A Knowledge Enforcement Network-based Approach for Classifying a Photographer's Images

Palaiahnakote Shivakumara[1], Pinaki Nath Chowdhury[2], Umapada Pal[2], David Doermann[3], Raghavendra Ramachandra[4], Tong Lu[5,] and Michael Blumenstein[6]

[1]Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. shiva@um.edu.my

[2]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India Email: pinakinathc@gmail.com, umapada@isical.ac.in.

[3]Department of Computer Science and Engineering, University at Buffalo, USA. doermann@buffalo.edu

[4]Faculty of Information Technology and Electrical Engineering, IIK, NTNU, Norway, raghavendra.ramachandra@ntnu.no

[5]National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China, lutong@nju.edu.cn.

[6]University of Technology Sydney, Australia, Email: michael.blumenstein@uts.edu.au

Abstract

Classification of photos captured by different photographers is an important and challenging problem in knowledge-based and image processing. Monitoring and authenticating images uploaded on social media are essential, and verifying the source is one key piece of evidence. We present a novel framework for classifying photos of different photographers based on the combination of local features and deep learning models. The proposed work uses focused and defocused information in the input images to extract contextual information. The model estimates the weighted gradient and calculates entropy to strengthen context features. The focused and defocused information is fused to estimate cross-covariance and define a linear relationship between them. This relationship results in a feature matrix fed to Knowledge Enforcement Network (KEN) for obtaining representative features. Due to the strong discriminative ability of deep learning models, we employ the lightweight and accurate MobileNetV2. The output of KEN and MobileNetV2 is sent to a classifier for photographer classification. Experimental results of the proposed model on our dataset of 46 photographer classes (46234 images) and publicly available datasets of 41 photographer classes (218303 images) show that the method outperforms the existing techniques by 5-10% on average. The dataset created for the experimental purpose will be made available upon publication.

Keywords: Focused region classification, defocused region classification, Low-high pass filters, Entropy features, Deep learning, Photographer identification

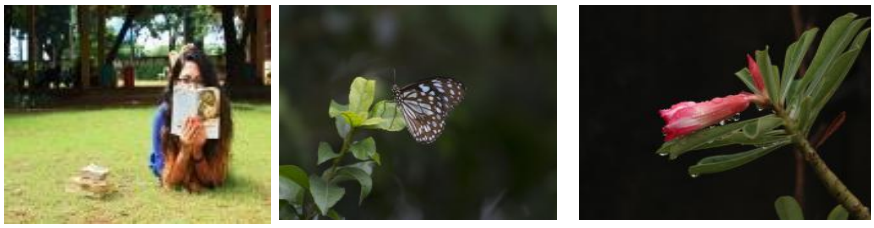
1. Introduction

In the case of watchdog photos uploaded on social media, validating pictures for authentication and monitoring the content has drawn particular attention from researchers [1, 2, 3]. Sharing sensitive

images, for example, often indicate fake news, blackmail, or terrorism-related activities. In addition, due to urbanization, crime rates are also increasing exponentially [4]. Person identification applications use caricatures generated by a given face photo. This is called photo-to-caricature translation [5]. As a result, one can expect many crime-scene photos captured by different people and cameras [6].

In the same way, there are other real-world applications to recommend and cluster photos uploaded on social media using geographical locations and camera parameters to enhance tourism [7]. One such feature is to identify the person who has taken the photos; hence, it is a key objective for developing a new model in this work [8, 9, 10]. In the past, methods have been developed for source camera identification [11] using images to identify the culprit. Still, these methods may not work well for the situation where a photographer can use multiple cameras to capture images. In addition, the quality of the photos depends on several parameters of the camera, weather conditions, viewing locations, projector screen configuration, and viewers' psychological factors [12, 13]. Therefore, this problem is considered an open issue and an elusive goal for video and image processing researchers.

There are approaches developed in the past to address image tampering and fake images by identifying the image source, such as camera identification and social network identification [2, 3, 11]. However, these methods work based on noise, distortion, or errors introduced from the camera sensor. These features are not necessarily relevant in classifying images of different photographers because the characteristics of the photos may change according to the habits and professional experience of the photographer [14]. As a result, one can expect significant variations in intra- and inter-image classes for different photographers. Similarly, digital images provide metadata like time, date, format, etc., which can be used to identify the kind of camera and the photographer's identity. But this data is vulnerable and can be changed or modified easily [11]. This is not reliable information for authentication and validation of images.



Photographer-1 Photographer-2 Photographer-3

Fig. 1. Examples of images of three different photographers.

In the same way, the methods [15, 16, 17, 18] proposed for scene categorization, scene image classification, and scene image recognition. However, these methods may not be effective for classifying photos captured by photographers because these methods require particular objects and shapes, which may not be prominent in the case of photographer images. Hence, we can infer that classifying a photographer's images using image content is an open and complex challenge. It is evident from the sample images of three photographer classes shown in Fig. 1, where we can see common

information in the photo, although captured by different photographers. In addition, since there is no constraint in capturing scenes, it makes photographer identification even more challenging.

To find a solution to this complex problem, we were inspired by the work on writer identification using handwriting analysis [19, 20], where the authors assume that each writer has a unique way of writing and significant variations in the writing of intra- and inter- classes. We propose a novel idea for classifying a photographer's images in this work. Similarly, the statistical analysis [14] shows that humans can classify images captured by different photographers into correct classes. This indicates common information for images captured by the same photographer and unique information for images captured by other photographers. This observation further motivates us to propose a new model for classifying photos captured by different photographers.

The proposed work is the first to classify the photos (images) of different professional and non-professional photographers. The contributions of our work are as follows: (1) Proposing a new method based on low and high pass filter kernels for separating focused and defocused regions. (2) Introducing Knowledge Enforcement Network (KEN) for extracting context features from focused and defocused images through a new fusion operation. (3) Introducing MobileNetV2 to extract features from input images and propose a new architecture for combining context and deep features for photographer identification. (4) The proposed work integrates handcrafted features through KEN and image features through MobileNet for achieving the best results is novel compared to the state-of-the-art methods.

The proposed work involves image processing pattern recognition techniques, such as high and low pass filters and clustering, for classifying focused and defocused regions in the input image. In addition, machine learning and deep learning techniques, such as knowledge enforcement networks and MobileNetV2 network, feature extraction, and classification of photos of different photographers. Therefore, one can summarize that image processing, and pattern recognition helps us to separate focused and defocused regions in the input image. In contrast, machine learning approaches help us define the relationship between focused and defocused regions so that high-level features can be extracted for classifying images of different photographers. Overall, the methods and concepts used in this work are all part of the artificial intelligence area in general.

The rest of the paper is organized as follows. Section 2 discusses related work. The proposed method is detailed in Section 3. Experimental results are provided in Section 4, and finally, Section 5 concludes the paper and gives directions for future work.

2. Related Work

The proposed work aims to classify the images captured by different photographers. Therefore, the approaches related to scene image classification, classification of images of personality traits, source camera identification, and photographers' image classification/identification are reviewed in the subsequent sections.

2.1. Scene Image Classification

Pan et al. [16] proposed a model for classifying scene images based on the foreground fisher vector. The idea is to separate class-relevant and class-irrelevant foreground and then use a descriptor for estimating feature vectors. Sun et al. [17] developed a method for scene categorization using a deep learning model and gaze shifting kernel. The approach detects the region of interest from each scenery to create a perceptual space that comprises color, texture, and semantic features. Wang et al. [18] used a hierarchical GAN tree and bi-directional capsules for scene image classification. The approach uses top-down and bottom-up criteria to find the relationship between the objects in the image for classification. Yang et al. [21] proposed a method based on a fully convolutional neural network for image scene analysis. The approach uses multi-scale fusion and weights of different sensitive channels for feature extraction and to restore spatial information. Zhang et al. [22] developed a method for classifying remote sensing scene images using co-evolution-based parameter learning. The technique proposes two population strategies to optimize the hyper and weight parameters. Finally, the method adopts parallel processing to achieve time efficiency.

In summary, although the scene classification methods are robust and powerful for the classification of natural scene images, these models are not suitable for classifying images captured by different photographers. The reason is that the features extracted by the methods are ineffective because one cannot expect common semantic information in the same class images. In other words, there is no correlation between the images of the same because the photographer can capture any scene images according to his interest and hobby.

2.2. Personality Traits Image Classification

There are methods for classifying images according to the person's personality traits. These methods can be used for person identification and behavior analysis, like photographer identification using image information. Using semantic analysis, Cheng et al. [23] developed a model for wedding event identification. The method extracts video information for wedding event identification. Since the technique requires video information as input, it is unsuitable for photographer identification. Beyan et al. [24] used deep visual activities for personality trait classification. The model's objective is to use non-verbal and spatio-temporal features for classification. This method is also limited to video but not still images. Krishnani et al. [25] proposed a structure function-based transformation to classify the social media image of different personality traits. The method works well for images containing faces with other expressions and emotions but not normal scene images. The same authors [26] developed an improved version of the method [25] for classifying different emotions in photographs. The model combines the Hanman transform and CNN for classification.

Zhang et al. [27] proposed a deep model to define the relationship between personality traits and emotions. The approach involves multi-task learning and works end-to-end for personality trait

identification. Sun et al. [28] used multimodal attention network learning for personality assessment. The approach combined gaze distribution and speech features for studying person behavior. However, the model is limited to video but not still images. Biswas et al. [2] used a multimodal for extracting features involving visual and textural features from images, profile pictures, and banners posted on social media to classify images of different personality traits. Liu et al. [29] extracted color and various texture features from the input images to classify personality traits images. The approach uses Twitter information and image information of profile pictures. Zhu et al. [30] proposed CNN for feature extraction from the input images of different personality traits. Finally, the probability distribution and the regression model are used to classify images of varying personality traits.

The above models use multimodal concepts to classify images of different personality traits. However, these features cannot be used to represent the unique region in the images captured by a different photographer. In addition, the methods are limited to images containing a person. This cannot be true in the images captured by other photographers.

2.3. Source Camera Identification

Since the objectives of the source camera identification and photographer identification from the images, such as assisting the forensic investigation team, are very similar, we reviewed the methods of camera identification in this section. Qiao et al. [10] explored a signal-dependent noise model for studying statistical distributions of pixels in JPG format for camera identification. Villalba et al. [31] used a PRNU-based method for manipulating smartphone image source identification. The method is based on sensor noise and a wavelet transform, identifying trace evidence indicating different cameras. Ding et al. [32] proposed domain knowledge-driven deep multi-task learning for camera device identification. It is noted from the above methods that the main objective of the approaches is to identify camera devices by studying noise introduced by cameras during image acquisitions at different levels and patterns of variations. These features may not be practical for classifying a photographer's images because they do not reflect the photographer's characteristics.

Amerni et al. [3] use image classification for social network identification. The work suggests that different social networks have other mechanisms for uploading and downloading images, which affect the content of the images. Zheng et al. [2] proposed a method for detecting forged information in images and source camera identification. The method is based on a Physical Unclonable Function (PUF) defined by the Bernoulli random space. Wang et al. [33] developed a unified framework for source camera identification. The method targets camera-specific artifacts based on preprocessing and residual calculation. Yuan et al. [34] aim at developing a model for cross-camera anchors detection application for person re-identification. The method defines judgment conditions to address scalability challenges by using association ranking.

The above discussions on the methods for camera source identification demonstrate that the approaches

consider that each camera introduces unique artifacts during image acquisition. Indeed, this artifact may not reflect the photographer's passion or the characteristics of the photographer. This is because image content changes according to the photographer's mind, focus, experience, and situation, including artifacts of multiple cameras. The same photographer can use different cameras to capture images in other conditions. Therefore, the approaches may not be practical for a photographer's image classification.

2.4. Photographer Image Classification/Identification

Since classifying photos captured by different photographers is a new research problem, we hardly find any methods in the literature. We thus discuss here some strategies with similar objectives. Birajdar et al. [35] present a systematic survey on photorealistic compute graphics and photographic image discrimination. However, this survey focuses on the classification of authentic photographic and computer graphics images but does not discuss the photos captured by different photographers. Rugna et al. [36] aim at developing a method to identify the country of origin for the tourists' using images. Angelov et al. [37] use images captured by mobile cameras for landmark recognition with the help of a Global Positioning System (GPS). The idea of the work is to group images based on landmarks such that if a person loses the location, they could recognize the location with the previous landmark images. The above two methods use tourist identification and location recognition images to help the person. Hoshen et al. [38] developed a technique for video photographer identification using egocentric features. The approach uses optical flow to study human motion and explores a convolutional neural network for identifying video photographers. However, the method works well for videos but not still images or photos without video information.

We found only one method, [14], specifically proposed to address professional photographer identification but not a mix of professional and non-professional considered in our work. Thomas et al. [14] extract image-based features, including low and high levels, and then extract features using a Convolutional Neural Network (CNN) to classify images captured by different photographers. This work assumes that the images captured by a particular photographer share common properties. This is not true in our work because the content of images can vary from one image to another. Our work involves images captured by professionals (experts in photography and their style and passion) and non-professional photographers (they capture images randomly). Therefore, the approach may not be robust on the proposed dataset.

Based on the above discussion, it is noted that none of the methods addressed the classification of photos captured by different photographers. However, Thomas and Kovashka [14] focus only on photos captured by a professional photographer. Hence, classifying photos (images) captured by photographers is an elusive goal for researchers. Therefore, this work aims to develop a new model for photographers' image classification.

3. Proposed Approach

We believe each photographer has their style of capturing images reflected in their photos. For example, in the case of an image with a bird and a river, the context or relationship between the bird and the river remains the same for images of birds with different rivers, lakes, ponds, etc. A photographer captures such images with the same style. Another photographer may capture similar images with different quality, focus, zoom in, zoom out, and point of view, for example, because each photographer has their style and way of capturing images. This observation motivated us to propose a method based on the relationship between focused and unfocused regions and the features extracted by a deep learning model for classification in this work. Inspired by the unique property of MobileNetv2 that it can achieve high accuracy with a smaller number of operations and parameters [39] in contrast to Alexnet [40] and ResNet [41], we explore MobileNetv2 for classification of images captured by a different photographer in this work. The statement is validated through ablation study experiments in the experimental section.

Our intuition is that the context between focused and defocused regions in images exhibits distinct cues for inter and intra-images of photographer classes. Focused areas are defined by high contrast pixels, while defocused regions are characterized by low contrast pixels [42]. Our method combines the properties of low- and high-pass kernels, which represent fine edges and non-edges, respectively. To separate those pixels, we use k-means clustering with k=2, which outputs two clusters, namely, Max cluster and Min cluster, for the respective focused and defocused images.

Motivated by the work in [43], where optimal weights are derived for fusing high and low-frequency coefficients to make the method robust to noise and multi-focused regions, we explore the same operation differently for combining the Max cluster of entropy for focused and defocused images, and the Min cluster of entropy for focused and defocused images. This results in two enhanced fused images: the Fused Max cluster and the Fused Min cluster. As expected, to use the context between focused and defocused regions, our method estimates cross-covariance between two fused images because the cross-covariance is meant for finding the linear relationship between objects in images [44]. This results in a feature matrix of the same size as the input image.

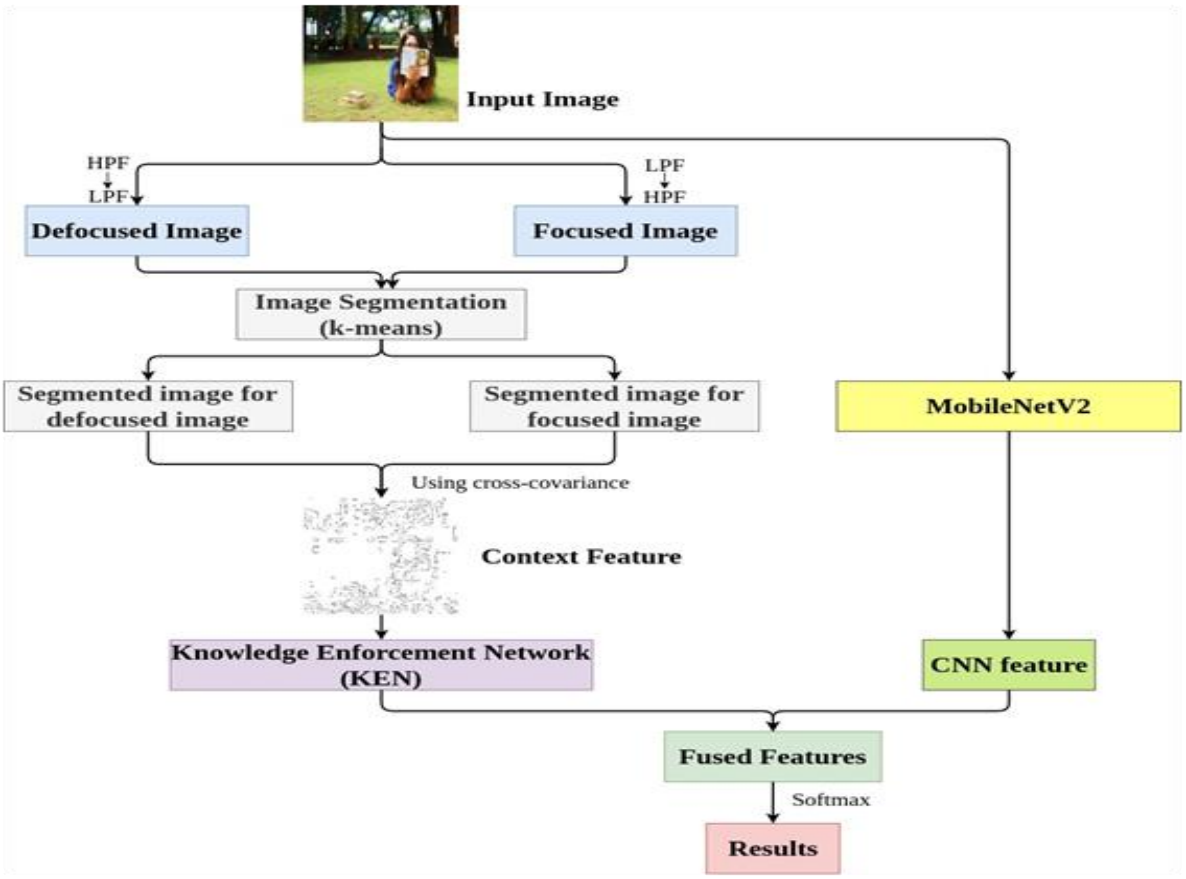
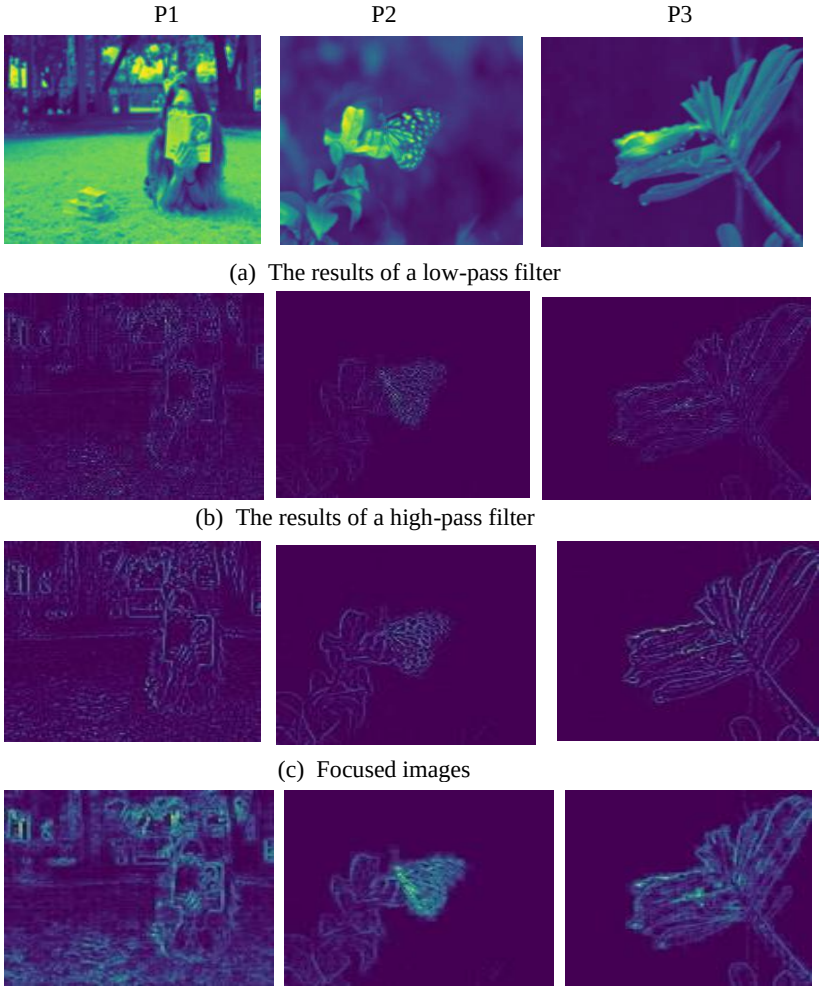


Fig. 2. The proposed framework for photographer identification. Here HPF and LPF are high and low pass filters.

Inspired by the success of classifying complex images by the deep learning models, we explore MobileNetv2 [39] to extract features and classification of photos in this work. The proposed model fuses the context features obtained by cross-covariance with the features extracted from the MobileNetv2 with the new architecture for classification. The unified framework can be seen in Fig. 2. Overall, we "force" the network to emphasize the fine-grained difference among photography styles. In this way, the extracted focused and defocused information is injected into the Knowledge Enforcement Network (KEN) for better classification.

It is noted that the considered problem is complex, and it is not so easy to obtain stable and reliable results. Therefore, to alleviate the problem, this work combines conventional features with the features of deep learning models. This makes sense because deep learning-based approaches may not be robust to diversified datasets or images compared to traditional feature-based models [45, 46]. The reason is that the performance of the deep learning model depends on the number of relevant samples, while the feature-based methods are not much dependent on samples. This is the advantage of the conventional feature-based method. But the deep learning models effectively achieve the best and high results for particular situations. This is the advantage of deep learning-based models. The proposed model integrates the merits of both models to achieve stable and consistent results for the classification of

images of different photographers.



(a) The results of a low-pass filter

(b) The results of a high-pass filter

(c) Focused images

(d) Defocused images

Fig. 3. Low and high pass kernels for obtaining focused and defocused images

3.1. Focused and Defocused Region Segmentation

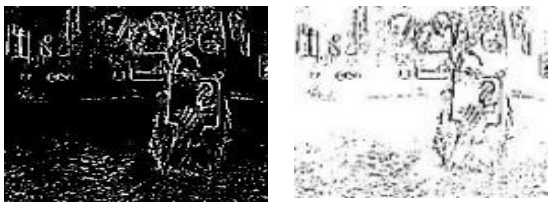
For each input image, our method uses low-pass and high-pass kernels as defined in Equation (1) for segmenting focused I_F and defocused I_{DF} regions in images as defined in Equation (2), where I denotes the input image. The operation defined for I_F indicates that it enhances pixels representing fine edges. In contrast, the operation defined for I_{DF} suggests that it enhances both the pixels representing fine edges and edges of the background. The effect of low and high-pass kernels can be seen in Fig. 3(a) and Fig. 3(b) for different photographer images, as shown in Fig. 1. It is observed in Fig. 3(a) and Fig. 3(b) that the low-pass kernel reduces the impact of high-value pixels, which represent noise. At the same time, it introduces blur while the high-pass kernel enhances the fine edge pixels and, at the same time, introduces noise. Due to this, pixels that represent fine edges are enhanced in the case of focused images. In contrast, pixels representing the fine edges of an object and its background are enhanced in the case of defocused images, as shown in Fig. 3(c) and Fig. 3(d), respectively.

$$f_1 = \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix},\quad f_2 = \frac{1}{4}\begin{bmatrix}0 & 1 & 0\\ 1 & -4 & 1\\ 0 & 1 & 0\end{bmatrix} \tag{1}$$

$$I_F = (I * f_1) * f_2, \qquad I_{DF} = (I * f_2) * f_1 \tag{2}$$

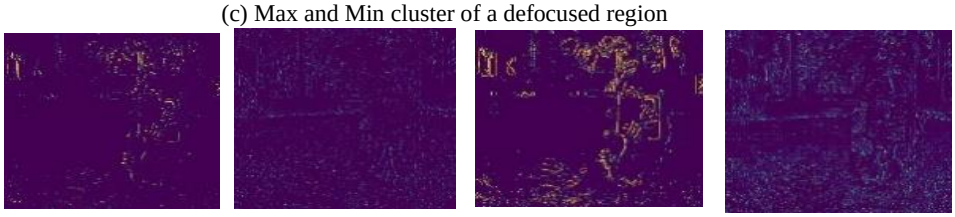
Where "*" indicates convolution operation.

(a) Photographer-1

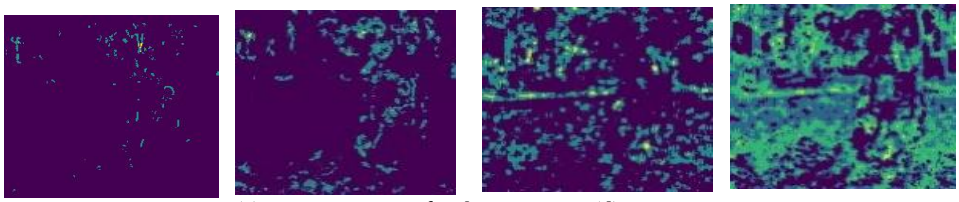


(b) Max and Min cluster of a Focused region

(c) Max and Min cluster of a defocused region



(d) Weighted gradient images for the images in (b) and (c)



(e) Entropy images for the images in (d)

Fused-Max Fused-Min Cross-covariance

(f) Fusing max clusters of entropy images of the respective focused and defocused regions to estimate cross-covariance.

Fig. 4. Estimating cross-variance for the input images.

To separate the pixels which play a prominent role in discrimination, we employ k-means clustering with k=2 on focused and defocused images. This outputs two clusters, namely, the Max cluster, which contains high values, and the Min cluster, which has low values for each image, as shown in Fig. 4(b) and Fig. 4(c), respectively, for the sample Photographer-1 input image shown in Fig. 4(a). Since the operation defined for obtaining focused and defocused regions introduces blur and noise, we can expect misclassification in the case of k-means clusters. To reduce this effect, we consider the value of a high-

pass kernel over the input image as weights. The values are multiplied by gradient values corresponding to pixels in the respective Max and Min clusters of focused and defocused images. This results in weighted gradient images, as shown in Fig. 4(d). Our method estimates entropy as defined in Equation (3) for each pixel in the weighted gradient images of focused imagesI_{FwEnt} and defocused images I_{DFwEnt} to strengthen the weighted gradient features. This is because entropy is a measure of energy and is estimated using neighboring pixels (3×3 window). The effect can be seen in Fig. 4(e).

$$H(x) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{3}$$

Motivated by the work in [43], where optimal weights are derived using high and low wavelet frequencies to overcome the challenges of multi-focus regions, we explore the same in a different way for deriving weights using entropy images as defined in Equation (4) and Equation (5). Based on the weights, our method fuses the Max cluster of focused and defocused images as defined in Equation (6) and Equation (7), which results in Fused Max cluster images as shown in Fig. 4(f)(1st image). In the same way, our method fuses the Min cluster of focused and defocused images, which results in Fused Min cluster images as shown in Fig. 4(f) (2nd image).

As mentioned in the Methodology Section, we aim to define the context between focused and defocused regions. We propose to estimate cross-covariance between Fused Max and Fused Min cluster images as defined in Equation (8) and Equation (9). The motivation for choosing cross-covariance is that, as stated in [44], cross-covariance extracts linear relationships between two objects. An element of the cross-covariance matrix (CM) in position (*i*, *j*) represents a degree of the linear relationship between object parts of the Fused Max cluster and the Fused Min cluster. A significant value of CM represents a stronger linear relationship between the two clusters. It is the same as defining the context between focused and defocused regions in the proposed work. The process gives a cross-covariance feature matrix for each input image, and the image format is shown in Fig. 4(f).

$$w_{max} = \frac{I_{FwEnt} * I_{DFwEnt}}{{I_{FwEnt}}^2 + {I_{DFwEnt}}^2} \tag{4}$$

$$w_{min} = 1 - w_{max} \tag{5}$$

if $I_{FwEnt} \geq I_{DFwEnt}$ then,

$$I_{Fused} = (I_{Fw} * w_{max}) + (I_{DFw} * w_{min}) \tag{6}$$

Else,

$$I_{Fused} = (I_{Fw} * w_{min}) + (I_{DFw} * w_{max}) \tag{7}$$

$$CM = \frac{\sum_{n=1}^{N} (X_n^1 - \mu)^T * (X_n^2 - \mu)}{N - 1} \tag{8}$$

The average mean vector μ is computed as follows:

$$\mu = \frac{\sum_{n=1}^{N} X_n^1}{N} \tag{9}$$

where, X_n^1 and X_n^2 are C_{max} and C_{min} images, respectively.

3.2. KEN and MobileNet for Classification of Photos

In addition to CM features which capture focused and defocused features, we observe that the feature extraction directly from the input image using the state-of-the-art Convolutional Neural Network (CNN) boosts the accuracy of the proposed model for classification of photos captured by a different photographer. The CM feature is applied to a novel Knowledge Enforcement Network (KEN) to transform and make it compatible with features extracted by the state-of-the-art CNN. The complete architecture of the proposed model can be seen in Fig. 5.

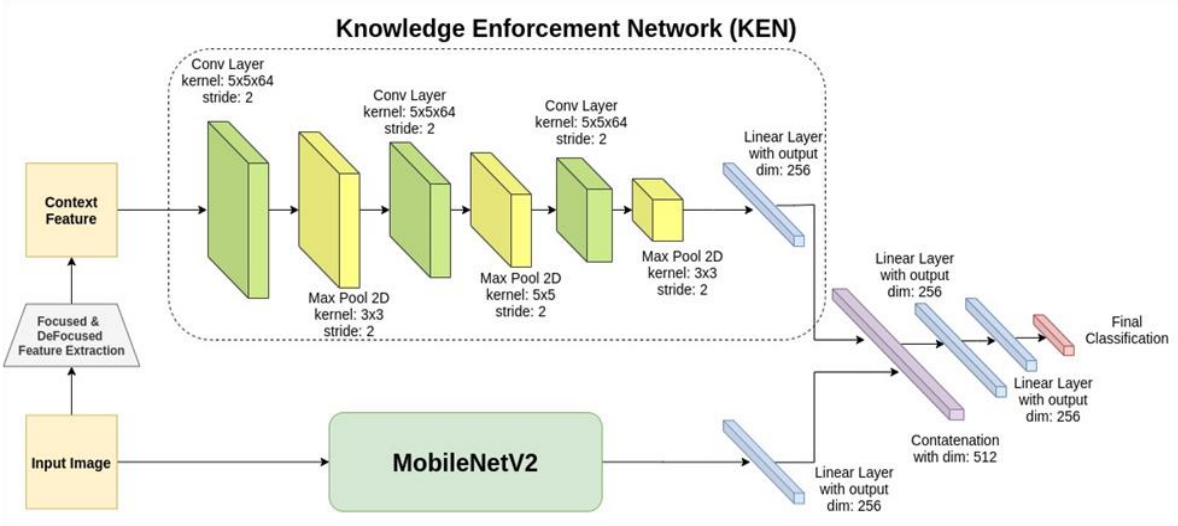


Fig. 5. The proposed new architecture by combining deep features and cross-variance features.

The inputs to the KEN are CM features, which are resized to the shape of (224× 224 ×1) and fed to a cascade of 3 modules, each consisting of Conv-ReLU-MaxPool layers. The first module consists of a conv layer with kernel size (5×5), stride and padding having value two, and a MaxPool layer with kernel size (3×3) and stride 2. Each of the remaining two modules in KEN comprises a Conv layer with a kernel size (3×3) with stride and padding value two and a MaxPool layer with kernel size (3×3) padding 2. Next, a flattened layer is added to obtain a 1-dimensional vector having dimension 576. The vector is fed to a fully connected layer, reducing the vector's dimension to 256.

As mentioned earlier, to boost the performance of our method for photographer identification, we employ three popular state-of-the-art CNN architectures such as AlexNet, ResNet-18, and MobileNet-V2. Input to each of these CNN architectures is an RGB image resized to the shape of (224×224×3). The output of each state-of-the-art CNN network is given to a specific Conv layer whose kernel size is uniquely chosen for a specific CNN architecture. The purpose of the Conv layer is to transform the

feature maps from any CNN architecture to the dimensions of (1×1×256). Hence, we add a Conv layer of kernel size (6×6) with output filter maps of 256 to AlexNet. A conv layer of kernel size (1×1) with 256 output filter maps is added to ResNet-18, and a Conv layer with a kernel size of 7×7 and 256 output filter maps is added to MobileNet-V2. Finally, the feature of shape (1×1×256) from the CNN is applied to a flattening layer to obtain a 1-dimensional vector of size 256.

The 1-dimensional vector from the KEN and the CNN branch, each having a dimension of 256, is concatenated in the Fusion Net to give a fused vector of shape 512. This is followed by a fully connected layer reducing the feature dimension to 256, followed by the ReLU activation, and a second fully connected layer to reduce the feature dimension to 128, which is again followed by a ReLU activation. Lastly, the 128-dimensional vector is applied to a final fully connected layer, which reduces it from 128 to 46 (or 41) units depending on the number of classes in the dataset.

Overall, the KEN network receives handcrafted features and processes the features to generate a 1-dimensional vector. In the same way, the CNN branch processes input images to generate feature vectors. Each vector has a dimension of 256 and is concatenated in the Fusion Net to give a fused vector of the shape of 512 dimensions. This is followed by a fully connected layer reducing the feature dimension to 256, followed by the ReLU activation. Here, the fusion operation is simple concentration.

The proposed architecture is trained using cross-entropy loss as the loss function and Adadelta as the optimizer with a learning rate of 0.001. We have also used ReduceLROnPlateau as the scheduler to control the learning rate and obtain a higher classification rate. The model is trained using a batch size of 50 sample images. Furthermore, the model is trained for 100 epochs using AlexNet/ResNet-18/MobileNetv2, which are pre-trained on ImageNet [39-42].

4. Experimental Results

To conduct experiments for evaluating the proposed model, we divide the experimental section into the following sub-sections. In Section 4.1, we present details of our dataset creation, standard dataset, and performance measures. Section 4.2 discusses the performance of the proposed and existing methods on different datasets. The ablation study to show the effectiveness of the critical steps of the proposed model to achieve the best results for the classification of photos is presented in Section 4.3. In Section 4.4., discussion on the merits and demerits of the proposed model as well as the possible solutions to overcome the limitation are presented.

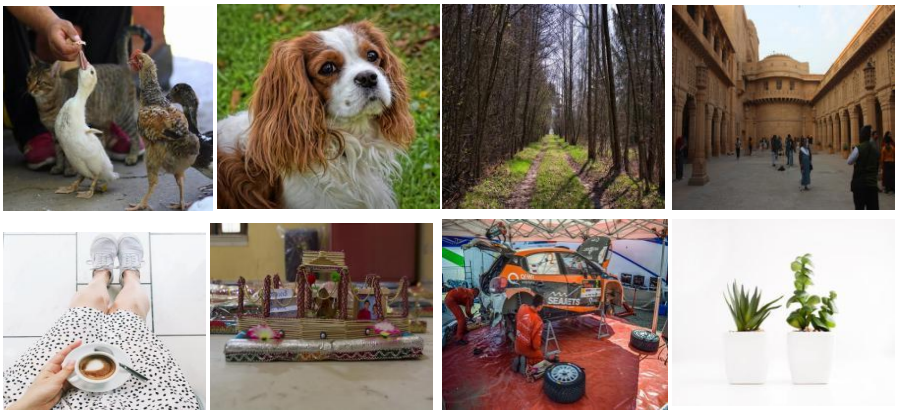
4.1. Dataset and Evaluation

To validate the performance of our method, we collected images captured by 46 photographers, resulting in a 46-class classification problem. To collect this data, we approached professional photographers online from various sources, irrespective of gender, age, or country. The photographers were asked to share their photos and attest to their authenticity. In our dataset, the main challenge is

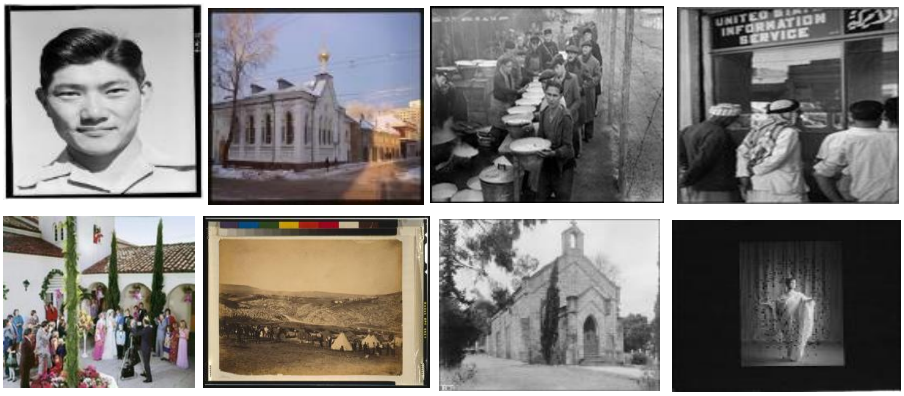
that photographers' images share properties because there is a high chance of overlap in the types of subjects and their skills in the art. Therefore, more variations in intra-images and fewer variations in inter-images are expected.

Furthermore, there is no constraint in this dataset collection that a photographer must use the same camera to capture the whole event or the same kind of scenes. The photographer can capture images according to their mind, habit, passion, and interest. This is another difficulty that makes the classification problem more complex and challenging. It is evident from the description given for each photographer of the 46 classes in Table 1, where we note that each photographer has their style of capturing images. The number of images for each class is listed in Table 2, where the number of images for each class varies from 533 to 1815. The dataset provides 46234 images, which is sizable for experimentation and evaluation of our method. Sample images for a few classes of our dataset are shown in Fig. 6(a), where we can see foreground and background variations in the images.

To show the effectiveness of the proposed model, we also considered a standard dataset, which is available to the public [14] and contains 41 photographer classes. The numbers of images for each class are listed in Table 3, where it is noted that the size of a class varies from 126 to 28475, which gives 218303 images in total. As noted in [14], these photographers are professionals and famous for photography. Therefore, the images captured by them are expected to have unique characteristics with less variation than expected in the intra-images of the classes. As a result, one can argue that our dataset is much more complex than the standard one, although the size is substantial compared to ours. Sample images of a few classes of the standard dataset are shown in Fig. 6(b), where we can sense that images of different classes represent unique scenes, which eases the complexity of the problem.



(a) Sample images of different classes from our dataset.



(b) Sample images of different classes from the standard dataset.

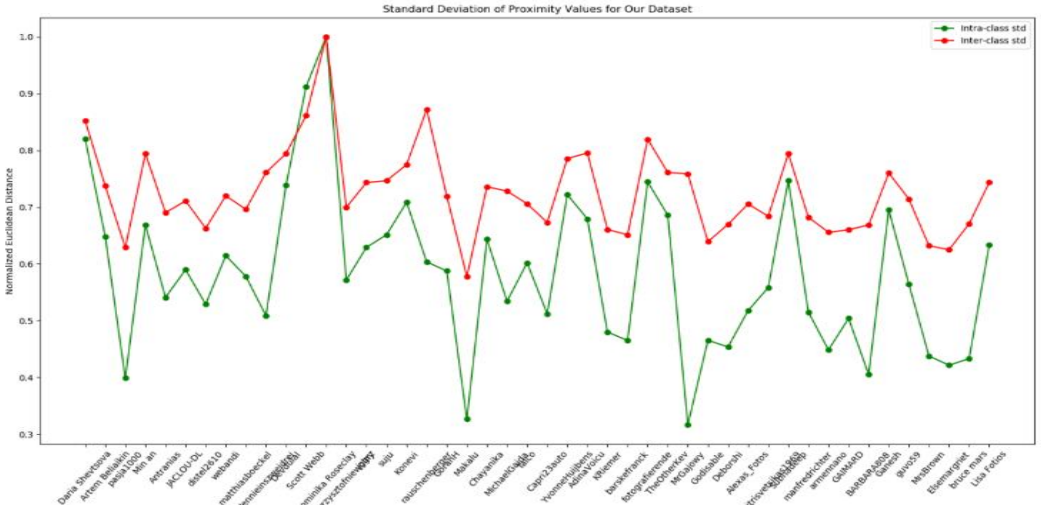
Fig. 6. Examples of different classes from ours and the standard datasets

To analyze the complexity of ours and the standard dataset, we generated proximity matrices by estimating the similarity between intra- and inter-images of the classes. The similarity is determined between the first image with all the other images in the same class, the second image with all the other images in the same class, and so on, using the Euclidean distance measure. This process produces proximity matrices for each class of intra-and inter-images. Then we compute the standard deviation for the proximity matrices, which results in two sets containing standard deviation values for each class of intra- and inter-images, respectively. If variations in intra- and inter-images are high, then a high standard deviation is expected with values for those classes. The classes can expect constant standard deviation values if variations are not high for intra- and inter-images. It is illustrated in Fig. 7(a) and Fig. 7(b), where it is confirmed that significant variations for intra- and inter-images are found in our dataset.

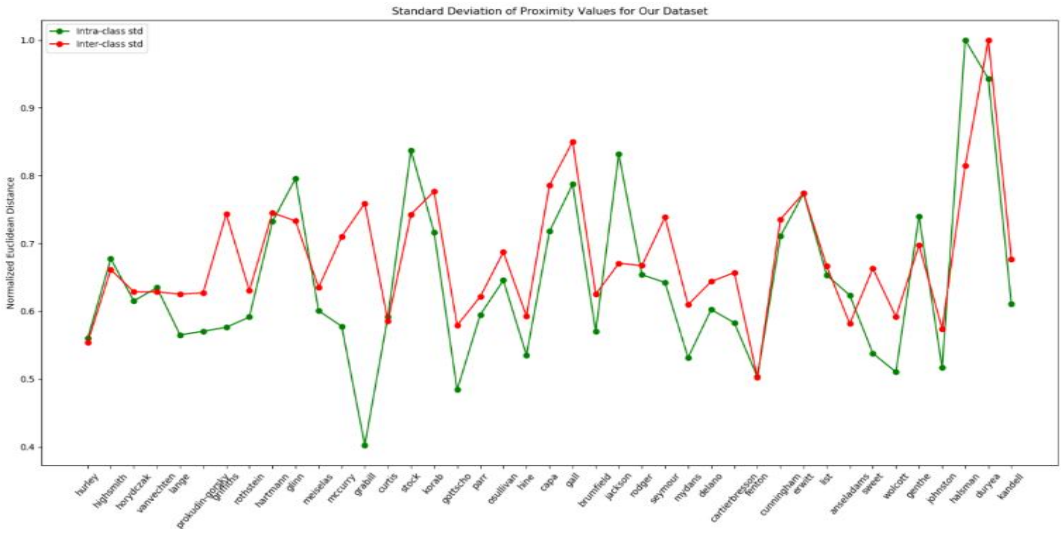
In contrast, the variations are low for intra- and inter-images of the standard dataset. Therefore, we find arbitrary behaviors for graphs of intra- and inter-images of our dataset, while there is smooth behavior for graphs of intra- and inter- images on the standard dataset. This demonstrates that our dataset is much more complex than the standard dataset.

Since only one method [14] is available in the literature for classifying a photographer's images, we have implemented the same for a comparative study in this work. The technique explores low-level

features extracted by descriptors and high-level features extracted by deep learning models for classification. The reason for choosing this method is that the objective of the method [14] is the same as ours, and it is the state-of-the-art method for classifying photographer's images. In addition, it uses a combination of handcrafted and deep features achieving results comparable to our method, which considers both handcrafted and deep features.



(a) The Standard deviation of the proximity matrix of intra- and inter-classes of our dataset. The red color represents inter-images, and the green color represents intra-images.



(b) The Standard deviation of the proximity matrix of inter- and intra-classes of the standard dataset. The red color represents inter-images, and the green color represents intra-images

Fig. 7. Similarity between inter- and intra-images of classes of our and standard datasets

Table 1. Description of each class of photographer from our dataset.

Photographer	**Description**
AdinaVoicu	Takes images of natural beauty
Alexas_Fotos	It takes images of animals in need - whether they were intended for slaughter, neglected, abandoned, or couldn't survive in the wild zoo. natural beauty
Antranias	Photos/pictures of all perspectives of man and nature

Armennano	Takes images of natural scene
ArtemBeliaikin	Lifestyle, event documentary photography.
Barbara808	Natural scene images, flowers, pet animals
Barskefranck	Amateur photographer, varieties images include cityscape, people, animals, flowers
Brucemars	Takes images of people working and involved in some actions
Capri23	Images of flowers, animals, birds
Chayanika	Photographer captures different places when they visit tourist places
Daria Shevtsova	Photoshoot images
Deborshi	A professional photographer usually captures wedding events and wildlife
Devdulal	Normally captures wildlife image
Dimitrisvetsikas	Images of sea, nature and the history and tradition, sculptures, paintings, graffiti, people's faces
Distel	Images are from varied fields, nature, landscape, flowers
Dominikaroseclay	Images of objects
Elsemargriet	Images of scenes, animals, flowers
Falco	Images of German structures
Fotografierende	Images of different objects
Gaimard	Images of flowers, birds, structures
Ganesh	A professional photographer usually captured images of institutional events
Godisable	Fashion photography
GoranH	Images of agricultural fields, flowers, statues
Guvo	Images of pet animals, landscapes, objects
Jaclou-Dl	Images of pet animals, flowers, landscapes, and people involved in the activity
Jggrz	Images of natural scenery
Konevi	Images of buildings,scenery,flower(majority rose)
Kriemer	Images of landscapes and buildings of Germany, flowers
Krzysztofniewolny	Images of mainly macro, sport, walking, travel, techno music, medieval castles, and animals
Lisa Fotios	Images of different objects
Makalu	Travel photography, flowers
Manfredrichter	Images of flowers, animals, fruits
Matthiasboeckel	Images of fruits, flowers, buildings
Michaelgaida	Images of different architectures, abandoned places, flowers, forests, landscapes, nature, railway stations, roads, trees
Min an	Studio Photoshoot images, outdoor photography
Mrgajowy	Images of flowers
MrsBrown	Images of monuments,church,statues,birds,petanimals,flower
Nennieinszweidrei	Images of flowers, birds, pet animals, Christmas
Pasja	Images of Switzerland
Rauschenberger	Images of scenes against light
Scott Webb	Images of buildings, plants, underwater
Subhadeep	Captures photographs randomly according to his interest
Suju	Images of birds, scenery, fruits
TheOtherkev	Images of birds, wildlife
Webandi	Images of bikes, flowers, mountains
YvonneHujibens	Images of nature and wildlife

Table 2. Different classes and their respective sizes in our dataset (N denotes the total number of images for the class, and

SLNo. denotes serial no).

SLNo.	Class	N	SLNo.	Class	N	SLNo.	Class	N
1	Daria Shevtsova	940	17	Konevi	1020	33	Deborshi	994
2	ArtemBeliaikin	1230	18	rauschenberger	825	34	Alexas_Fotos	1080
3	pasja1000	872	19	GoranH	1020	35	dimitrisvetsikas1969	1020
4	Min an	1815	20	Makalu	1020	36	Subhadeep	1000
5	Antranias	1020	21	Chayanika	746	37	manfredrichter	1020
6	JACLOU-DL	1020	22	MichaelGaida	1500	38	armennano	533
7	distel2610	1020	23	falco	1020	39	GAIMARD	803
8	webandi	1020	24	Capri23auto	1020	40	BARBARA808	902
9	matthiasboeckel	1020	25	YvonneHuijbens	1215	41	Ganesh	1224
10	Nennieinszweidrei	959	26	AdinaVoicu	1020	42	guvo59	904
11	Devdulal	1158	27	KRiemer	1020	43	MrsBrown	1019
12	Scott Webb	638	28	barskefranck	1020	44	Elsemargriet	975
13	Dominika Roseclay	708	29	fotografierende	772	45	bruce mars	644
14	krzysztofniewolny	1020	30	TheOtherKev	1164	46	Lisa Fotios	1314
15	jggrz	1175	31	MrGajowy	1020			
16	suju	1020	32	Godisable	765			

Table 3. Different classes and their respective sizes in the standard dataset (N denotes the total number of images for the class, and SLNo. denotes serial no).

SLNo.	Class	N	SLNo.	Class	N	SLNo.	Class	N
1	Adams	245	15	Halsman	1310	29	Parr	20635
2	Duryea	152	16	Johnston	6962	30	Van Vechten	1385
3	Grabill	189	17	Mydans	2461	31	Curtis	1069
4	Hurley	126	17	Stock	3416	32	Glinn	4529
5	McCurry	6705	19	Bresson	4693	33	Hine	5116
6	Rothstein	12517	20	Gall	656	34	Lange	3913
7	Brumfield	1138	21	Hartmann	2784	35	Prokudin-Gorsky	2605
8	Erwitt	5173	22	Kandell	311	36	Wolcott	12173
9	Griffiths	2000	23	O'Sullivan	573	37	Delano	14484
10	Jackson	881	24	Sweet	909	38	Gottscho	4009
11	Meiselas	3051	25	Cunningham	406	39	Horydczak	14317
12	Seymour	1543	26	Genthe	4140	40	List	2278
13	Capa	2389	27	Highsmith	28475	41	Rodger	1204
14	Fenton	262	28	Korab	764			

We use the standard measures, namely, Precision (P), Recall (R), and F-Measure (F), for measuring the performance of our and existing methods. The measures are defined in Equation (9) to Equation (11). In addition, we also calculate the Overall Score, which considers images of all the classes classified correctly by our method divided by the actual number of images in the dataset. More instructions to define the measures mentioned above can be found in [14].

The measures used to calculate accuracy for each class as defined in Equation (10) to Equation (12) are: Precision (P_i), Recall (R_i) and F-Measure (F_i) for Class i. Let $\varphi_{i,j}$ represent the number of images belonging to class i and being predicted as class j by the network. Therefore:

$$P_i = \frac{\varphi_{i,i}}{\sum_{\forall j} \varphi_{j,i}} \tag{10}$$

$$R_i = \frac{\varphi_{i,i}}{\sum_{\forall j} \varphi_{i,j}} \tag{11}$$

$$F_i = \frac{2 * P_i * R_i}{(P_i + R_i)} \tag{12}$$

Mean Precision (P_{mean}) is the average of all class-wise Precision values, Mean Recall (R_{mean}) is the average of all class-wise Recall values, and Mean F-Measure (F_{mean}) is calculated using the Mean Precision and Mean Recall values.

If the F-measure is the performance of the proposed method is good. If Recall is high while the precision is low, the proposed model is good in quantity but not quality. On the other hand, if the Recall is low and precision is high, the proposed model is not good in quantity, but the system is perfect in terms of quality.

4.2. Evaluating the Proposed Classification Method

Quantitative results of the proposed model and existing methods for our dataset of all 46 classes and standard dataset of 41 classes are reported in Table 4. It is observed from Table 4 that our method is the best in the overall score compared to the existing method. The current approach's poor performance is because the extracted features cannot handle our dataset's complexity. The technique is developed for images captured by professional photographers. This constraint is not valid in the case of our dataset, where we can see classes of professional and non-professional photographers.

On the other hand, the proposed model involves features extraction from the fusion of focused and defocused regions through KEN and MobileNetV2 yields a difference compared to the existing method. However, the overall score of our method on our dataset is not very high. This is evident that the dataset is complex because of diversity; hence, further investigation is needed to achieve good results, which is beyond the scope of this proposed work.

It is observed from the results of the proposed and existing model on a standard dataset in Table 5 that the proposed approach is the best in terms of the overall score compared to the existing method. Although the existing method was developed for 41 classes of this dataset, it reports poor results compared to the proposed approach. This is because the features used in the existing method have inherent limitations, which are sensitive to variations in images. At the same time, the proposed approach extracts high-level features, namely, context, which is the relationship between focused and defocused regions through the KEN model. However, when we compare the results of our dataset and the standard dataset, both the proposed and the existing methods score poorly on our dataset but demonstrate better results for the standard dataset. This shows that our dataset is complex compared to the standard dataset. As illustrated in Fig. 7(a) regarding the standard deviation of the proximity matrix of inter- and intra-classes, our dataset involves more heterogeneous images due to a greater diversity of photographers. Our dataset is complex and challenging compared to the standard dataset. In the same way in a standard dataset, since images captured by professional photographers with fewer classes compared to our dataset, the dataset does not involve more diversified images than our dataset, as

illustrated in Fig. 7(b)in terms of the standard deviation of the proximity matrix of inter- and intra-classes.

Table 4 shows that the proposed model achieves consistent precision, Recall, and F-measure for our and standard datasets. Therefore, one can infer that the proposed model provides stable and reliable results for different datasets.

Table 4. Mean precision, mean Recall, and mean F-Measure of the proposed and existing models on our and standard dataset

Datasets	Proposed Model			Existing Model [14]		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Our Dataset	**0.50**	**0.50**	**0.50**	0.44	0.44	0.44
Standard Dataset	**0.76**	**0.73**	**0.74**	0.69	0.66	0.67

4.3. Ablation Study

The proposed work comprises vital steps: feature extraction from focused and defocused regions, combining focused and defocused regions, and exploring MobileNetV2 for extracting deep features. We conduct the following experiments to show that the key steps effectively achieve better photographer classification.

Experiment (i): Features are extracted from the segmented focused regions of the input images. Then the extracted features are supplied to the proposed KEN for classification. (Table 5(i)).

Experiment (ii): Features are extracted from the segmented defocused region, and then the feature matrix is fed to the proposed KEN for classification. (Table 5 (ii)).

Experiment (iii): Extracted features from focused and defocused regions are combined using the fusion operation defined in our methodology through a cross co-variance matrix, resulting in the proposed context feature matrix. These features are fed to KEN to select dominant features, and then the input image is passed to MobileNetV2 to extract the deep features. (Table 5(iii)).

Furthermore, the proposed context and deep features are combined to classify the photographer's images. The measures are calculated for each experiment on our benchmark datasets [14], and the results are reported in Table 6. It is observed from Table 6 that the focused and defocused features almost contribute equally in terms of measures for the classification of the photographer. Therefore, one can infer that the focused and defocused features effectively achieve better results for classification. It is evident from the results of our method, which combines both focused, defocused, and deep features reported in Table 6, that our method is better than individual features. The ablation experiments also noted that individual features alone are inadequate to achieve the best results for classification compared to our method.

Table 5. Analyzing the contribution of the critical steps of our method using our dataset (46 classes) and benchmark dataset (41 classes)

Steps	(i) Only Focused Features			(ii) Only Defocused Features			(iii) Proposed (Focused +Defocused + Deep Features)		
Measures	P	R	F	P	R	F	P	R	F
Our dataset	0.37	0.38	0.38	0.41	0.42	0.42	0.50	0.50	0.50
Benchmark	0.66	0.62	0.64	0.71	0.68	0.69	**0.76**	**0.73**	**0.74**

Experiment (iv): To show that the use of MobileNetV2 [39] is effective for photographer identification compared to other popular networks, we calculated precision, recall, and F-measure for well-known architectures, namely, AlexNet [36], ResNet [47], and SqueezeNet [48] using our dataset (Table 6). The AlexNet and ResNet are basic deep learning models for extracting distinct features for the classification problem, while SqueezeNet is popular in terms of speed, similar to MobileNetV2. However, the MobileNetV2 is better in terms of accuracy and the number of computations than other architectures. It is evident from the results reported in Table 6 for our dataset, where we can see the overall score of MobileNetV2 is better than those of the other three architectures. The reason for obtaining good results with MobileNetV2 and poor results with different architectures are as follows. Unlike AlexNet and ResNet, which rely on 3×3 convolutional layers, MobileNetV2 relies on 1×1 convolution layers. This results in a reduction in computational cost. This is coupled with Depth-wise Convolution, where the convolution operation is performed independently for each input channel, thereby reducing computational complexity by omitting convolutions in the channel domain.

In summary, the generalization ability increases due to the lower training parameters of MobileNetV2 compared to AlexNet and ResNet-18. On the other hand, although, SqueezeNet is faster in computations, the accuracy of the models depends on a large number of samples, and there are high chances of causing overfitting for fewer samples, especially for a complex problem. In the case of MobileNetV2, the model can solve a complex classification problem without demanding many samples. Thus, MobileNetV2 is better for the proposed complex photographer's photos classification in terms of accuracy and number of computations than the other architectures.

Table 6. Mean precision (P), mean recall (R), and mean F-measure (F) of the proposed model and baseline architectures on our dataset (Experiment (iv))

Proposed architecture			AlexNet			ResNet			SqueezeNet		
P	R	F	P	R	F	P	R	F	P	R	F
0.50	**0.50**	**0.50**	0.44	0.44	0.44	0.37	0.38	0.38	0.36	0.37	0.37

4.4. Limitation of the Proposed Model

It is noted in Table 4 that the proposed method does not achieve high results, like more than 90%. Instead, it achieves good results for both datasets. Although the proposed model is consistent for different photo classes by different photographers, the results are not very high. The key reason is that the classification of photos captured by different photographers is a complex problem because of diversified images. In addition, the photographer can capture any scene depending on their interest, mind, target, and hobby. Therefore, this classification problem involves more subjectivity than

objectivity. In the case of our dataset, we have considered photos captured by both professional and non-professional photographers. As a result, the dataset can include images affected by adverse factors, such as noise, blur due to defocus, and motion. However, the standard dataset includes photos captured by a professional photographer. It is evident from Table 4 that the proposed and existing models do not report high results for our dataset while high results for the standard dataset. As a result, we can conclude that achieving stable, convincing, reliable, and meaningful results for the above situations is beyond the scope of the proposed work. Hence, it is an open challenge for the researchers. Therefore, there is a scope for further improvement soon. To improve the results of the proposed method, we need to focus on local information in the focused and defocused regions rather than considering fully focused and defocused regions. This is because it is true that fully focused and defocused regions may include unimportant information, which may lead to poor performance. Furthermore, we believe that if we design models by considering the photographer's characteristics, hobbies, and interests along with the visual features of the photos, we can achieve better results. Therefore, we plan to design a transformer to integrate the photographer's interest, hobby, and objective and photos' visual features to improve the results.

5. Conclusion and Future Work

In this work, we have proposed a novel approach for classifying photos captured by different photographers. Our model segments the given input image into focused and defocused regions by exploring low and high-pass kernels in a new way. It also performs new fusion operations to combine features extracted from focused and defocused regions as a single feature vector through cross-covariance features, which extract context information between focused and defocused regions. Furthermore, this feature vector combines features extracted from the input images using MobileNetV2 through a new architecture for classifying photographers' photos. Experimental results of evaluating the proposed model and existing approaches on ours and the standard datasets demonstrate that our method outperforms the existing method for both datasets in terms of the overall score. However, the results obtained from our dataset are low due to the diverse range of images and more classes. Our future target is to identify images captured by male and female photographers and professional and non-professional photographers to improve the results.

Acknowledgment

Ministry of Higher Education of Malaysia funded this project with a generous grant from the Fundamental Research Grant Scheme (FRGS) with code number FRGS/1/2020/ICT02/UM/02/4. Grants from the Natural Science Foundation of China under No. 61672273 and the ISI-UTS Joint Research Cluster (TIH) also supported this project. The authors thank Chandrakant Saha, Monalisa Nayak, and Swati Kanchan for their help in creating, annotating datasets, and experimenting.

References

[1] L. Nandanwar, P. Shivakumara, U. Pal. T. Lu, D. Lopresti, B. Seaogi and B. B. Chaudhuri, "A new for detecting altered text in document images", International Journal of Pattern Recognition and Artificial Intelligence, 35(12), 2021.

[2] K. Biswas, P. Shivakumara, U. Pal, T. Chakraborti, T. Lu and M. N. B. Ayub, Fuzzy and genetic algorithm-based approach for classification of personality traits oriented social media images, Knowledge-Based Systems, 241, 2022. https://doi.org/10.1016/j.knosys.2021.108024.

[3] I. Amerini, C. T. Li and A. R. Caldelli, Social network identification through image classification with CNN, IEEE Access, 7, pp 35264-35273, 2019.

[4] Q. Wang, G. Jin, X. Zhao, Y. Feng and J. Huang, CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale, Knowledge-Based Systems, 189, 2020.

[5] L. Yan, W. Zheng, C. Gou and F. Y. Wang, IPGAN: Identify-preservation generative adversarial network for unsupervised photo-to-caricature translation, Knowledge-Based Systems, 241, 2022.

[6] K. Pastra, H. Saggion and Y. Wilks, Extracting relational facts for indexing and retrieval of crime scene photographs, Knowledge-Based Systems, pp 313-320, 2003.

[7] Y. S. Rawat and M. Kankanhalli, ClickSmart: A context-aware viewpoint recommendation system for mobile photography, IEEE Trans. CSVT, pp 149-158, 2017.

[8] L. Zhang, M. Song, Q. Zhao, X. Liu, J. Bu, and C. Chen, Probabilistic graphlet transfer for photo cropping, IEEE Trans. IP, pp 802-815, 2013.

[9] X. Qian, C. Li, K. Lan, X. Hou, Z. Li, and J. Han, POI summarization by aesthetics evaluation from crowd source social media, IEEE Trans. IP, pp 178-01189, 2018.

[10] D. Li, H. Wu, J. Zhang, and K. Huang, Fast A3RL: Aesthetics-aware adversarial reinforcement learning for image cropping, IEEE Trans. IP, pp 5105-5120, 2019.

[11] T. Qiao, R. Retraint, R. Cogranne and T. H. Thai, Individual camera device identification from JPEG images, Signal Processing: Image Communication, 52, pp 74-86, 2017.

[12] M. Sun, Z. Liu, J. Qiu, Z. Zhang and M. Sinclair, Active lighting for video conferencing, IEEE Trans. CSVT, pp 1819-1829, 2009.

[13] T. Yan, Y. Mao, J. Wang, W. Liu, X. Qian and R. W. H. Lau, Generating stereoscopic images with convergence control ability from a light field image pair, IEEE Trans. CSVT, pp 1435-1450, 2020.

[14] C. Thomas and K. Kovashka, Seeing behind the camera: Identifying the authorship of a photograph, In Proc. CVPR, pp 3494-3502, 2016.

[15] N. Sun, W. Li, G. Han and C. Wu, Fusing object semantics and deep appearance features for scene text recognition, IEEE Trans. CSVT, 29, pp 1715-1727, 2019.

[16] Y. Pan, Y. Xia and D. Shen, Foreground fisher vector: Encoding class-relevant foreground to improve image classification, IEEE Trans. IP, 28, pp 4716-4729, 2019.

[17] X. Sun, L. Zhang, Z. Wang, J. Chang, Y. Yao, P. Li and R. Zimmermann, Scene categorization using deeply learned gaze shifting kernel, IEEE Trans. Cybernetics, 49, pp 2156-2167, 2019.

[18] B. Wang, X. Hu, C. Zhang, P. Li and P. S. Yu, Hierarchical GAN-Tree and Bi-directional capsules for multi-label image classification, Knowledge-Based Systems, 238, 2022.

[19] V. Venugopal and S. Sundaram, Modified Sparse Representation Classification Framework for Online Writer Identification, in IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp 34-325, 2018.

[20] V. Venugopal and S. Sundaram, Online writer identification with sparse coding-based descriptors, IEEE Trans. IFS, 13, pp 2538-2552, 2018.

[21] W. Yang and Y. Hui, "Image scene analysis based on improved FCN model", International Journal on Pattern Recognition and Artificial Intelligence, Vol. 35, pp 1-17, 2021.

[22] D. Zhang, Y. Zhou, J. Zhao and Y. Zhou, "Co-evolution-based parameter learning for remote sensing scene classification", International Journal on Pattern Recognition and Artificial Intelligence, Vol. 20, pp 1-15, 2022.

[23] W. H. Cheng, Y. Y. Chuang, Y. T. Lin, C. C. Hsieh, S. Y. Fang, B. Y. Chen and J. L. Wu, "Semantic analysis for automatic event recognition and segmentation of wedding ceremony videos", IEEE Transactions on Circuits and Systems for Video Technology, pp 1639-1650, 2008.

[24] C. Beyan, A. Zunino, M. Shaid and V. Muino, "Personality traits classification using deep visual activity-based nonverbal features of key-dynamic images", IEEE Transactions on Affective Computing, pp 1084-1099, 2021.

[25] D. Krishanani, P. Shivakumara, T. Lu, U. Pal and R. Ramachandra, "Structure function-based transform features for behavior-oriented social media image classification", In Proc. ACPR. 2019.

[26] D. Krishnani, P. Shivakumara, T. Lu, U. Pal, D. Lopresti, and G. H. Kumar, Multimedia Tools and Applications, 2021.

[27] L. Zhang, S. Peng and S. Winkler, "PersEmoN: A deep network for joint analysis of apparent personality, emotions and their relationship", IEEE Transaction on Affective Computing, pp 298-305, 2022.

[28] X. Sun, J. Huang, S. Zheng, X. Rao and M. Wang, "Personality assessment based on multimodal attention network learning with category-based mean square error", IEEE Transactions on Image Processing, pp 2162-2174, 2022.

[29] L. Liu, D. Preoţiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice", In Proc. AAAI, 2016.

[30] H. Zhu, L. Li, S. Zhao, and H. Jiang, "Evaluating attributed personality traits from scene perception probability", Pattern Recognition Letters, vol. 116, pp. 121–126, 2018.

[31] L. G. Villalba, A. L. S. Orozco, J. R. Corripio and J. H. Castro, A PRNU based counter forensic method to manipulate smartphone image source identification techniques, Future Generation Computer Systems, 76, pp 418427, 2017.

[32] X. Ding, Y. Chen, Z. Tang and A. Y. Huang, Camera identification based on domain knowledge driven deep multi-task learning, IEEE Access, 7, pp 25878-25890, 2019.

[33] B. Wang, K. Zhong, Z. Shan, M. N. Zhu and X. Sui, A unified framework of source camera identification based on features, Forensic Science International, 307, 2020.DOI: 10.1016/j.forsciint.2019.110109.

[34] X. Yuan, H. Han and L. Huang, "Association loss and self-discovery cross-camera anchors detection for unsupervised video-based person re-identification", International Journal on Pattern Recognition and Artificial Intelligence, Vol. 35, 2021.

[35] G. K. Birajadar and M. D. Patil, "A systematic survey on photorealistic computer graphics and photographic image discrimination", International Journal on Pattern Recognition and Artificial Intelligence, Vol. 22, pp 1-35, 2023.

[36] J. D. Rugna, G. Chareyron and B. Branchet, Tourist behavior analysis through geotagged photographs: A method to identify the country of origin, In Proc. ISCII, pp 347-351, 2012.

[37] P. Angelov, J. Andreu and T. Vuong, Automatic mobile photographer and picture diary, In Proc. CEAIS, 2012.

[38] Y. Hoshen and S. Peleg, An egocentric look at video photographer identity, In Proc. CVPR, pp 4284-4292, 2016.

[39] M Sandler, A Howard, M Zhu, A Zhmoginov and L C Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks", In Proc. CVPR, pp 4510-4520, 2018.

[40] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, pp. 1097-1105, 2012.

[41] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, In Proc. CVPR, pp 248-255, 2009.

[42] R. Annamaria, V. Koczy, A. Rovid and T. Hashimoto, Gradient based synthesized multiple exposure time color HDR image, IEEE Trans. TIM, 57, pp 1779-1785, 2008.

[43] J. Dou, Q. Qin and Z. Tu, Image fusion-based wavelet transform with genetic algorithms and human visual system, Multimedia Tools and Applications, 78, pp 12491-12517, 2019.

[44] L. Qi, X. Lu and X. Li, Exploiting spatial relation for fine-grained image classification, Pattern Recognition, 91, pp 47-55, 2019.

[45] A. N. J. Raj. C. Junmin, R. Nersission, V. G. V. Mahesh and Z. Zhuang, "Bilingual text detection from natural scene images using faster R-CNN and extended histogram of oriented gradients", Pattern Analysis and Applications, pp 1-13, 2022.

[46] R. Nersission, T. J. Iyer, A. N. J. Raj and V. Rajanagm, "A dermoscopic skin lesion classification technique using YOLO-CNN and traditional feature model", Arabian Journal for Science and Engineering, pp 9797-9808, 2021.

[47] K. He, X. Zhang, S. Re and J. Sun, Deep Residual Learning for Image Recognition, In Proc. CVPR, pp 770-778, 2016.

[48] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally and K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50X fewer parameters and M 0.5 MB Model size. arXiv:1602.07360.

Palaiahnakote Shivakumara is an Associate Professor at the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. Previously, he was with the Department of Computer Science, School of Computing, the National University of Singapore from 2008-2013 as a Research Fellow on a Video text extraction and recognition project. He received B.Sc., M.Sc., M.Sc Technology by research and Ph.D. degrees in computer science, respectively, in 1995, 1999, 2001, and 2005 from the University of Mysore, Karnataka, India. He has been serving as Editor-in-Chief for Artificial Intelligence and Applications (AIA) since 2022. He has been serving as Associate Editor for IEEE Transactions on Multimedia (TMM), Pattern Recognition (PR), International Journal on Document Analysis and Recognition (IJDAR), Springer Nature of Computer Science (SNCS), ACM Transactions Asian and Low-Resource Language Information Processing (TALLIP), Malaysian Journal of Computer Science (MJCS) and CAAI-Transactions on Intelligence in Technology. He received a prestigious "Dynamic Indian of the Millennium" award from the KG foundation, India for his research contribution to computer science. He has published more than 300 papers in conferences and journals. His research interests are image processing, Document image analysis, and Video text processing.

Pinaki Nath Chowdhury worked as a project-linked person at the computer vision and pattern recognition unit, Indian Statistical Institute, Kolkata, India. Currently, he is pursuing a Doctor of Philosophy (Ph.D.) at the SketchX Lab of Center of Vision, Speech and Signal Processing (CVSSP), University of Surrey, United Kingdom.

Dr. Umapada Pal has been a Faculty member of the Computer Vision and Pattern Recognition Unit (CVPRU) of the Indian Statistical Institute, Kolkata, since 1997. He is currently a Professor at CVPRU. His fields of research interest are different pattern recognition and computer vision problems like Digital document analysis, Camera/video text processing, Biometrics, Image retrieval, Keyword spotting, Video analysis, Medical image analysis, Pose estimation, Image/video generation, etc. He has published more than 470 research papers in various international journals, conference proceedings, and edited volumes. Because of his significant impact on Document Analysis research, in 2003, he received the "ICDAR Outstanding Young Researcher Award" from International Association for Pattern Recognition (IAPR). He also received many fellowships from different countries. He is actively engaged with the PR community. Prof. Pal is serving as an EIC of the Springer Nature Computer Science journal and AE of 6 journals, including the Pattern Recognition journal. He is a Fellow of the International Association for Pattern Recognition (IAPR), the Asia-Pacific Artificial Intelligence Association (AAIA), West Bengal Academy of Science and Technology, etc. Also, he is among the top two percent of scientists in the world, as listed by Stanford University in 2020 and 2021.

Dr. David Doermann is a Professor of Empire Innovation and the Institute for Artificial Intelligence and Data Science Director at the University at Buffalo (UB). Before coming to UB, he was a Program Manager with the Information Innovation Office at the Defense Advanced Research Projects Agency (DARPA). At DARPA, he developed, selected, and oversaw research and transition funding in computer vision, human language technologies, voice analytics, and media forensics. From 1993 to 2018, David was a research faculty member at the University of Maryland, College Park. In his role in the Institute for Advanced Computer Studies, he served as Director of the Laboratory for Language and Media Processing and as an adjunct member of the graduate faculty for the Department of Computer Science

and the Department of Electrical Engineering. He and his group of researchers focused on many innovative topics related to document images and video analysis and processing, including triage, visual indexing and retrieval, enhancement, and recognition of visual media's textual and structural components. His recent research has focused on advanced AI techniques applied to computer vision, medical image analysis, federated learning, neural architectural search, binary neural networks, and detecting false and misinformation in multimedia content. David has over 300 publications in conferences and journals, is a fellow of the IEEE and IAPR, has numerous awards, including an honorary doctorate from the University of Oulu, Finland, and is a founding Editor-in-Chief of the International Journal on Document Analysis and Recognition.

Dr. Raghavendra Ramachandra obtained a Ph.D. in computer science and technology from the University of Mysore, Mysore, India, and Institute Telecom and Telecom Sudparis, Evry, France (carried out as collaborative work) in 2010. He is currently a full professor at the Institute of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Gjøvik, Norway. He is also working as R\&D chief at MOBAI AS. He was a researcher with the Istituto Italiano di Tecnologia, Genoa, Italy, where he worked with video surveillance and social signal processing. His main research interests include deep learning, machine learning, data fusion schemes, and image/video processing, with applications to biometrics, multimodal biometric fusion, human behavior analysis, and crowd behavior analysis. He has authored several papers and is a reviewer for several international conferences and journals. He also holds several patents in biometric presentation attack detection and morphing attack detection. He has also been involved in various conference organizing and program committees and as an associate editor for various journals. He has participated (as a PI, co-PI, or contributor) in several EU projects, IARPA USA, and other national projects. He is an editor of the ISO/IEC 24722 standards on multimodal biometrics and an active contributor to the ISO/IEC SC 37 standards on biometrics. He has received several best paper awards and is also a senior member of IEEE.

Tong Lu received a Ph.D. in computer science from Nanjing University in 2005. He received his M.Sc. and B.Sc. degrees from the same university in 2002 and 1993, respectively. He served as an Associate Professor and Assistant Professor in the Department of Computer Science and Technology at Nanjing University from 2007 and 2005. He is now a full Professor at the same university. He also has served as Visiting Scholar at the National University of Singapore and the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, respectively. He is also a member of the National Key Laboratory of Novel Software Technology in China. He has published over 130 papers, authored two books in his area of interest, and issued more than 20 international or Chinese invention patents. His current interests are multimedia, computer vision, and pattern recognition algorithms/systems.

Professor Michael Blumenstein is currently the Deputy Dean (Research and Innovation) in the Faculty of Engineering & IT at UTS, where he previously held the positions of Associate Dean (Research Strategy and Management) and, before that, the Head of the School of Software (now Computer Science). Michael is a nationally and internationally recognized expert in automated Pattern Recognition and Artificial Intelligence. His current research interests include Marine Animal and Wildlife Detection from Video Imagery, Document Image Analysis, Video-based/multi-lingual Text Detection, and Signature Verification. He has published over 300 papers in refereed books, conferences, and journals. His research also spans various projects applying Artificial Intelligence to Engineering, Environmental Science, Neurobiology, and Coastal Management. Michael has secured internal/nationally competitive research grants to undertake these projects with funds exceeding AUD$6.5 Million. Components of his research into the predictive assessment of beach conditions have been commercialized for use by local government agencies, coastal management authorities, and commercial applications. In addition, Michael's team has developed SharkSpotter, a world-first AI-based technology for detecting Sharks in the ocean from Unmanned Aerial Vehicles (UAVs). SharkSpotter has attracted numerous industry awards, including recognition at the National iAwards for the 2018 AI or Machine Learning Innovation of the Year. Following the success of SharkSpotter, Michael's team developed world-first technology to protect North Queensland's waterways from predators via the launch of CrocSpotter in 2019. In 2020, Michael's Team conferred the Australian Association for Unmanned Systems (AAUS) Industry Champions Innovation Award for research on their AI-Spotter technology suite. In 2009 Michael was named as one of Australia's Top 10 Emerging Leaders in Innovation in Australia's Top 100 Emerging Leaders Series supported by Microsoft. Michael is a Fellow of the Australian Computer Society and a Senior Member of the IEEE.

.