Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW) 2017 10-14 July 2017

IMAGE BLUR CLASSIFICATION AND BLUR USEFULNESS ASSESSMENT

Mingyuan Fan[1,2], *Rui Huang*[1,2*], *Wei Feng*[1,2], *Jizhou Sun*[1,2]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] Key Research Center for Surface Monitoring and Analysis of Cultural Relics, State Administration of Cultural Heritage, China

{myfan,ruihuang,wfeng,jzsun}@tju.edu.cn

ABSTRACT

Blur classification is an important and widely-studied problem in computer vision. State-of-the-art blur classification methods are designed and verified using man-made blur images of known blur types and blur kernels. In reality, natural blur occurs under wild-conditions, and thus cannot be simply simulated by several hand-crafted blur kernels. Hence, conventional blur classification methods cannot deal with complex real-world blur classification tasks. In this paper, we propose a new blur classification model, which learns from real-world images by convolutional neural network. On the basis of the blur classification network result, we further propose an interesting and useful problem, called blur usefulness assessment, which assesses the usefulness of blur image. To support blur classification and blur usefulness assessment, we establish a useful blur image classification dataset, U-BICD, which contains $1,000$ sharp images and $1,000$ blur images (500 useful and 500 useless images). Compared with state-of-the-art blur classification methods, our method have achieved the highest blur classification accuracy of 98.1%. Our blur usefulness assessment also achieves an accuracy of 89.1%.

Index Terms— Image blur classification, blur usefulness assessment, blur classification dataset

1. INTRODUCTION

Blur classification is a conventional and useful problem in computer vision and multi-media processing, and this technique has been studied for several years. Unlike image quality estimation [3–7] and image blur assessment [8–10], image blur classification pays more attention on classifying images into blur or not. Recent works [11,12] further classify the blur types into motion, out-of-focus, etc. However, most conventional blur classification methods are designed based on man-made blur images with known blur kernel parameters [10,13].

*is the corresponding author. Tel: (+86)-22-27406538. This work is supported by NSFC (61671325, 61572354) and the National Science and Technology Support Project (2014BAK07B03).

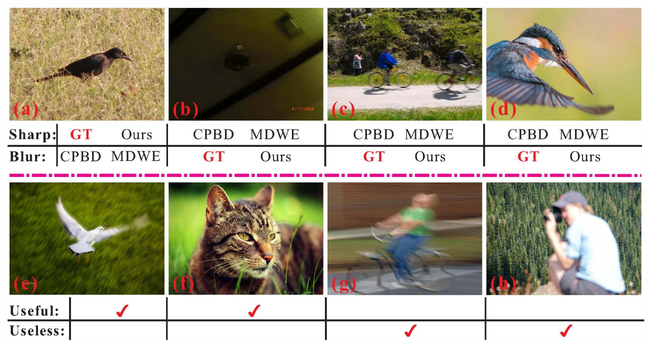

Fig. 1. Examples of image blur classification and blur usefulness assessment. In the top row, we show four images and their corresponding blur classification results of CPBD [1], MDWE [2] and our method. Note, GT denotes ground truth. In the bottom row, we show four images for blur usefulness assessment and the results of our method. (a) is a sharp image, (c), (e) and (g) are motion blur images, (b), (d), (f) and (h) are out-of-focus blur images.

In real-world cases, blur cannot be simply simulated by several hand-crafted blur kernels. Thus, previous methods, such as CPBD [1] and MDWE [2], fail to distinguish real blur images from sharp images when treating the blur classification in reality as shown in the first row of Fig. 1.

In this paper, we propose a blur classification model, which is a deep learning model and has powerful capacity of learning how to distinguish blur and sharp images of complex real-world scenes. However, training a data-hungry deep convolutional neural network (CNN) for image-level blur classification is impossible, because a blur classification dataset with sufficient labeled images is unavailable. To solve the data scarceness problem, we train a deep CNN with image patches on the basis of blur segmentation dataset [14]. This dataset contains 296 motion images and 704 out-of-focus blur images. Each of them has pixel-level label. The trained network is called BCnet. BCnet uses image patch as input and outputs three probabilities of three types, namely, sharp, motion blur, out-of-focus blur. On the basis of the BCnet output,

978-1-5386-0560-8/17/$31.00 ©2017 IEEE ICMEW 2017

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.

we use a simple yet effective method that thresholds the number of sharp image patches sampled from image pyramid to classify an image into sharp or blur. As shown in the top row of Fig. 1, our blur classification method can accurately classify the images captured in real-world scenes.

However, simply classifying images into blur and sharp cannot fit the general applications. For example, photographers often use out-of-focus and motion blur to highlight the subjects. We define this type of blur image as useful blur image. Useful blur images always have clean purpose and expression, such as a flying bird waving its wings in Fig. 1 (e) and a cozy cat staring at something in Fig. 1 (f). By contrast, the images with meaningless blur as shown in Fig. 1 (g) and Fig. 1 (h), belong to useless blur images. To avoid removing useful blur images, we propose a novel problem, called blur usefulness assessment. Different from aesthetic estimation, that aims to rank the aesthetic of an image without considering whether the image is blurry, blur usefulness assessment is a further classification step of blur image. Thus blur usefulness assessment can be used to remove the meaningless blur images as preprocessing step for many computer vision applications [15].

Useful blur images usually contain salient objects that are highlighted by intentional blur. Thus, we first attempt to incorporate saliency detection and blur classification to solve the blur usefulness assessment problem. We obtain image-level usefulness features for training a support vector machine (SVM) classifier for blur usefulness assessment by stacking the features of image patches. To facilitate the evaluation, we build a useful blur image classification dataset (UBICD) with real-world images. UBICD consists of $2,000$ images, including $1,000$ sharp images, 500 useful blur images and 500 useless blur images. We will share our UBICD after the approval of our paper. Our method achieves the highest accuracy of 98.1% on UBICD for image blur classification compared with state-of-the-art blur classification methods and an accuracy of 89.1% for blur usefulness assessment.

2. RELATED WORK

Image quality estimation, blur assessment are related to blur classification. In general, image quality estimation focuses on assessing the distortion degree of JPEG, JPEG2000, Gaussian blur, noise and fast-fading channel distortions on the basis of natural scene statistic [3–7]. Most image quality estimation methods only treat Gaussian blur, which is only a small portion of blur types. Image blur assessment, is aim at endowing a blurriness score to an image [13, 16–18]. This kind of methods often partition images into non-overlap blocks, then calculate the blur degree of the edges. The final blurriness of an image can be calculated by the cumulative probability of blur detection [1] or by averaging all local blur values [2]. Unlike previous studies, image blur classification pays more attention on classifying images into blur or not [8–12]. In this paper, we propose a blur image classification method based on the state-of-the-art deep learning framework. Our method learns blur classification deep convolutional neural network (BCnet) by real-world blur images. Based on BCnet, we propose image-level blur classification and blur usefulness assessment.

3. METHOD

Our image blur classification and blur usefulness assessment methods are shown in Fig. 2. The purpose of blur classification is classifying images into two types, i.e., sharp and blur. Note, since our BCnet outputs three probabilities of sharp, motion blur and out-of-focus blur, we classify an image patch to be blur when it contains one blur type. Besides, we can easily get the blur type of the blur image. After obtaining the blur images, we further conduct blur usefulness assessment by incorporating BCnet and saliency detection to decide which blur image is useful or useless. In following section, we first introduce the architecture, training data and training setting of our blur classification network (BCnet). Then we present how to employ BCnet to obtain image-level blur classification. Finally, we present our blur usefulness assessment.

3.1. BCnet

Network architecture. Our BCnet has three convolutional layers and three fully connected layers. The first convolutional layer contains 64 filters of size 11×11 . The second convolutional layer contains 128 filters of size 5×5. The third convolutional layer contains 256 filters of size 3×3. Each convolutional layer is followed by a 2×2 max pooling layer that reduces the spatial resolution to achieve translation invariance and reduce computation cost. The fully connected layers 4 and 5 contains 1024 and 256 filters, respectively. We use dropout with probability of 0.5 to avoid overfitting in the two fully connected layers. The last layer has 3 filters with three-way softmax layer for three types classification (i.e., sharp, motion blur and out-of-focus blur), which are used for blur image classification and blur usefulness assessment. BCnet inputs 48×48 image patches and outputs 3 probabilities.

Training data. Directly training a deep CNN for image-level blur classification needs a large number of labeled images. However, the present labeled blur classification images are not sufficient for this data-hungry training. To solve scarceness of training data of image blur classification, we use the patch-level training strategy. Shi *et al.* [14] shared a blur detection dataset with 296 motion images and 704 out-of-focus blur images. Each of the image has pixel-level label. We randomly select 80% of the two kinds of blur images to build a training set. We use 800 training images, which contains 236 motion images and 564 out-of-focus images. For each image in the training set, we collect samples by cropping image patches with size of 48×48 and a stride of 25 pixels in

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.

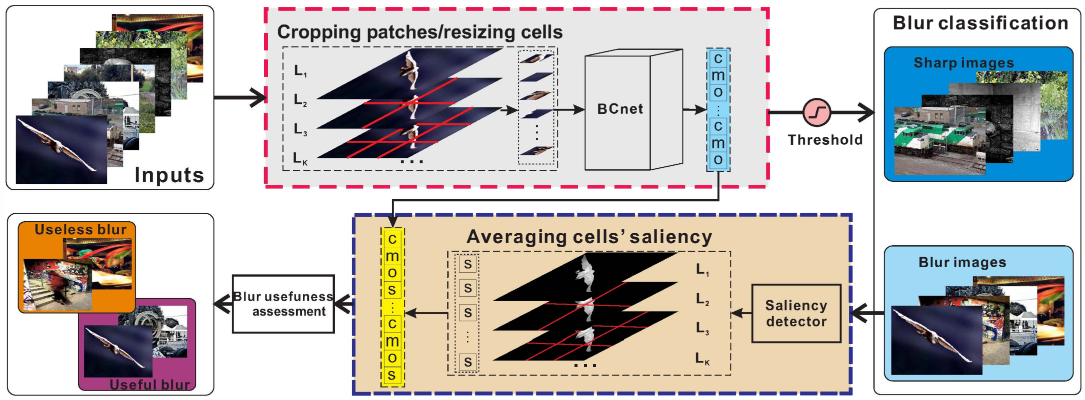

Fig. 2. The framework of our blur classification and blur usefulness assessment. Note, **c**, **m**, **o** and **s** denote the sharp probability (to distinguish saliency **s**), motion blur probability, out-of-focus blur probability, separately. Please see text for details.

sliding window manner. To enlarge the training samples, we expand the image data by resizing the images with different scales (i.e., 1, $7/8$, $5/8$, $3/8$ and $1/8$). The ratio of the number of training samples of each type is $1:1:1$. Finally, we collect $10,569,774$ training samples. We use 80% of these samples for training, and the remaining 20% of the samples are used for validation.

Training setting. Our model is implemented under the deep learning framework CAFFE [19]. We train BCnet using stochastic gradient descent with the batch size of 256, momentum of 0.9, and weight decay of 0.0005. The base learning rate is set to 0.001. We drop it by multiplying 0.1 when the validation accuracy does not improve.

3.2. Image blur classification

In real-world, motion blur and out-of-focus blur may coexist in an image. Thus, classifying an image into motion and out-of-focus blur may confuse the classifier. In this paper, we classify images into two types, i.e., sharp and blur. Note, when we say an image is blurry, we mean that there exist blur in the image without considering its blur type. Due to the specialization of BCnet, we can easily get the blur type of an image, even both of motion and out-of-focus blur present. Thus, motion and out-of-focus blur are not distinguished in our blur classification.

To conduct image-level blur classification, we use statistical method, which counts the number of sharp/blur image patches randomly sampled from cells of pyramid structure of image. Note, we use pyramid structure for the purpose that extracting patches uniformly from image, which also increases the number of patches for statistic. We partition an image into K pyramid levels, without downsampling the image. In the kth pyramid level, we partition the image into k^2 cells. Thus, there is only one cell in the first pyramid level, i.e., whole image. We randomly crop M image patches with size of 48×48. In our experiment, we experimentally set $M = 10$.

Then we classify each patch into sharp or blur (i.e., motion blur and out-of-focus blur) using BCnet. We count the number of the sharp image patches and set a threshold to classify an image to be sharp or blur. Formally, we define

$$l = \begin{cases} 1 & \text{if } N_{\mathtt{sharp}} >= N_{\text{patch}} * \tau \\ 0 & \text{otherwise} \end{cases}, \qquad (1)$$

where l is image label, $l = 1$ indicates image is sharp and $l = 0$ indicates image is blurry. $N_{\mathtt{sharp}}$ is the number of sharp image patches. $N_{\text{patch}} = \sum_{k=1}^{K} k^2 * M$ is the number of all image patches. Threshold τ is experimentally set to a constant of 0.9. We analyze the running time and classification accuracy using pyramid level $K = 1$ to 6. And we set $K = 3$ for the computation efficiency and satisfied accuracy.

To compute the sharpness/blurriness of an image, first, we define $\rho = N_{\mathtt{sharp}}/N_{\text{patch}}$. Then we calculate $\overline{\rho}$ by normalizing ρ to $[0, 1]$ on whole blur classification dataset. Thus, $\overline{\rho}$ denotes sharpness, $1 - \overline{\rho}$ denotes blurriness.

3.3. Blur usefulness assessment

In Section 3.2, we can obtain the sharp images. However, we find that some blur images are intentionally shot by the photographers to highlight the subjects, in which, blur is useful. This inspires us to propose a novel problem, i.e., blur usefulness assessment.

Observed from blur images, we find that useful blur highlights salient objects, while useless blur often distracts the attention from the salient objects. Meanwhile, useful blur images always have clean purpose and expression. In this paper, we try to incorporate salient object detection with blur classification to solve this novel and interesting problem. The basic idea is extracting blur usefulness related features and learning a classifier to assess the usefulness of the blur image.

Usefulness related features. Blur usefulness is highly related to blur types, image content and place that blur occurs.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.

We use salient object detector wCtr [20] to catch main subjects of the image. Saliency map always presents high saliency value on salient objects which cause more attentions than background regions. We build connections between blur and saliency to depict the blur usefulness. Specially, we exploit the multilevel pyramid to extract global and local usefulness related features. In kth level pyramid, first, we partition the original image into $k \times k$ cells and resize each cell to 48×48. We use BCnet to obtain the probabilities (sharp, motion blur, out-of-focus blur) to get a 3d deep blur features. Then we derive the corresponding saliency map of the cell and average the saliency values of the cell as 1d saliency feature. Thus, for each cell, we concatenate 3d deep blur features and 1d saliency features and obtain a 4d (sharp, motion blur, out-of-focus blur and saliency) usefulness related features. Finally, the features of all cells are concatenated together to generate image-level usefulness related features. When using pyramid level K, we can obtain a $n_{dim} = 4 * \sum_{k=1}^{K} k^2$ dimensional features each image.

Usefulness assessment. Given a training image set $\mathbb{I} = \{I_1, I_2, \ldots, I_N\}$, where N is image number. Let $\mathbb{F} = \{F_1, F_2, \ldots, F_N\}$ denotes the feature set, where $F_i \in \Re^{1 \times n_{dim}}$. The corresponding label set is $\mathbb{L} = \{l_1, l_2, \ldots, l_N\}$, where $l_i \in \{-1, +1\}$, $+1$ means useful blur, and -1 means useless blur. We train a classical SVM [21] to assess the usefulness of a blur image. We implement our training and testing by LIBLINEAR [22].

4. EXPERIMENT

4.1. UBICD

We build a dataset named useful blur image classification dataset (i.e., UBICD). UBICD consists of $2,000$ images, wherein half images are sharp images and half images are blur images. We manually select sharp images from PASCAL VOC2012 dataset [23]. The blur images are collected from Internet. We classify the blur images into useful and useless blur images by five people. According to the voting results, we select 500 useful and 500 useless blur images.

4.2. Setup

Baselines. We compare our method with five methods including BIQI [3], BRI [6], JNB [17], CPBD [1] and MDWE [2]. For all these methods we used the parameters suggested in the their papers. The above methods output scores in a pre-specified range (e.g., [0,1]), which indicating the level of blur, rather than making a binary decision. To make a binary decision for blur classification, we transform the scores into binary sharp/blur decisions by searching an appropriate threshold for each method, respectively.

Criteria. Following [13], we use overall Accuracy, F-measure for evaluation. For F-measure, we set $\beta = 1$. Since

Table 1. Quantitative comparisons of different blur classification methods on UBICD.

Methods	Ours	BIQI [3]	BRI [6]	JNB [17]	CPBD [1]	MDWE [2]
Accuracy	**98.1%**	51.75%	50.8%	69.95%	94.35%	95.1%
F-measure	**0.981**	0.6411	0.052	0.7	0.9414	0.9506

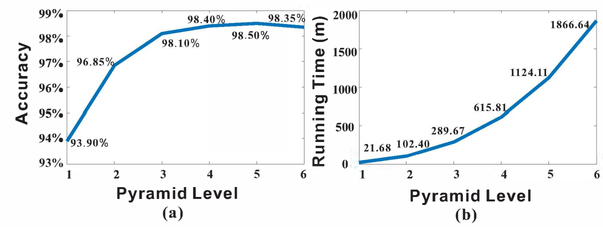

Fig. 3. (a) shows average accuracy change curve of our blur image classification method with different pyramid level K on UBICD. (b) shows the total running time of our blur classification method with different pyramid level K for all images on UBICD. Note, we use minute as time unit.

F-measure is a balance metric of Precision and Recall, in following result analysis, we mainly analyze the improvement on Accuracy and F-measure.

4.3. Result analysis

Image blur classification. We first analyze the blur classification accuracy and computation cost of setting pyramid level K to $[1-6]$. As shown in Fig. 3, as the increasing of K from 1 to 6, the blur image classification accuracy increases from 93.9% to 98.5%, and the total running time increases from 21.68m to 1866.64m. We obtain the highest blur image classification accuracy 98.5% at pyramid level $K = 5$. Note, when $K = 3$, the accuracy already reaches to 98.1%. And the accuracy increases only 0.4% from $K = 3$ to 6. Thus, we propose to set $K = 3$ for blur image classification, which considers the tradeoff between the time and accuracy.

In Table 1, we compare our method with some state-of-the-art blur image classification methods. We can find that our method outperforms other methods by a large margin. Specially, we find that MDWE [2] is the best method among the compared methods, which achieves 95.1% and 0.9506 for accuracy and F-measure, respectively. Our method obtains improvements of 3.2% and 3.2% over MDWE [2], with absolute incasement of 0.03 and 0.0304 of Accuracy and F-measure, respectively.

In Fig. 4, we show our sharpness measurement $\overline{\rho}$ of motion blur, out-of-focus blur and sharp images. We can find that as the sharpness increases, the blur extent decreases and the images are more prone to be sharp.

Blur usefulness assessment. For blur usefulness assessment, we use the subset of UBICD, which consists of 500 useful blur images and 500 useless blur images. Since we are the

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.

Fig. 4. Examples of image sharpness measurement $\overline{\rho}$ for motion blur (1st row), out-of-focus blur (2nd row) and sharp (3rd row) images. The sharpness scores are highlighted by the pink rectangles.

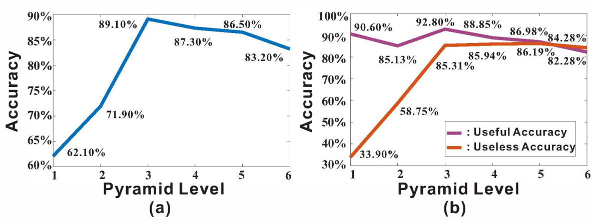

Fig. 5. Average accuracy changes of blur usefulness assessment with different pyramid level settings. We use six pyramid levels for evaluations. (a) shows average accuracy change curve on all test images for blur usefulness assessment. (b) shows average accuracy change curves of classifying useful and useless blur images separately.

first to deal with blur usefulness assessment, we do not have other methods to compare. We train different SVM models with our usefulness related features that extracted from pyramid levels (i.e., K) 1 to 6. We randomly select 400 useful blur images and 400 useless blur images for training, and the remaining 200 images are used for testing. We conduct experiments 10 times for each K and report the average accuracy.

As shown in Fig. 5 (a), the accuracy of blur usefulness assessment initially increases with pyramid level increasing. The accuracy reaches the highest value of 89.1% at pyramid level 3. However, as the pyramid level increases, the feature dimension also increases, which results in heavy training burden. Our blur usefulness related features can capture the blur and corresponding image content information more thoroughly, thus can support blur usefulness assessment. When using $K > 3$, the accuracy decreases slowly. Although the representation capability of the blur usefulness related features increases when increasing pyramid level, the capacity of the SVM model has already reached the limitation.

As shown in Fig. 5 (b), the accuracy of useful blur image classification presents a smooth changes as pyramid level

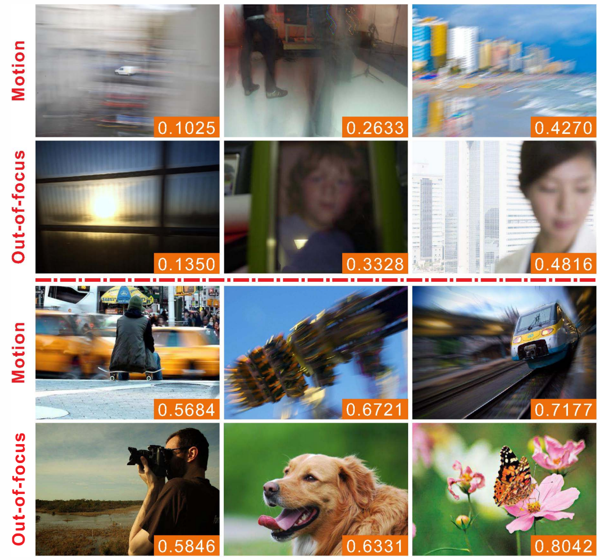

Fig. 6. Examples of blur image usefulness assessment. The blur useful scores are highlighted by orange rectangles. Top two rows show the images with useless blur. Bottom two rows show the images with useful blur. The corresponding blur type are labeled at the leftmost.

increases, while the accuracy of useless blur image classification changes obviously. We argue that on UBICD dataset, useless blur images are highly related to the spatial position of blur that occurs. Note, we just consider only two cues for blur usefulness assessment, which is most restricted by the small number of the training images.

In Fig. 6, we shows some examples of blur image usefulness assessment. The usefulness score is normalized SVM confidence. We can find that the lower usefulness score, the main subject of the image is more likely distracted by useless blur. On the contrary, the higher usefulness score, the main subject of the image is more likely highlighted and purpose and expression are cleaner.

5. CONCLUSION

In this paper, we have proposed a learning-based method for image blur classification. The proposed method learns from real-world blur images and achieves best classification accuracy. In addition, we have proposed a novel problem, called blur usefulness assessment, which further classifies blur images into useful and useless. To facilitate evaluations and comparisons, we established a large and high quality dataset (i.e., UBICD). On UBICD we achieve 98.1% accuracy for image blur classification and 89.1% accuracy for blur usefulness assessment. In our future work, we plan to enlarge our UBICD to include large number of sharp images and blur images. On the basis of this large dataset, we plan to train a deep CNN for image blur classification and blur usefulness assessment directly from images.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (cpbd)," *IEEE TIP*, vol. 20, no. 9, pp. 2678–83, 2011.

[2] S . Winkler P. Marziliano, F. Dufaux and T. Ebrahimi, "A no-reference perceptual blur metric," in *ICIP*, 2002.

[3] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE S-PL*, vol. 17, no. 5, pp. 513–516, 2010.

[4] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: a natural scene statistics approach in the dct domain," *IEEE TIP*, vol. 21, no. 8, pp. 3339–52, 2012.

[5] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Making image quality assessment robust," in *CSC*, 2012.

[6] A. K. Moorthy A. Mittal and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE TIP*, vol. 21, no. 12, pp. 4695–4708, 2012.

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

[8] B. T. Koik and H. Ibrahim, "A literature survey on blur detection algorithms for digital imaging," in *ICAIMS*, 2013.

[9] G. L. Sun, G. Q. Li, and J. Yin, "Blurred image classification based on adaptive dictionary," *IJM*, vol. 5, no. 1, pp. 1–9, 2012.

[10] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *EI*, 2007.

[11] H. Ping and B. Y. Chen, "Blurred image detection and classification," in *IMMC*, 2008.

[12] R. Yan and L. Shao, "Image blur classification and parameter identification using two-stage deep belief networks," in *BMVC*, 2013.

[13] E. Mavridaki and V. Mezaris, "No-reference blur assessment in natural images using fourier transform and spatial pyramids," in *ICIP*, 2014.

[14] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," in *CVPR*, 2014.

[15] W. Liu, T. Mei, Y. Zhang, C. Che, and J. Luo, "Multi-task deep visual-semantic embedding for video thumbnail selection," in *CVPR*, 2015, pp. 3707–3715.

[16] M. J. Chen and A. C. Bovik, "No-reference image blur assessment using multiscale gradient," *EJIVP*, vol. 2011, no. 1, pp. 1–11, 2011.

[17] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur," *IEEE TIP*, vol. 18, no. 4, pp. 717–728, 2009.

[18] T. Oh and S. Lee, "Blind sharpness prediction based on image-based motion blur analysis," *IEEE TOB*, vol. 61, no. 1, pp. 1–15, 2015.

[19] Y. Q. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv*, 2014.

[20] W. J. Zhu, S. Liang, Y. C. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[22] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "Liblinear: A library for large linear classification," *JMLR*, vol. 9, no. 9, pp. 1871–1874, 2008.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *IJCV*, vol. 88, no. 2, pp. 303–338, 2010.

Authorized licensed use limited to: Access paid by The UC Irvine Libraries. Downloaded on March 24,2021 at 15:25:25 UTC from IEEE Xplore. Restrictions apply.