No-Reference Image Quality Assessment for Intelligent Sensing Applications

Zhuobin Yuan, Ademola Ikusan, Rui Dai
Department of Electrical and Computer Engineering
University of Cincinnati
Cincinnati, OH, USA
{yuanzn,ikusanaa}@mail.uc.edu, rui.dai@uc.edu

Junjie Zhang
Department of Computer Science and Engineering
Wright State University
Dayton, OH, USA
junjie.zhang@wright.edu

NAECON 2024 - IEEE National Aerospace and Electronics Conference | 979-8-3503-6762-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/NAECON61878.2024.10670634

***Abstract*—Many intelligent sensing systems rely heavily on automatic analysis tools to extract high level information from the raw videos or images captured by cameras. In particular, deep-learning-based computer vision solutions have shown promising results in analysis tasks ranging from image segmentation to object detection and recognition. In practical systems, image distortions due to factors such as noise and blur may degrade the accuracy of these analysis tools. This paper proposes a no-reference image quality assessment model for predicting the quality of images from the perspective of three major computer vision tasks: image segmentation, image classification, and object detection. A data set is constructed that considers distortions including noise, blur, and bad lighting, which commonly occur during the image acquisition process in diverse applications. Three widely used deep-learning-based algorithms are considered to label the quality of the images in the dataset. A set of light-weight features are extracted to characterize the structure of the content in an image. Based on the data set and the extracted features, a classification model is built to predict the quality of images used in computer vision tasks. Experimental results show that the proposed model offers more accurate predictions than common image quality measures such as BRISQUE, NIQE, and PIQE.**

***Index Terms*—image quality assessment, no-reference, computer vision, image classification, object detection, image segmentation**

I. Introduction

Cameras have become vital components in a variety of intelligent sensing applications such as intrusion management, crowd detection, traffic monitoring, and augmented reality. Many of these applications rely heavily on deep-learning-based computer vision (CV) tools that can automatically analyze the images captured by cameras and extract high level information from them. Due to environmental or human factors, the sensed images may suffer from different types of distortions like noise, blur, or bad light. The accuracy of a CV algorithm could degrade if the quality of an input image is not satisfactory. Therefore, it is necessary to assess how and to what extent image distortions affect the performance of common CV tools.

In the field of image quality assessment (IQA), there are extensive studies on modelling the perceptual image quality which is evaluated by human users [1]–[3]. Traditional perceptual quality assessment methods take advantage of known characteristics of the human visual system (HVS) such as luminance sensitivity, color perception, spatial resolution [4]. However, the quality of an image evaluated by a computer vision algorithm is not necessarily sensitive to the same factors that drive human perceptions. In our recent work on object detection quality [5], we found that the performance of classical object detection algorithms could be influenced by the quality of background, whereas human beings can easily focus on a moving object even with a blurred background. It has also been shown in [6] that some characteristics of the HVS are useless for CNN-based methods of computer vision tasks. For example, images are typically presented to CNNs as a static rectangular pixel grid with fixed spatial resolution but the primate eye has an eccentricity dependent spatial resolution.

There are a few studies on the problem of quality evaluation for different computer vision tasks. For example, five quality factors, including contrast, brightness, focus, sharpness, and illumination, were used to evaluate the performance of face recognition [7]. The degradation of the performance of face detectors was quantified considering different factors including noise, blur, and compression in [8]. An image quality prediction model for object detection was proposed based on features like image gradient, edge, and estimated object size [9]. For target tracking, the image quality for tracking in airborne reconnaissance systems was studied in [10], and it has been found that the accuracy of target detection is impacted by factors such as jitter, level of noise, and edge sharpness, but it is less sensitive to spatial resolution.

In this paper, we aim to advance existing studies by tackling the challenge of building a more general quality prediction model for a wide range of intelligent sensing applications considering common types of distortions that may occur. We propose to study three representative computer vision methods: image classification, object detection, and image segmentation, because almost all of the existing image analysis tasks are based on at least one of these three methods. We consider image distortions caused by noise, blur, and bad lighting. We propose a no-reference image quality assessment model based on local features in an image such as edge as well as global features like contrast and estimated object size.

This work was supported by the National Science Foundation under Grant No. TI-2234596.

(a) Gaussian Noise (b) Gaussian blur (c) Motion blur

(d) Lens blur (e) Brighten (f) Darken

Fig. 1: Samples of distortions.

The model is trained using a large number of images with different distortions considering three representative deep-learning-based computer vision algorithms. The accuracy of the proposed model is then evaluated on a separate test data set and compared with commonly used IQA measures.

II. Dataset and Performance Metrics

A. Dataset Generation

We have generated a distorted images dataset based on three different datasets, ImageNet [11], PASCAL VOC2012 [12], and COCO [13], corresponding to the three CV tasks being considered. We selected 685 correctly classified images which are used in the ImageNet challenge for image classification. We selected 1225 images from PASCAL VOC2012, which contains 20 object categories for object detection. For image segmentation, we chose 772 images from COCO, which is a large dataset for object detection, segmentation, and captioning published by Microsoft. For each of the 2,682 selected original images, we simulated 6 specific types of distortion: Gaussian noise, Gaussian blur, motion blur, lens blur, brighten, and darken. And for each distortion type, 5 distortion levels were simulated (low, mid-low, mid, mid-high, and high).

Samples of distorted images in our dataset are shown in Fig.1. Gaussian noise was added to the original images, where variances were set to be 0.001, 0.002, 0.003, 0.005, and 0.01. The 2D circularly symmetric Gaussian blur kernel was applied to generate the blurring effect of each image by using standard deviations of 0.1, 0.5, 1, 2, and 5. Motion blur was simulated to approximate the linear motion of a camera by 1, 2, 4, 6, and 10 pixels with an angle of 45 degrees. Circular averaging filter was used to simulate lens blur by adjusting filter radius to 1, 2, 4, 6, and 8, where a higher radius means a high level of lens blur. For the brighten and darken distortions, we non-linearly adjusted the luminance channel by keeping extreme values fixed and increasing or decreasing others. For brighten effects, 0.1, 0.2, 0.4, 0.7, and 1.1 were used for increasing from low to high distortion levels, and for darken effects, 0.05, 0.1, 0.2, 0.4, and 0.8 were used for decreasing from low to high distortion levels.

B. Deep-Learning-Based Computer Vision Tasks and Metrics

We considered three core tasks in computer vision: image classification, object detection, and image segmentation. These

(a) Object Detection (b) Image Classification (c) Image Segmentation

Fig. 2: Different computer vision tasks.

foundational tasks underpin almost all other computer vision processes. Specifically, given an image, it is crucial to segment the various parts of the scene, as illustrated in Fig. 2(c). Once the different segments are identified, it is essential to detect and isolate each meaningful object in the scene (localization) from the background, as shown in Fig. 2(a). Finally, classification, as shown in Fig. 2(b), allows us to build more complex tasks such as action recognition. We propose to study the following deep-learning-based methods because of their efficiency and popularity in the computer vision field:

- ResNet-50 for image classification [14];
- YOLOv3 for object detection [15];
- Mask-RCNN for image segmentation [16].

There are a variety of metrics for evaluating the performance of different CV tasks. Towards a general quality model, we chose a set of metrics that can generate normalized values ranging from 0 to 1.

For image classification, the accuracy measures the number of correct predictions over a total number of predictions:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}. \quad (1)$$

Apart from the accuracy of each class in an image, we are also interested in the probability values associated with the classification accuracy for each class. For correctly classified images, the confidence in the prediction is reflected in the probability score while the score for misclassified images is effectively set to zero, resulting in a *cTE* (classification evaluator) ranging from 0 to 1, which is given by

$$cTE = Accuracy * Probability\ Score. \quad (2)$$

For object detection, the evaluation metric is the mean average precision (*mAP*) score that is calculated by taking the mean AP (average precision) over all the classes in an image:

$$mAP = \frac{1}{N}\sum_{N}^{i=1} AP_i. \quad (3)$$

The performance of image segmentation can be evaluated using the intersection-over-union (*IoU*), also known as the Jaccard Index which basically determine the pixels common between the ground truth and the prediction divided by the total pixels present across both masks:

$$IoU = \frac{groundtruth\ \cap\ prediction}{groundtruth\ \cup\ prediction}. \quad (4)$$

To analyze the impact of the studied distortions on the performance of CV tasks, we applied the three chosen deep-learning-based methods on each image in our distorted dataset.

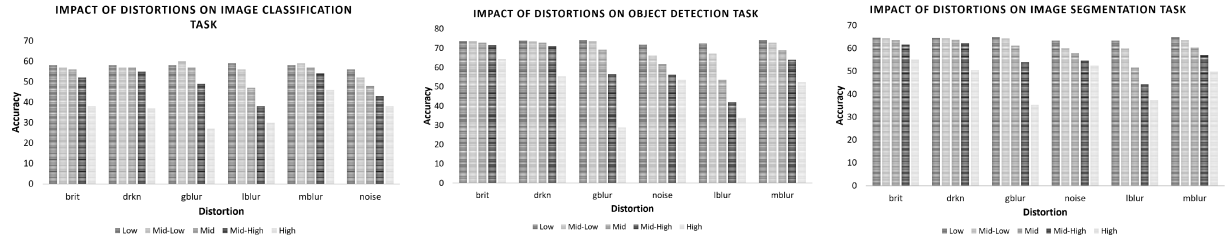

Fig. 3: Impact of distortions on the accuracy (%) of computer vision tasks.

The average accuracy (*cTE/mAP/IoU*) for different groups of data are shown in Fig. 3. It is evident that different types and levels of distortions can affect the performance of these tasks. The lower the distortion in an input image, the better the results for the computer vision tasks.

Each image was labeled as either "good" or "bad" to indicate its general usefulness for different CV tasks. The distribution of accuracy for the three CV tasks was considered, which is shown in Fig. 4. First, for the Mask-RCNN model used for image segmentation, when it was trained, it gave a positive response to an *IoU* of 0.5 and above when compared with the ground truth. Furthermore, we analyzed the distribution of accuracy values in our entire dataset, and we found that the images resulting in an *IoU* of 0.5 and above can consistently produce decent *mAP* values larger than 0.50 and *cTE* scores as high as 0.90 or above. Therefore, the images with $IoU \geq 0.5$, $mAP \geq 0.50$, and $cTE \geq 0.90$ were labeled as "good", and the rest of images in our dataset were labeled as "bad". Our goal was to predict the quality of an image as either "good" or "bad" without any prior knowledge of the type or the extent of distortions in it.

III. No-Reference IQA for Computer Vision Tasks

We introduce a classification model for assessing the quality of images for CV tasks. It operates directly on a possibly distorted image and falls into the category of no-reference IQA methods. The classification model is based on 11 features that describe the structure of the content in an image. These features fall into 8 categories: edge, image gradient, colorfulness, contrast, blur, brightness, resolution, and object size. The structure of the proposed model is shown in Fig. 5.

Edge: This represents the boundary information which plays a great role in segmenting, detecting, and observing patterns, and this was obtained using the Canny operator [17]. We propose to include three features in this category, using three thresholds for broad, moderate, and narrow edge analysis [17], respectively.

Image gradient: This is also a quality-contributing factor and it has its magnitude and direction. For an image $f(x, y)$, the gradient of f at location (x, y) is defined as the two dimensional column vector: $[\partial f/\partial x, \partial f/\partial y]^T$, where $\partial f/\partial x = f(x+1, y) - f(x-1, y)$, and $\partial f/\partial y = f(x, y+1) - f(x, y-1)$ using finite difference filters. The magnitude and direction of this gradient at location (x, y) are:

$$mag(x, y) = \sqrt{(\partial f/\partial x)^2 + (\partial f/\partial y)^2}, \tag{5}$$

$$dir(x, y) = tan^{-1}\left[\frac{\partial f/\partial y}{\partial f/\partial x}\right]. \tag{6}$$

The statistical properties of gradient could be used to depict the characteristics of an image. We calculate 2 features in this category: meanGmag (the average gradient magnitude) and meanGdir (the average gradient direction) [9].

Colorfulness: This is also an important indicator of image quality. To compute it, we can separate an image into its RGB color components and then calculate the difference between the red and green channels as well as the yellow-blue mask, which is given by

$$rg = R - G \tag{7}$$

$$yb = \frac{1}{2}(R + G) - B \tag{8}$$

We calculate the mean and standard deviation of each mask and then compute the overall mean and standard deviation, as outlined in (9) and (10). Using these values, we can determine the colorfulness according to (11) [18].

$$\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}. \tag{9}$$

$$\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2}. \tag{10}$$

$$\breve{I} = \mu_{rgyb} + (0.3 * \sigma_{rgyb}). \tag{11}$$

Contrast: We considered the RMS contrast of an image because it does not depend on the angular frequency content or the spatial distribution of contrast in an image. This is calculated as the standard deviation of the pixel intensities of an image, given by [19]:

$$I_{contrast} = \sqrt{\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(I_{ij} - \bar{I})^2}, \tag{12}$$

where $\bar{I}$ is the mean of the pixel intensities, I_{ij} is each pixel, and M and N indicate the resolution of the image.

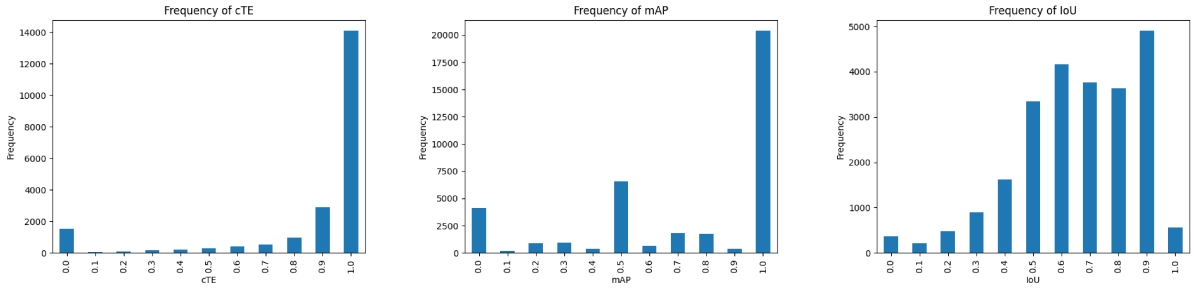

Fig. 4: Distribution of accuracy values for different computer vision tasks.

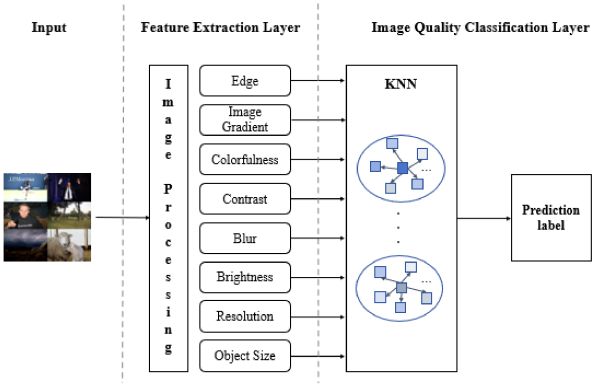

Fig. 5: Structure of the proposed model

Blur: The level of blur in an image can be represented by the reduction of the variance in gray level [20]. Given an image, I, the local point variance at (m, n) is given by

$$l_v(m,n) = \frac{1}{w_x w_y} \sum_{i}^{w_x} \sum_{j}^{w_y} \left[I(m+i, n+j) \hat{I} \right]^2, \qquad (13)$$

where $\hat{I}$ is the mean gray level defined as

$$\hat{I} = \frac{1}{w_x w_y} \sum_{i}^{w_x} \sum_{j}^{w_y} (I(m+i, n+j)), \qquad (14)$$

and w_x and w_y denote the size of a window centered on the point (m, n). The focus based on the local variance will be given by a global variance as follows:

$$VAR(I)_{blur} = \frac{1}{MN} \sum_{m}^{M} \sum_{n}^{N} \left[l_v - \hat{l_v} \right]^2, \qquad (15)$$

where $\hat{l_v}$ is given by

$$\hat{l_v} = \frac{1}{MN} \sum_{m}^{M} \sum_{n}^{N} l_v(m,n). \qquad (16)$$

Brightness: This is assessed by converting an image to grayscale and analyzing its histogram. The overall brightness can be calculated using the standard method in [21]. To find the brightness ratio, we divide each frequency by the total pixel count. The overall brightness is then calculated by scaling each ratio according to its intensity and summing the results.

Resolution: This is very crucial because when an image is resized to smaller dimensions, all objects shrink proportionally which can affect the accuracy of detection. The resolution is easily calculated by multiplying the image's width and height.

Object size: It could affect the extent to which an object can be separated from the background. We can apply the method in [22] to perform a quick estimation of object size. Initially, a contour-based spatial prior is extracted based on the layout of edges in the given image along a non-selective pathway. Then, local features such as color, luminance are gathered via a selective pathway. Lastly, Bayesian inference is used to auto-weight and integrate the local cues to predict the exact locations of objects.

We used the ensemble subspace KNN classifier [23] to train a model to predict image quality based on the extracted features. The algorithm basically chooses without replacement a random set of predictors from the possible total predictors, then trains a weak learner using the chosen predictors. This is done for the specified number of learners using the randomly chosen predictors. Then it predicts by taking an average of the prediction of the weak learners and classify the category with the highest average score. The tuning parameters include the number of learners, dimensions and learner types. The proposed model was chosen instead of a CNN-based classification model, because it just requires light-weight computation and it is more clear to interpret than black-box CNN-based models.

IV. Performance Evaluation

Our dataset was divided into a training set and a testing set for building and evaluating the proposed classification model. The entire dataset contains 83142 images, and 75% of the dataset (62357 images) were used for training and the rest 25% of them (20785 images) were used for testing. In the training process, 32-fold cross validation, 512 base learners of nearest neighbor learner type, and 8 subspace dimensions are set to train the ensemble subspace KNN classifier.

The classification performance of the proposed model on the test set is exhibited in the confusion matrix in Table I. It predicted 14260 samples of "good" labeled data out of a total of 15500 "good" samples and 4017 of "bad" labeled data out of 5285. The overall accuracy of classification on the test set

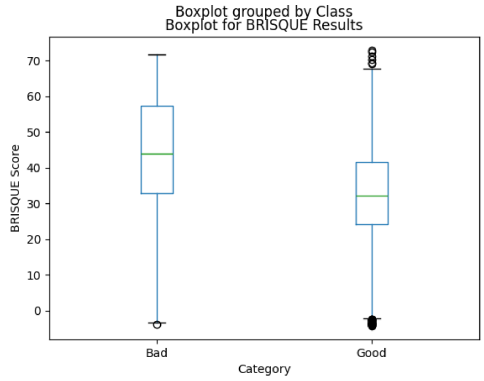

Fig. 6: Distribution of BRISQUE scores.

is 90.6%. The proposed model was also compared with three other popular no-reference IQA models including BRISQUE [1], NIQE [2], and PIQE [3]. Because these three models are regression models, we slightly modified them to classification models to enable fair comparison. Take BRISQUE for example, we generated a box diagram as shown in Fig. 6 to show the distribution of BRISQUE scores in different label types. As we know that for BRISQUE, the higher score means the worse image quality, and from the diagram, the mode of the "bad" is obviously higher than the mode of the "good". We used the average of the 2 modes as the threshold to label the predictions. The same modification was applied for NIQE and PIQE. The overall classification accuracy for these prediction models are presented in Table II, which shows that the proposed method performed better than the other three IQA models.

TABLE I: Confusion matrix

Caterogy	Bad	Good
Bad	4017	1240
Good	1268	14260

TABLE II: No-Reference classification comparison

Algorithms	BRISQUE	NIQE	PIQE	Proposed
Accuracy	76.7%	75.7%	76.7%	**90.6%**

V. Conclusion

We have proposed a no-reference model that can predict the quality of an image from the perspective of deep-learning-based CV tasks. The model was built based on a comprehensive dataset that includes common types of distortions, and it considered three fundamental CV tasks. The model has achieved good prediction accuracy and it is lightweight and easy to implement. It serves as a general and effective quality assessment solution for a wide range of camera-based intelligent sensing applications.

References

[1] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, 2012.

[2] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[3] V. N, P. D, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*, 2015, pp. 1–6.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[5] L. Kong, R. Dai, and Y. Zhang, "A new quality model for object detection using compressed videos," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3797–3801.

[6] G. F. Elsayed, S. Shankar, B. Cheung, N. Papernot, A. Kurakin, I. Goodfellow, and J. Sohl-Dickstein, "Adversarial examples that fool both computer vision and time-limited humans," 2018.

[7] A. Abaza, M. A. Harrison, and T. Bourlai, "Quality metrics for practical face recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 3103–3107.

[8] S. Gunasekar, J. Ghosh, and A. C. Bovik, "Face detection on distorted images augmented by perceptual quality-aware features," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2119–2131, 2014.

[9] L. Kong, A. Ikusan, R. Dai, and J. Zhu, "Blind image quality prediction for object detection," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2019, pp. 216–221.

[10] J. M. Irvine and R. J. Wood, "Real-time video image quality estimation supports enhanced tracker performance," in *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications X*, vol. 8713, International Society for Optics and Photonics. SPIE, 2013, pp. 302 – 313. [Online]. Available: https://doi.org/10.1117/12.2016174

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[12] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," 2019.

[13] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: http://arxiv.org/abs/1405.0312

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[15] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2017. [Online]. Available: https://arxiv.org/abs/1703.06870

[17] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[18] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Human vision and electronic imaging VIII*, vol. 5007. International Society for Optics and Photonics, 2003, pp. 87–95.

[19] E. Peli, "Contrast in complex images." *Journal of the Optical Society of America. A, Optics and image science*, vol. 7 10, pp. 2032–40, 1990.

[20] J. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 3, 2000, pp. 314–317 vol.3.

[21] S. Bezryadin, P. Bourov, and D. Ilinih, "Brightness calculation in digital image processing," in *International symposium on technologies for digital photo fulfillment*, vol. 1. Society for Imaging Science and Technology, 2007, pp. 10–15.

[22] K.-F. Yang, H. Li, C.-Y. Li, and Y.-J. Li, "A unified framework for salient structure detection by contour-guided visual search," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3475–3488, 2016.

[23] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.