Evaluating LIQE a BIQA model on PIQ a portrait image quality dataset

Joshua Sherwood

April 25 2 24

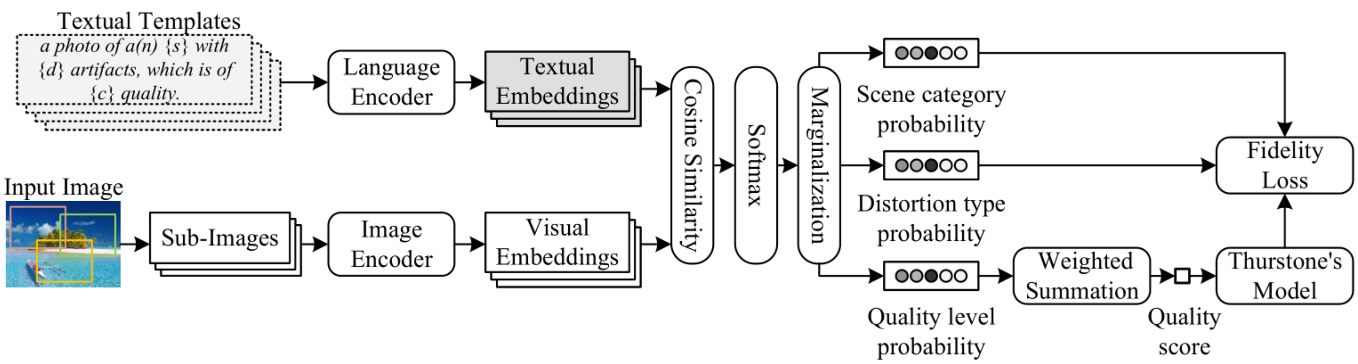

Figure 1: The LIQE model

Introduction

Quality computer vision requires quality image datasets. But what makes an image one of quality? Subjective opinions are not deterministic[Chahine et al.(2023)], so it is impossible to perfectly rate the quality of an image. Photos are taken in a variety of scenes (scene = fixed visual content), which a ect perceived quality for reasons including viewing conditions, the observer's cultural and professional backgrounds, and image content[Chahine et al.(2023)]. Blind image quality assessment (BIQA) must work to overcome these barriers to accurately predict the human perception of image quality without any reference information[Zhang et al.(2023)]. Accurately predicting perceived image quality is a current research topic with many potential use cases, including marketing, social media, and image generation. All of these applications could benefit from portrait quality assessment, so BIQA models should be evaluated in part on their ability to determine portrait image quality.

The Portrait Image Quality (PIQ) dataset was developed with the help of over 30 image quality experts who performed pair-wise comparisons based on specified criteria for face detail preservation, face target exposure, and overall portrait image quality. They did this in a controlled setting, with images from 100 smartphone devices across 14 brands and various camera lens focal lengths. The images they annotated contain portraits of various ethnicities and genders. The resulting dataset was the first-of-its-kind smartphone portrait quality dataset[Chahine et al.(2023)], and o ers a unique opportunity to validate BIQA models.

The Language-Image Quality Evaluator (LIQE), a model presented at the Conference on Computer Vision and Pattern Recognition (CVPR) 2023, predicts the human perception of image quality without any reference information(BIQA), taking an image and outputting a quality score. At time of release (27 March 2023), it scored state-of-the-art (SoA) on several BIQA datasets, including LIVE, CSIQ, KADID-10k, BID, CLIVE, and KonIQ-10k. This project evaluates the LIQE model on the PIQ dataset.

2 Method

The PIQ dataset was downloaded from the DXOMARK website, and includes source images and quality labels for each of the 5116 images. These labels include a Just-Objectionable-Di erence(JOD) score, which the PIQ paper used to produce a Spearman's rank correlation coefficient (SRCC) score for existing BIQA models. SRCC is used in evaluating quality labeling because it evaluates the order in which images are ranked, and does not directly consider the quality number assigned to the image. To evaluate LIQE, quality scores were produced for every image in the PIQ dataset with the PyTorch implementation of LIQE. This implementation uses the same model as described in the paper, but retrains on KonIQ-10, according to the LIQE GitHub repository. Then, the SRCC metric was obtained by comparing the image quality rankings produced by LIQE with the ground truth, JOD scores. This was done with the spearmanr function from the scipy.stats Python module. Along with an overall SRCC score for all images in the PIQ dataset, scores were produced for each of the four scene types included in PIQ (indoor, outdoor, low light, night). This can be easily reproduced by downloading the PIQ dataset and using the LIQE PyTorch model (which by default is not trained on PIQ, as of April 25 2024).

3 Conclusion

The results are shown in Figures 2 and 3. LIQE performs similarly to other top BIQA models on PIQ, but has a larger standard deviation. Interestingly, LIQE performs better for some light levels, as shown in Figure 2.

LIQE trains on image-label pairs, where the labels consist of a scene $\in$ {”animal”, ”cityscape”, ”human”, ”indoor scene”, ”landscape”, ”night scene”,

Method	Overall	
BRISQUE	0 192	
ILNIQE	0 214	
NIQE	0 298	
LIQE	0 577	0 20
DB-CNN	0 555	0 07
MUSIQ	0 589	0 07
HyperIQA	0 611	0 06
SEM-HyperIQA	0 621	0 06
SEM-HyperIQA-CO	0 621	0 07
SEM-HyperIQA-SO	0 642	0 08

Figure 2: Median SRCC results for models tested on PIQ for the Overall category, taken from the PIQ paper, with LIQE added

Scene	Overall
Outdoor	0 45
Indoor	0 51
Low Light	0 62
Night	0 72

Figure 3: LIQE performance on each scene type in PIQ

”plant”, ”still-life”, ”others”}, a distortion type ∈ {”blur”, ”color-related”, ”contrast”, ”JPEG compression”, ”JPEG2000 compression”, ”noise”, ”over-exposure”, ”quantization”, ”under-exposure”, ”spatially-localized”, ”others”}, where the ”others” category includes pristine quality images, and a quality score ∈ {”bad”, ”poor”, ”fair”, ”good”, ”perfect”}. The images are encoded using ViT-B/32 and the labels with GPT-2. When evaluated on images without a label, LIQE first encodes sub-images of the input image, since the CLIP model used only accepts input images with a fixed spatial size(224 x 224 x 3)[Zhang et al.(2023)]. It then outputs a quality score based on which labels the visual embedding is closest to, considering all sub-images. This is a simplification of the process, the full process involves math beyond my understanding, and the analysis is accordingly limited. The processing of sub-images may be one of the reasons LIQE does not perform SoA on the PIQ dataset, which has a specified region of interest[Chahine et al.(2023)].

LIQE does not rank images with quality scores that closely align with the JOD provided by the PIQ dataset (the LIQE SRCC being 0.577, and the SoA being 0.642 with less standard deviation). Explaining the reasons behind this is beyond the scope of this project. Regardless, LIQE does not appear to be ready for the applications of BIQA mentioned in the introduction, given its lack of SoA performance on the PIQ dataset.

References

[Chahine et al.(2023)] Nicolas Chahine, Stefania Calarasanu, Davide Garcia-Civiero, Théo Cayla, Sira Ferradans, and Jean Ponce. 2023. An Image Quality Assessment Dataset for Portraits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR).* 9968–9978.

[Zhang et al.(2023)] Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. 2023. Blind Image Quality Assessment via Vision-Language Correspondence: A Multitask Learning Perspective. In *IEEE Conference on Computer Vision and Pattern Recognition.* 14071–14081.