Adaptive Patch Labeling and Multi-label Feature Selection for 360-Degree Image Quality Assessment

Abderrezzaq Sendjasni, Mohamed-Chaker Larabi, Seif-Eddine Benkabou

Adaptive Patch Labeling and Multi-label Feature Selection for 360-Degree Image Quality Assessment

Abderrezzaq Sendjasni[1], Mohamed-Chaker Larabi[1] and Seif-Eddine Benkabou[2]

[1] CNRS, Univ. Poitiers, XLIM, UMR 7252, France
[2] LIAS, Univ. Poitiers, France

***Abstract*—Assessing the quality of 360-degree images based on individual regions presents a challenging task. The lack of ground truth opinion scores (MOS) for specific regions makes it difficult to evaluate image quality accurately. Existing datasets only provide MOS for entire 360-degree images, which limits the granularity of assessment. To overcome this challenge, we propose a novel framework that employs adaptive patch labeling techniques. We leverage a set of 2D-IQA methods to generate quality score distributions for each patch in the 360-degree images. These distributions, combined with the available MOS, serve as labels for individual patches, providing a more comprehensive characterization of patch quality. Furthermore, we use these labels to adaptively select and refine deep neural features. By selectively choosing label-specific features, we enhance the accuracy and effectiveness of patch-based 360-degree image quality assessment. This approach allows us to focus on the most relevant and informative features for each patch, resulting in improved assessment performance. The experimental results on two benchmark datasets demonstrate that adaptive patch labeling and feature selection achieve accurate and reliable performances, thus advancing the field of 360-degree image quality assessment.**

***Index Terms*—360-degree, Image quality assessment, Convolutional neural networks, Adaptive Patch labeling, Feature selection, Multi-regression.**

I. Introduction

The use of immersive media has become tremendously popular in recent years. Several reasons are behind such growth, particularly the development of more accessible and affordable immersive technologies [1], such as head-mounted displays (HMDs). As technology continues to improve, more people have the opportunity to engage with immersive content. A popular and widely accessible form of immersive content is 360-degree imagery. The proliferation of such content imaging technology has significantly expanded the possibilities of immersive visual experiences. However, the challenges surrounding its acquisition and processing inflict a loss of visual quality, particularly in compression and transmission. To determine a reliable and appreciable quality of experience (QoE), image quality assessment (IQA) can be used to account for visual quality requirements. Nevertheless, accurately assessing the quality of 360-degree images poses unique challenges due to their panoramic nature and distorted projection.

IQA for 360-degree images refers to the process of evaluating and quantifying their perceived quality. Due to the unique characteristics of such images, including their spherical projection and immersive viewing experience, appropriate IQA techniques are required to accurately assess their quality. These techniques should take into account factors such as geometric distortions, image artifacts, viewing behaviors, and overall visual fidelity. One way to achieve this is through subjective assessment, which involves human observers rating the quality of 360-degree images based on their visual perception. This approach provides reliable quality evaluations, although it can be time-consuming and costly. In addition, it is still a unique process to build datasets and collect data for training, evaluating, and benchmarking objective IQA models [2]. These models aim to automatically assess image quality without relying on the opinions collected from human observers, with the goal of mimicking their visual perception. To achieve reliable accuracy, the IQA model for 360-degree images is designed to consider the unique characteristics of the spherical projection and the specific challenges associated with it.

In the literature, 360-degree IQA models, particularly deep neural-based ones, use viewports or patches [3]–[10], since 360-degree images have high resolution (4k+). This implies that the quality evaluation is carried out on particular portions rather than the complete image. However, available datasets for 360-degree IQA provide ground truth labels, *i.e.* mean opinion scores (MOS), per entire image [11], [12]. Given this, IQA solutions require labeling regions, particularly patch-based ones where each patch is seen as separate content. Although it appears quite straightforward as a process for various image processing tasks such as classification, detection, etc., it is quite challenging for IQA. Since the MOS represents the global quality of 360-degree images, patch labels should represent the local quality. In existing methods [4], [5], [7], patch quality is directly inherited from the corresponding image, generating redundancy among quality labels and assigning the same MOS for all patches. This may be problematic since local quality is not always consistent with global quality due to the high diversity of content within the 360 degrees and the intricate interactions between content and distortions [13]. Moreover, some processing, such as compression, may create impairments that are inequally distributed over the sphere. Consequently, the use of MOS as labels for sampled patches is somewhat questionable. Finding a way to label the patches that cope with the above drawbacks is important to reaching reliable results.

This work is funded by the Nouvelle Aquitaine regional Council under projects SIMOREVA360 2018-1R50112 and PERCEIVE.

Given that deep learning is used in the majority of current attempts for 360-degree IQA, input images are encoded into latent spaces, resulting in deep neural features that are regressed to quality scores. These features are obtained using backbones, usually pre-trained convnets such as ResNets [14]. The features' encoding is an entrusted task, where the resulting features are used as they are to derive the quality scores. Still, the resulting features may contain some noise and irrelevant information, particularly since the backbones used are trained for other tasks. Refining these features, by selecting the most prominent ones according to quality labels, would boost robustness. This is overlooked in existing approaches for IQA in general, and 360-degree in particular.

In this paper, we explore efficient ways to train patch-based 360-degree IQA models, with a focus on local quality labeling and label-specific feature selection. A framework is designed to leverage deep neural network feature selection and adaptive patch labeling (APL). First, we introduce the concept of APL by exploiting the rich state-of-the-art 2D IQA. This technique involves dividing the 360-degree image into local patches and assigning quality labels as a quality distribution to each patch based on its content. Several no-reference 2D IQA models are used to generate a quality distribution for each patch. By employing APL, we address the inherent variability in image quality across different regions of the panoramic scene. Then, a feature selection is performed on encoded features obtained using the ResNet-50 model. The selection is conducted using mFILS [15], an algorithm that adaptively select and refine the most useful feature sets from encoded features by means of convenets. To evaluate the proposed study, we collect two datasets of 360-degree images encompassing various visual content and quality levels. The proposed study strives to enhance the accuracy and effectiveness of IQA models for evaluating the quality of 360-degree content.

II. Proposed Study

The proposed study focuses on exploring efficient methods for training patch-based 360-degree IQA models. The key objectives include local quality labeling and label-specific feature selection to improve the accuracy.

A. Adaptive patch labeling (APL)

We employed a set of m 2D-IQA models to serve as virtual observers, simulating the process of subjective experiments where observers provide their opinion scores. The selected 2D-IQA models, as listed in Table I, represent a diverse range of state-of-the-art approaches commonly used in the field, ranging from traditional to deep-learning-based approaches. By processing patches extracted from 360-degree content as 2D images, we leverage the extensive knowledge and advancements in 2D-IQA, allowing us to generate quality score distributions at patch-level. This approach enables us to capture the variability and nuances in image quality within the 360-degree scene, providing a more comprehensive and accurate representation of the patch quality.

TABLE I: 2D-IQA models for estimation of the quality score distributions at patch level.

Type	Model
Traditional	NIQE, PIQUE, BRISQUE, BIQAA [16], BLIINDS [17], CPBD [18], FADE [19]
Deep learning	DB-CNN [13], MUSIQ [20], TRES [21], NIMA [22], PaQ-2-PiQ [23]

The structure of the proposed adaptive patch labeling (APL) framework is illustrated in Fig. 1. The framework involves the application of m 2D-IQA models to evaluate and label each patch P_i. This process generates a quality score distribution (QSD), consisting of m individual scores. The steps involved in this process can be described as follows:

- Sampling n patches from each 360-degree image I. In this study, we specifically focused on extracting patches from the Equirectangular projection (ERP) content of the image. This choice was made to mitigate the introduction of any sampling biases that could potentially arise from alternative patch sampling methods.
- Application of 2D-IQA models, where each P_i is fed into m 2D-IQA models, which individually assess and label the quality of the patch. Each 2D-IQA model produces a quality score for P_i, resulting in m individual scores.
- Combining the generated labels to form QSD. This distribution captures the variability in quality provided by different 2D-IQA models for the given patch P_i.

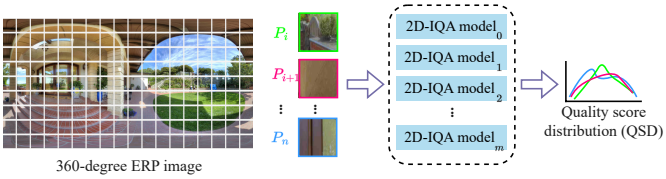

Fig. 1: Patch sampling and labeling using the adaptive patch labeling (APL).

By employing the APL, we can capture multiple perspectives on the quality of each patch within the 360-degree image by objectively generating the $QSDs$. In contrast, MOS is obtained through subjective evaluations by human observers, where each has explored the scene and based his opinion on the perceived quality. This means that each portion of the scene may contribute to the observer's judgment. Combining QSDs with MOS allows for a more comprehensive and accurate representation of patch quality. Therefore, the label Y_i associated with P_i is obtained as:

$$Y_i = \{MOS_I, norm(QSD_i)\}, \quad (1)$$

where $norm(.)$ is a normalization function employed to standardize the elements constituting QSD_i onto a uniform scale.

B. Features selection

Feature selection plays a crucial role in developing accurate IQA models by identifying and retaining the most relevant features. We believe that incorporating adaptive feature selection techniques, specifically designed for refining encoded features from deep neural networks, can significantly enhance the accuracy of IQA models. The process of feature selection involves choosing a subset of features that are highly discriminative and have a significant impact on quality prediction. In our approach, we consider label-specific feature selection algorithms, which are particularly effective in scenarios where each label correlates with the encoded features. We have used and adapted a generic framework, mFILS [15], which allows the tri-selection of features, instances, and labels. mFILS adaptively selects and refines the most useful feature sets from the encoded features, eliminating potential noise and irrelevant information that may hinder the accuracy of the IQA model. This process is illustrated in Fig. 2. As we are only interested in feature selection in our context, we have adapted the mFILS framework for the feature selection task while taking into consideration the latent semantics of the multi-label information.

Let $\mathbf{X} \in \mathbb{R}^{C \times D}$ be a set of features representing a dataset content, and $\mathbf{Y} \in \mathbb{R}^{C \times \bar{m}}$ their corresponding labels. Here, $\mathbf{X}$ is generated using the ResNet-50 [14] and $\mathbf{Y}$ are obtained by means of the previously describe APL framework. In order to select the best features, we have to minimize the following problem[1]:

$$\min_{\mathbf{W}} \|\mathbf{XW} - \mathbf{V}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1}, \tag{2}$$

where λ is a regularization hyper-parameter, $\mathbf{W} \in \mathbb{R}^{D \times k}$, with $k << \bar{m}$ and $\bar{m} = m + 1$, represents the feature coefficient matrix and $\mathbf{V} \in \mathbb{R}^{C \times k}$ is a low-dimensional latent semantics matrix, which can be obtained by the decomposition of the label space $\mathbf{Y}$ by (3).

$$\min_{\mathbf{V} \geq \mathbf{0}, \mathbf{B} \geq \mathbf{0}} \|\mathbf{Y} - \mathbf{VB}\|_F^2 \tag{3}$$

Once $\mathbf{W}$ is learned, the D features can be sorted and selected according to the $\ell_{2,1}$-norm of the D^{th} rows of $\mathbf{W}$.

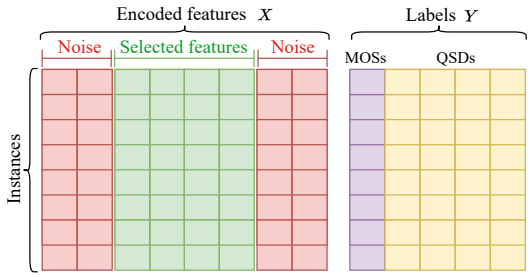

Fig. 2: Illustration of features selection according to the assigned labels using the mFILS [15] framework.

[1] $\|\cdot\|_F$ and $\|\cdot\|_{2,1}$ are Frobenius and $\ell_{2,1}$ norms respectively.

C. Multi-regression training

In order to estimate a distribution, a multi-regression training strategy is required. This allows the model to learn a combined representation of global and local quality by regressing the feature representation X_i to Y_i. Therefore, features are fed into a regression block, which consists of two fully connected layers. A first FC layer with 512 neurons followed by a Rectified Linear Unit (ReLU) activation function, and a second one with two $m+1$ neurons, where m represents the elements on $QSDs$. The architecture of the training strategy is visualized in Fig. 3, showcasing the regression block.

During training, the regression block is optimized to minimize a loss criterion across the multi-outputs. The $L1$ loss is used in this study.

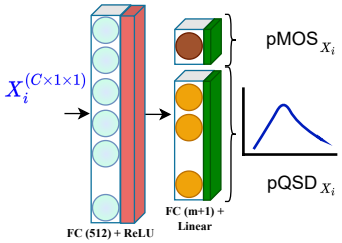

Fig. 3: Structure of the used regression block. For each X_i associated with P_i, multiple outputs are generated, including predicted MOS and QSD.

D. Quality scores pooling

In a patch-based model for IQA, a quality score pooling stage is essential to derive a single representative score for the entire image. In our approach, we adopt a double-stage pooling process to handle the predicted distributions for each patch. Firstly, the scores composing the patch $pQSD$ are pooled together using a pooling function $f(.)$. This pooling operation combines the individual scores within the $pQSD$ to generate a single score representation. Secondly, the resulting pooled scores from the $pQSD$ are added to the predicted $pMOS$. This addition takes into account both the aggregated patch-level quality scores and the overall perception of quality captured by the $pMOS$. Finally, a second pooling operation is performed on the combined scores to obtain the final quality score for the entire image I. This second pooling consolidates the information from the $pQSD$ and $pMOS$ into a single score that represents the overall quality of the image, as illustrated in Fig. 4. By using this double-stage pooling process, we effectively combine the patch-level quality information with the global perception captured by the $pMOS$, enabling us to derive a comprehensive and representative quality score for the 360-degree image.This is achieved as follows:

$$pMOS_I = f(pMOS_i + (\alpha \times f(pQSD_i))) \tag{4}$$

For this study, we use three pooling strategies, including average (Avg), Percentile (Perc.) and agreement-based pooling (AP) [24]. Percentile pooling emphasizes the lowest quality

scores among the set of scores based on a threshold k, known as the $k-$th percentile. Here, K is set to 75%. AP implements the outlier rejection paradigm, where scores falling outside an agreement range are discarded, considering only scores that agree and depict less variability.

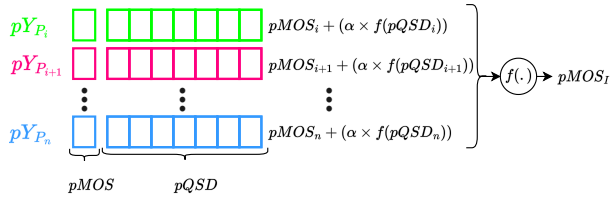

Fig. 4: Quality scores pooling: from local to global quality

III. Experimental results

Datasets: Two benchmark 360-degree IQA datasets, including OIQA [25] and CVIQ [3] are used to validate the performances. OIQA includes 320 distorted 360-degree images derived from 16 pristine ones. It covers four distortions, including JPEG compression (JPEG), JPEG 2000 compression (JP2K), Gaussian blur (BLUR), and Gaussian white noise (WN). Each distortion type is applied at five different levels of distortion. CVIQ focuses only on compression-related distortion, generated by JPEG, H.264/AVC (AVC) and H.265/HEVC (HEVC). These compressions are applied to 16 360-degree images with eleven levels each.

Implementation Details: The validation of the proposed study is conducted on a computer equipped with 32GB of RAM and an Nvidia Quadro T2000 MAX-Q 4GB GPU. A 4-fold cross-validation is used within each of the selected datasets. Each fold is trained for 100 epochs. We split the datasets into $80\%/20\%$ for training and testing, respectively. To ensure complete separation of the training and testing sets, the distorted samples associated with the same pristine image are allocated to the same set.

Evaluation Criteria: Three recommended evaluation metrics are used, including the Pearson linear correlation coefficient (PLCC) to evaluate accuracy, the Spearman rank-order correlation coefficient (SRCC) for monotonicity, and the root mean squared error (RMSE) to evaluate prediction errors.

A. Performance evaluation

Table III summarizes the performances on OIQA and CVIQ of the considered strategies (*see* Table II) using three pooling techniques, namely average pooling (Avg), percentile pooling (Perc.), and agreement-based pooling (AP). These pooling techniques aggregate and summarize the predictions in different ways. The choice of pooling technique can influence the models' overall performance and their ability to capture specific aspects of image quality.

The first observation that emerges is that the performance of the models varies across different versions and pooling strategies. Specifically, when evaluating on the OIQA dataset, versions 4 and 5 exhibit higher PLCC/SRCC values compared to the other versions. It is worth mentioning that these versions utilize feature selection, which contributes to a stronger linear correlation with the quality regression task. Additionally, the feature selection process is conducted based on the quality labels assigned to each patch, using the APL framework. This approach enables the selection and refinement of encoded features, focusing only on those that contribute to more accurate predictions. A notable improvement of up to 2.5% can be observed with versions 4 and 5 compared to the other versions, independently of the used pooling strategy. By incorporating quality labels and feature selection, the PLCC/SRCC values significantly increased from 0.9468/0.9406 (when considering only the MOS for model training) to 0.9599/0.9549.

TABLE II: Labels for the considered training strategies with respect to the scores pooling methods.

Version	training		Scores pooling		Feature selection
	MOS	QSD	pMOS	pQSD	
1	✓		✓		
2	✓	✓	✓		
3	✓	✓	✓	✓	
4	✓	✓	✓		✓
5	✓	✓	✓	✓	✓

TABLE III: Performance evaluation on OIQA and CVIQ. The median over 4-fold is provided and the best performances are highlighted as Avg, Perc., and AP.

Version	Pooling	OIQA			CVIQ		
		PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
1	Avg	0.9468	0.9288	0.0469	0.9531	0.9392	0.0419
	Perc.	0.9421	0.9278	0.0488	0.9482	0.9301	0.0439
	AP	0.9550	0.9453	0.0422	0.9541	0.9408	0.0415
2	Avg	0.9444	0.9336	0.0474	0.9653	0.9559	0.0361
	Perc.	0.9382	0.9228	0.0496	0.9626	0.9518	0.0375
	AP	0.9447	0.9318	0.0472	0.9654	0.9543	0.0361
3	Avg	0.9429	0.9315	0.0477	0.9639	0.9537	0.0368
	Perc.	0.9421	0.9330	0.0478	0.9609	0.9472	0.0383
	AP	0.9461	0.9362	0.0464	0.9634	0.9510	0.0371
4	Avg	0.9576	0.9534	0.0412	0.9542	0.9289	0.0415
	Perc.	0.9543	0.9499	0.0427	0.9506	0.9261	0.0432
	AP	0.9597	0.9535	0.0402	0.9549	0.9321	0.0412
5	Avg	0.9599	0.9549	0.0401	0.9540	0.9285	0.0416
	Perc.	0.9596	0.9539	0.0403	0.9532	0.9270	0.0420
	AP	0.9623	0.9577	0.0389	0.9548	0.9312	0.0413

On the same dataset, it is clear that versions 2 and 3 demonstrate a decrease in performance compared to version 1, irrespective of the pooling strategy employed. Surprisingly, this decline in performance is observed even when using the generated quality labels to train a multi-regression model. The primary objective of using a quality distribution as labels is to enhance the representativeness of the patch quality. However, the evaluation on OIQA reveals that the performances slightly dropped, indicating that the use of quality labels alone is not sufficient. It is only when combined with feature selection and refinement that the performances showed improvement. Furthermore, it is worth noting that pooling together the

predicted MOS and QSD in version 5 of the model resulted in further improved accuracy compared to considering only the predicted MOS in version 4. This suggests that incorporating the additional information provided by the QSD enhances the models' ability to accurately assess image quality. By combining both the inherited quality (MOS) and the assigned one (QSD), version 5 achieves outperformed the others.

Regrading the evaluation on CVIQ, it is observed that using only the MOS as the basis for training and evaluation yielded the poorest performance. However, the best performance was achieved with version 2 of the model, followed by version 3. This indicates that the generated quality labels contribute to enhancing the accuracy of the IQA model compared to relying solely on the MOS. This observation aligns with the challenges associated with using MOS as the sole metric for labeling individual patches. The introduction of quality labels, derived from the generated quality distribution, addresses the limitations of relying solely on MOS and provides more nuanced and detailed information about image quality. By incorporating these quality labels into the training and evaluation process, versions 2 and 3 demonstrate improved performance, highlighting the importance of leveraging quality labels to accurately assess image quality of 360-degree images. However, it is worth noting that the adopting feature selection on CVIQ led to a slightly worse performance compared to using 100% of the encoded features. This implies that in the specific case of the CVIQ dataset, the feature selection did not provide significant improvements in the model's accuracy. While feature selection is generally considered a valuable technique for reducing dimensionality and enhancing model performance, its effectiveness can vary depending on the dataset and the specific characteristics of the features being selected. In the case of CVIQ, it appears that using the full set of encoded features without selection yielded better results than employing a feature selection process. This observation suggests that the encoded features obtained from ResNet-50 encompass essential information for accurately assessing image quality. Hence, the additional step of feature selection did not provide substantial benefits in this particular context.

Through the analysis of the pooling strategies' impact on the achieved performances, it is evident that the AP strategy demonstrates the highest overall performance. The unique characteristic of AP is its consideration of the distance between each score and the median. This approach proves to be particularly effective in improving the accuracy of the IQA model's performance. Furthermore, the use of AP also captures the variability among predicted scores. A lower variability indicates a higher level of agreement among the scores. This aligns with the processing of human opinions, where the objective is to consider only opinions that agree according to specific criteria in order to establish consistent judgments.

B. Ablation study

To further investigate the significance and impact of feature selection rates and the α parameter on the model's performance, an ablation study was conducted. The purpose of this study is to analyze and assess the effects of varying these factors on the overall performance of the model. This section presents the findings of the ablation study, providing valuable insights into each component.

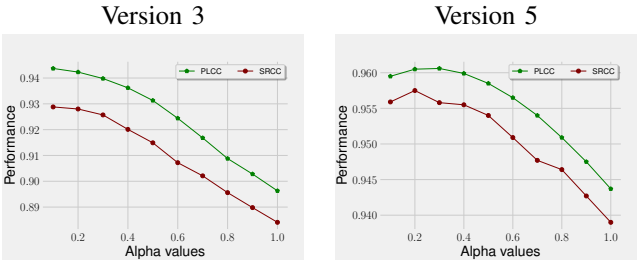

Fig. 5: Effect of the parameter α on the performance of the scores pooling.

Fig. 5 illustrates the accuracy (PLCC) and monotonicity (SRCC) performances of the Avg pooling strategy on the OIQA dataset, considering different values of α within the range of $\{0.1, 0.2, ..., 1\}$, as described in Sec. II-D. The results demonstrate that for version 3 of the model, the best performances are achieved with the lowest value of α, specifically $\alpha = 0.1$. For version 5, the best performances are observed with α values within the range of $\{0.2, 0.3\}$. These findings indicate that the optimal value of α varies depending on the specific version of the model, underscoring the importance of selecting appropriate values for α to maximize the performance of the IQA model. Furthermore, an interesting observation from these results is that increasing the value of α leads to a decrease in performance for version 3 of the model. This finding suggests that the use of the $pQSD$ does not contribute to the accuracy of the predictions for this particular version. It indicates that the information provided by $pQSD$ may not be as crucial or influential in improving accuracy or monotonicity. In contrast to version 3, an improvement can be observed for version 5 of the model when using α values up to 0.3. This highlights the effectiveness of incorporating the predicted $pMOS$ along with $QSDs$ in the feature selection process. By leveraging $pMOS$ and $QSDs$, the model is able to select and refine the most suitable features for the multi-regression task, resulting in improved performance.

To evaluate the effectiveness of feature selection on the OIQA dataset, we further investigated the impact of the selection rate on the model's performance. In this analysis, we varied the percentage rate of feature selection from 90% down to 30%, and the corresponding results are depicted in Fig. 6. This figure provides insights into how the selection rate affects the accuracy of the model's predictions. By examining the performance trends across different selection rates, we can gain a deeper understanding of the optimal feature selection rate that maximizes the model's performance. As it can be seen, the results demonstrate a consistent trend where the PLCC/SRCC values tend to increase as the feature selection rate decreases, regardless of the evaluation strategy employed. Whether using only $pMOS$ (version 4) or incorporating both

$pMOS$ and $\alpha \times pQSD$ (version 5), the performance metrics exhibit an upward trend with lower feature selection rates. This indicates that a lower percentage of retained features leads to a better overall performance. Therefore, selecting a lower rate of features during feature selection proves to be more effective in improving the overall performances.

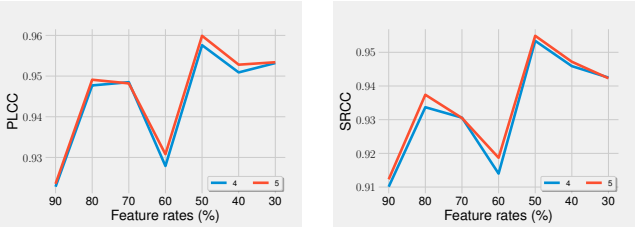

Fig. 6: Effect of feature rate selection on the performance of versions 4 and 5 on OIQA.

The best performances are achieved when selecting 50% of the encoded features, which corresponds to half the amount of features that need to be regressed to quality scores. Aprox. 5% and 4% gain compared to using 90% of features are achieved in terms of PLCC and SRCC, respectively. This finding suggests that a careful and efficient feature selection process can significantly improve the accuracy of the IQA model's predictions. By retaining a reduced but representative subset of features, the model can effectively capture the essential information necessary for accurate quality score regression.

IV. Conclusions

In conclusion, the findings of this study highlight the importance of local quality labeling and label-specific feature selection in training patch-based 360-degree IQA models. By leveraging deep neural network feature selection and adaptive patch labeling techniques, the model can effectively address the inherent variability in image quality across different regions of 360-degree scenes. The evaluation on both the OIQA and CVIQ datasets revealed that using generated quality labels, such as quality score distributions, enhances the accuracy of the IQA models compared to relying solely on MOS. This demonstrates the value of considering agreed opinions and capturing the variability among predicted scores, similar to how human opinions are processed in psychophysical studies. Furthermore, the results showed that feature selection plays a crucial role in improving the performance of the IQA models. The analysis of different feature selection rates indicated that selecting a lower rate of features leads to better overall performance, achieving higher PLCC and SRCC values. The optimal feature selection rate was found to be 50%, representing half the amount of features that need to be regressed to quality scores. Overall, these findings contribute to the development of efficient patch-based 360-degree IQA models and provide insights into the impact of pooling strategies, feature selection, and the utilization of quality labels on the performances. This knowledge can guide future research and improve the accuracy and reliability of IQA models for 360-degree images.

References

[1] A. Perkis, C. Timmerer, S. Baraković, J. Husić, and *et al.*, "Qualinet white paper on definitions of immersive media experience (imex)," *ENQEMSS, 14th QUALINET meeting (online)*, 2020.

[2] W. Zhou and B. Alan, "Modern image quality assessment," *Syn. Lectures on IVMP*, vol. 2, no. 1, pp. 1–156, 2006.

[3] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan, and S. Ma, "MC360IQA: A multi-channel CNN for blind 360-degree image quality assessment," in *IEEE JSTSP*, 2020, vol. 14, pp. 64–77.

[4] TQ. Truong, TH. Tran, and TC. Thang, "Non-reference quality assessment model using deep learning for omnidirectional images," in *IEEE ICAST*, Morioka, Japan, 2019, pp. 1–5.

[5] L. Yang, M. Xu, X. Deng, and B. Feng, "Spatial attention-based non-reference perceptual quality prediction network for omnidirectional images," in *IEEE ICME*, Shenzhen, China, 2021, pp. 1–6.

[6] Y. Zhou, Y. Sun, L. Li, K. Gu, and Y. Fang, "Omnidirectional image quality assessment by distortion discrimination assisted multi-stream network," *IEEE TCSVT*, pp. 1–1, 2021.

[7] A. Sendjasni and MC. Larabi, "SAL-360IQA: A saliency weighted patch-based cnn model for 360-degree images quality assessment," in *IEEE ICMEw*, Taipei City, Taiwan, 2022, pp. 1–6.

[8] N. Tofighi, MH. Elfkir, N. Imamoglu, C. Ozcinar, E. Erdem, and A. Erdem, "ST360IQ: No-reference omnidirectional image quality assessment with spherical vision transformers," in *ICASSP*, Rhodes Island, Greece, 2023, pp. 1–5.

[9] C. Zhang and S. Liu, "No-reference omnidirectional image quality assessment based on joint network," in *30th ACM ICM*, Lisboa, Portugal, 2022, p. 943–951.

[10] M. Zhou, L. Chen, X. Wei, X. Liao, and Q. Mao *et al.*, "Perception-oriented U-Shaped transformer network for 360-degree no-reference image quality assessment (*Early access)*," *IEEE TBC*, pp. 1–10, 2023.

[11] J. Gutiérrez, P. Pérez, M. Orduna, and A. Singla *et al.*, "Subjective evaluation of visual quality and simulator sickness of short 360° videos: ITU-T Rec. P.919," *IEEE TMM*, vol. 24, pp. 3087–3100, 2022.

[12] A. Sendjasni, *Objective and subjective quality assessment of 360-degree images*, Ph.D. thesis, Univ. de Poitiers; NTNU, 2023.

[13] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE TCSVT*, vol. 30, no. 1, pp. 36–47, 2020.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, Las Vegas, NV, USA, 2016, pp. 770–778.

[15] DEK. Mansouriand, K. Benabdeslem, SF. Benkabou, S. Chaib, and M. Chohri, "mFILS: Tri-selection via convex and nonconvex regularizations," *IEEE TNNLS*, pp. 1–14, 2023.

[16] G. Salvador and C. Gabriel, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A*, vol. 24, no. 12, pp. B42–B51, 2007.

[17] MA. Saad, AC. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE TIP*, vol. 21, no. 8, pp. 3339–3352, 2012.

[18] N. Narvekar and L. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *IEEE QoMEX*, San Diego, USA, 2009, pp. 87–91.

[19] LK. Choi, J. You, and AC. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE TIP*, vol. 24, no. 11, pp. 3888–3901, 2015.

[20] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in *IEEE ICCV*, Montreal, Canada, 2021, pp. 5128–5137.

[21] A. Golestaneh, S. Dadsetan, and K. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *IEEE/CVF WCACV*, 2022, pp. 3209–3218.

[22] H. Talebi and P. Milanfar, "Nima: Neural image assessment," *IEEE TIP*, vol. 27, no. 8, pp. 3998–4011, 2018.

[23] Z. Ying and H. Niu *et al.*, "From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality," in *IEEE CVPR*, Seattle, USA, 2020, pp. 3572–3582.

[24] A. Sendjasni and MC. Larabi, "PW-360IQA: Perceptually-weighted multichannel cnn for blind 360-degree image quality assessment," *Sensors*, vol. 23, no. 9, 2023.

[25] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual Quality Assessment of Omnidirectional Images," in *IEEE ISCAS*, Florence, Italy, 2018, pp. 1–5.