating local quality.

6. ACKNOWLEDGMENTS

This work is partially supported by the National Basic Research Program of China (973 Program) (2015CB351803), National Science Foundation of China (61572042, 61390514, 61421062, 61210005, 61527084).

7. REFERENCES

[1] Weilong Hou and Xinbo Gao, "Be natural: A saliency-guided deep framework for image quality," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[2] Huixuan Tang, Neel Joshi, and Ashish Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 2877–2884.

[3] Le Kang, Peng Ye, Yi Li, and David Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 1733–1740.

[4] Peng Ye, Jayant Kumar, Le Kang, and David Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 1098–1105.

[5] Zhou Wang and Xinli Shang, "Spatial pooling strategies for perceptual image quality assessment," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 2945–2948.

[6] Anush K Moorthy and Alan Conrad Bovik, "Visual importance pooling for image quality assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 2, pp. 193–201, 2009.

[7] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[8] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, M Carli, and F Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.

[9] Wufeng Xue, Lei Zhang, and Xuanqin Mou, "Learning without human scores for blind image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 995–1002.

[10] Lin Zhang, David Zhang, and Xuanqin Mou, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[11] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[12] Qingbing Sang, Xiaojun Wu, Chaofeng Li, and Yin Lu, "Universal blind image quality assessment using contourlet transform and singular-value decomposition," *Journal of Electronic Imaging*, vol. 23, no. 6, pp. 061104–061104, 2014.

[13] Wufeng Xue, Lei Zhang, and Xuanqin Mou, "Learning without human scores for blind image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 995–1002.

[14] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[15] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio, "Theano: A cpu and gpu math compiler in python," in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 1–7.