Acta Univ. Sapientiae Informatica **14,** 2 (2022) 248–272

DOI: 10.2478/ausi-2022-0015

Rendering automatic bokeh recommendation engine for photos using deep learning algorithm

Rakesh KUMAR
Chandigarh University, Punjab, India
email: rakesh77kumar@gmail.com

Meenu GUPTA
Chandigarh University, Punjab, India
email: gupta.meenu5@gmail.com

JAISMEEN
Chandigarh University, Punjab, India
email: jaistiwana.231@gmail.com

Shreya DHANTA
Chandigarh University, Punjab, India
email: shreyadhanta1507@gmail.com

Nishant KUMAR PATHAK
Chandigarh University, Punjab, India
email: 19bcs3060@gmail.com

Yukti VIVEK
Chandigarh University, Punjab, India
email: yukti.vivek11@gmail.com

Ayush SHARMA
Chandigarh University, Punjab, India
email: ayushsharma1931@gmail.com

DEEPAK
Chandigarh University, Punjab, India
email: 19bcs1187@gmail.com

Gaurav RAMOLA
Samsung Research Institute, Banglore, India
email: gauravramola007@gmail.com

Sudha VELUSAMY
Samsung Research Institute, Banglore, India
email: Sudha.v@samsung.com

Computing Classification System 1998: G.2.2
Mathematics Subject Classification 2010: 68R15
Key words and phrases: bokeh, recommendation, photography, deep learning, inceptionV3, VGG16, mobileNetV2, effects

Abstract.
Automatic bokeh is one of the smartphone's essential photography effects. This effect enhances the quality of the image where the subject background gets out of focus by providing a soft (i.e., diverse) background. Most smartphones have a single rear camera that is lacking to provide which effects need to be applied to which kind of images. To do the same, smartphones depend on different software to generate the bokeh effect on images. Blur, Color-point, Zoom, Spin, Big Bokeh, Color Picker, Low-key, High-Key, and Silhouette are the popular bokeh effects. With this wide range of bokeh types available, it is difficult for the user to choose a suitable effect for their images. Deep Learning (DL) models (i.e., MobileNetV2, InceptionV3, and VGG16) are used in this work to recommend high-quality bokeh effects for images. Four thousand five hundred images are collected from online resources such as Google images, Unsplash, and Kaggle to examine the model performance. 85% accuracy has been achieved for recommending different bokeh effects using the proposed model MobileNetV2, which exceeds many of the existing models.

1 Introduction

'Bokeh' (i.e., Boke) is a word which is originated from the Japanese language (that translates as "Blur") and is popular in photography to provide visually appealing blurry images [6]. A user (s) must rely on different software to create beautiful bokeh effects on their images. With the significant number of bokeh types, it is difficult for users to choose the appropriate bokeh effect for their images [12]. The bokeh effect is usually used in photography to generate artistic images. Bokeh image focuses on the main object and the other regions that are out of focus (i.e., blurred). Most smartphone cameras cannot capture bokeh images due to small-sized apertures. Already existing methods [16] [18] [20] have worked on human portrait images by imposing depth estimation and image segmentation. In [11], Y. Liu and C. Yuan have proposed a constraint called background de-focused images for some particular class of pictures. This image shows an object portraying a blurred background; the main task was to separate the object's shape spontaneously. To find the segmentation of the undetermined image and for smoothing, they used the Graph cut (s-t cut) algorithm.

With the increase in social media usage, many images are shared and uploaded daily. Even though these social media platforms provide many filters, they still do not provide the professional quality images that the users require.

So, in [17], the authors have proposed category-aware aesthetic learning for filter recommendation. A bundle of filter images was generated by applying various types of filters. Further, the images were loaded into the system, which learns aesthetic images using pairwise comparison. Then various ranks were provided to the filters, and the top-ranked filter was recommended [1] [8]. A large aperture lens offers a good bokeh, making this effect unachievable for mobile cameras since they have tiny sensors and compact optics. In [7], the authors proposed Machine Learning (ML) based synthetic bokeh effect rendering algorithm. They applied a Convolutional Neural Network (CNN) model to segregate the people from the image, and then the remaining background of the image was blurred evenly. A predicted depth map was also used to blur the images, which was obtained with a stereo vision where multiple images were captured synchronously [7]. Aesthetic photos have become an important part of photography these days. But the ability of machines to convert normal images into aesthetic images is yet to be formulated. The Bokeh-effect rendering algorithm is used to overcome this issue. It was initially used for portraits where the background was blurred to make the person stand out. A depth estimation-based bokeh rendering method is used to satisfy the following characteristics: out-of-focus objects are blurred, and the focal plane provides a clear view of the object away from the focal plane (more blurred object). They have adjusted the degree of bokeh with the help of a blurred radius factor [19] [4]. The main objective of this work (i.e., bokeh recommendation) is to achieve the highest-ranked bokeh that is suitable for visual image content and the user's notice. The proposed model generates a bokeh effect on the random images based on different parameters such as color pop, spin, zoom, low key, high key, big bokeh, silhouette, and blur.

This paper is further classified into different sections. Section 2 provides a detailed description of the different researcher's views on bokeh recommendation with different techniques applied. This section also consists of the different examples of bokeh recommendation engines of existing systems. Next, section 3 discusses the methodology used for this proposed work, followed by a collection of datasets, a proposed model, and mathematical formulations. Section 4 discusses the experimental result analysis of testing and training with confusion matrices. Finally, this work is concluded in section 5, along with its future work.

2 Background study

This section discusses the automatic detection of recommendation systems proposed by different researchers. The important information about MobileNetV2 was extracted and used in a way that could be helpful for the proposed model.

In [2], S. Dutta proposed the blending of smoothed pictures with the help of a depth-aware technique for generating the Bokeh effect. The AIM 2019 Challenge on Bokeh Effect synthesis presented an approach contrasted to a saliency detection-based baseline, and many methods were offered. Single Lens Reflex cameras with a small depth of focus were commonly used to capture bokeh images. They employed a lightweight network with the depth-of-field or Bokeh effect to create visually appealing images. The AIM 2019 Bokeh effect challenge-perpetual Track ranked this technique second. In [9], R. Jain et al. proposed CNN models to classify Covid-19 afflicted patients with the assistance of chest X-ray scans. They used Deep Learning (DL), a very effective technique of Machine Learning (ML), to study massive amounts of X-rays. They compared the performance of InceptionV3, Xception Net, and ResNeXt in the model. As a result, Xception Net had the highest accuracy for recognizing Chest X-ray images compared to other models. In [3], V. Gajarla and A. Gupta proposed a model for emotion detection and sentiment analysis. They looked at the probability of utilizing DL to identify the emotion portrayed by an image. They used several classification algorithms on the data, including SVM on some good features of VGG-ImageNet, and made small adjustments to the already trained models, including RESNET, Places205-VGG16, and VGG-ImageNet. As a result, the authors concluded that neural networks could learn the emotion elicited by a picture. These types of prediction can be used in areas such as automatic tag predictions for photographs sent to media websites for determining public attitude and mood.

In [13], B. Ramzan et al. proposed a CF recommendation approach that could manage diversified data like lexical evaluations, votes, rankings, and views on videos. They used polarity identification and opinion-based sentiment analysis to create a hotel feature matrix. They used a combination of lexical, syntactic, and semantic analysis to acknowledge the point of view toward hotel characteristics and guest type profiling (solo, couple, family, etc.). The system introduced a recommendation of hotels based on the hotels' features and types of guests for a customized recommendation. In [11], Y. Liu and C. Yuan have suggested a method based on blur measuring, which is dependent on an edge width (precise) computation algorithm. They also used the Graph cut (s-t cut) algorithm to detect segmentation for a flat and undecided area. They experimented with a huge blur in the foreground, where the captured object had high accuracy. The authors concluded that the proposed model is used for picture editing, segmentation, automatic annotation, and recognition systems. In [10], H. H. Li et al. proposed a machine learning model to recommend personalized skin care products. All the concepts they used to focus on machine learning and deep learning algorithms in developing a recommendation platform for human face and skin intelligence. YOLOv4 was primarily utilized for face skin identification. The result received from the proposed model is sent to a recommendation platform for prediction, which examines the best skin care product for the consumer. They concluded that customers will have a good grasp of their skin's condition and can select more appropriate products to avoid skin damage and harm originating from some incorrect products.

In [17], W. T. Sun et al. proposed an approach for filter suggestion that learns aesthetic representations. Filter Aesthetic Comparison Dataset (FACD) used by the authors consisting of twenty-eight thousand one hundred sixty filtered pictures dependent on the AVA dataset and forty-two thousand two hundred and forty trustworthy image pairings with crowd-sourcing labels for pairwise filter comparison. This dataset included over 40K picture pairs filtered by the users' preferences. The results showed that the suggested category-aware aesthetic learning outperforms old classification methods. In [15], F. Saxen et al. proposed an idea for Attribute Detection of the face with the help of MobileNetV2 and NasNet-Mobile model. They addressed an issue of facial attribute approximation for mobile devices from RGB pictures. Face attribute approximation is prominent for human-machine interaction (HCI) systems because it provides useful context information such as the interacting user's age, sex, and ethnicity. Their comparison of processing time and accuracy revealed that it is quicker and better than the state-of-the-art model. Their evaluation showed that resource-limited applications could outperform current top-performing applications. They evaluated a methodology that can be executed on mobile devices using two lightweight architectures of CNN. In [6], A. Ignatov et al. proposed a technique for the practical rendering task of the bokeh effect. They recommended learning the bokeh effect straight from real photographs taken with a DSLR camera to imitate realistic camera bokeh. They gathered a large dataset of ranged broad and facile depth-of-field image pairs taken with a Canon 7D camera and a 50mm f/1.8 fast lens. The authors trained the suggested PyNET-based result on the data and found that it outperformed existing deep learning systems in terms of quality. As a result, they concluded that deep learning techniques do not require a camera or any specific hardware device for implementation.

In [5], R. Guha proposed DL neural nets for collaborative filtering in recommender system design. They used two recommender systems: Content-Based Filtering (CBF) and Collaborative Filtering (CF). It indicates that deep learning has the potential to improve recommender system performance. In [19], F. Wang et al. proposed a Bokeh effect rendering in which three modules: estimation of depth, a subdivision of background, and the rendering of bokeh. The depth estimation of the image is needed to render a normal image with the bokeh effect. They also presented a method for blurring image background subregions. By selecting multiple focal planes, their methods produced a variety of bokeh photos. Since picture-depth estimation and bokeh-effect rendering were independent, the bokeh image would be more in line with the natural bokeh effect as the accuracy of the image-depth-estimation model becomes better and better. CNN and GAN models performed well on the images. In [14], M. Sandler et al. proposed a network architecture that was easy to understand and permitted users to prepare a group of extremely structured phone models. ImageNet dataset consists of an architecture that improves the new methods for a diverse range of performance points. On the COCO dataset, the network surpasses the new method of real-time detectors in terms of accuracy and model complexity for object detection. The architecture combined with the SSD Lite detection module is 20× less computation and 10× fewer parameters than YOLOv2.

Based on the studies mentioned above, it has been concluded that the authors have worked on various techniques and algorithms. They have focused on InceptionV3, VGG 16, Xception net and MobileNetV2, etc. The MobileNetV2 model has not been used much for different Bokeh effects. So, in this work MobileNetV2 model is proposed for various Bokeh effects.

3 Proposed methodology

This section discusses the dataset description, data annotation, image categorization, proposed methodology, and data preprocessing. Further, it also covers the model formulation along with the proposed algorithm.

3.1 Dataset description

The dataset used in the proposed model has been consists of 4500 portrait images derived from three different sources, i.e., google images (2500 images), Unsplash (1200 images), and Kaggle (800 images). The collected dataset is stored in one drive, available at (https://drive.google.com/drive/u/1/folders/1qKJ5Q7hokIGITb7O1RdDnxsx-eK85bjq). Various parameters are used for categories such as Blur, Color-point, Zoom, Spin, Big Bokeh, Color Picker, Low-key, High-Key, and Silhouette (shown in figure 1). After this, Images are resized into 224 x 224 for perfect filter recommendation of the images and divided into training, testing, and validation, 90%, 5%, and 5%, respectively.

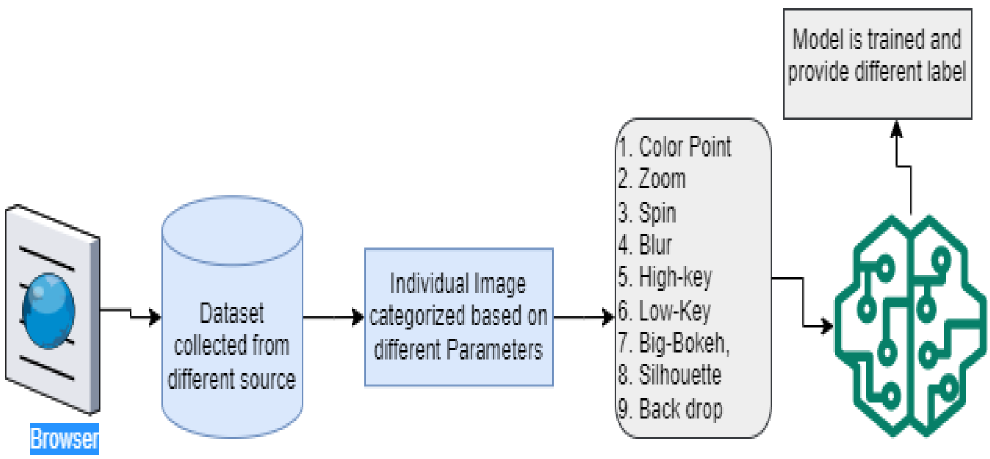


Figure 1: Preliminary processing of collected images categorized by different parameters

Table 1 shows the Parameters of Zoom, Spin, Big Bokeh, Color Point, and Low-Key Categories. Table 2 shows the Parameters of High Key, Blur, Silhouette, and Back Drop Categories. Here, NA refers to Not Applicable, i.e., doesn't matter, YES describes Preferred parameters, NO refers to the characteristics that should be absent, and [IMP] Yes describes the essential parameters. Artificial Light preferred explains that there should be light from artificial resources only, and it can not be from natural resources like sunlight. DUSK represents a time of the day which is closest to sunset time.

3.2 Data annotation and image categorization

Augmentation is applied to the images and trained model on various parameters (Zoom, Color pop, Blur, Spin, Big Bokeh, Backdrop, High key, Low Key, and Silhouette) to increase the model's accuracy. MobileNetV2, InceptionV3,

S.No.	Parameters	Zoom	Spin	Big Bokeh	Color Point	Low Key
1.	Static Background	YES	YES	NA	NA	NA
2.	Perceivable depth	YES	YES	NA	NA	NA
3.	Symmetry	YES	YES	NA	NA	NA
4.	Clutter	NO	NO	NO	NA	NO
5.	Bland Background	NO	NO	YES	NO	YES
6.	Light Source in the background	NO	NO	[IMP] YES	NA	NO
7.	Light sources sparse	NA	NA	NO	NA	NA
8.	Foreground Obstructed	NA	NA	NO	NO	NO
9.	Light source dominating/ obscuring foreground	NA	NA	NO	NA	NO
10.	Time of day	NA	NA	NIGHT	DAY	ALP
11.	Vanishing point within foreground pixel	[IMP] YES	NA	NA	NA	NA
12.	Normal vanishing points	[IMP] YES	NA	NA	NA	NA
13.	Small disturbances causing objects in the background	YES [IMP]	YES	NA	NA	NA
14.	Asymmetry due to few large objects	NO	NO	NA	NA	NA
15.	Unnatural pose	NO	NO	NA [IMP] YES	NA	NA
16.	Centered Subject	NA	YES	NA	NA	YES
17.	Uniform textured/ patterned background	NA	YES	NA	NA	NA
18.	Quadrant Homogeneity	NA	YES	NA	NA	NA
19.	The mood of the subject - Intense	NA	YES	NA (if low saturation) YES	YES	NA
20.	The mood of the subject - Happy	NA	NA	NA	(if low saturation YES)	NO
21.	The mood of the subject - Neutral	NA	NA	NA	(if low saturation YES)	YES

Table 1: Parameters of Zoom, Spin, Big Bokeh, Color Point, and Low-Key categories. *Note: Artificial Light Preferred:ALP.*

Sr. No	Parameters	High Key	Blur	Silhouette	Backdrop
1.	Static Background	NA	YES	NA	[IMP]YES
2.	Perceivable depth	NA	NA	NA	NA
3.	Symmetry	NA	NA	NA	NA
4.	Clutter	NO	[IMP]YES	NO	NO
5.	Bland Background	YES	NO	NO	YES
6.	Light Source in the background	NO	NA	[IMP]YES	NO
7.	Light sources sparse	NA	NA	NA	NA
8.	Foreground Obstructed	NO	NA	NO	NO
9.	Light source dominating/ obscuring foreground	NO	NA	NO	NA
10.	Time of day	ALP	NA	DUSK	NA
11.	Vanishing point within foreground pixel	NA	NA	NA	NO
12.	Normal vanishing points	NA	NA	NA	NA
13.	Small disturbances causing objects in the background	NA	YES	NA	NO
14.	Asymmetry due to few large objects	NA	NA	NA	NA
15.	Unnatural pose	NA	NA	NA	NA
16.	Centered Subject	NO	NA	NA	NA
17.	Uniform textured/ patterned background	NA	NA	NA	NA
18.	Quadrant Homogeneity	NA	NA	NA	NA
19.	The mood of the subject - Intense	NO	NA	NA	NA
20.	The mood of the subject - Happy	YES	NA	NA	NA
21.	The mood of the subject - Neutral	NO	NA	NA	NA

Table 2: Parameters of Zoom, Spin, Big Bokeh, Color Point, and Low-Key categories

and VGG16 models are applied to the collected dataset to achieve good accuracy. For a better understanding, a detailed description of all the parameters has been discussed below:

- **Zoom**:The model considers a symmetric image with perceivable depth and a static background. The model can categorize images with converged backgrounds and small obstructions.
- **Spin**: The model seems to be learning that a centered subject image with an intense mood and static background is to be sorted in this filter.
- **Big Bokeh**: The model seems to take out-of-focus and blurred images showing the lenses' aperture shape. Images with bright, colorful circles are often seen in this category.
- **Color Point**: This category has a vibrant collection of images. The model seems to have no difficulty categorizing this filter and can classify happy, sad, and intense faces in the hours of sunlight.
- **High Key**: The model can sort high key images that are entirely very bright with no dark shadows and an overexposed background. The model creates a light, pleasant mood as opposed to a shadow image which often creates a scary mood.
- **Low Key**: This filter is quite similar to high-key, and differentiating factor is that the subject's mood should be neutral and intense, and the subject should be centered on a bland background.
- **Blur**: The filter is usually used for images that have a high amount of clutter in the background. Some other important characteristics include static background, and there should be minor disturbances that cause objects in the background.
- **Silhouette**: The model arranges a solid, dark image of a subject against a brighter background in this category. Images with dark outlines of subjects in front of contrast or fully bright scenes such as studio lights or sunsets will be grouped in this filter by the model.
- **Back Drop**: This category has simple and weak patterns or solid colors so that eyes are not drawn off the subject. This type of filter shows a dark background with artificial lights.

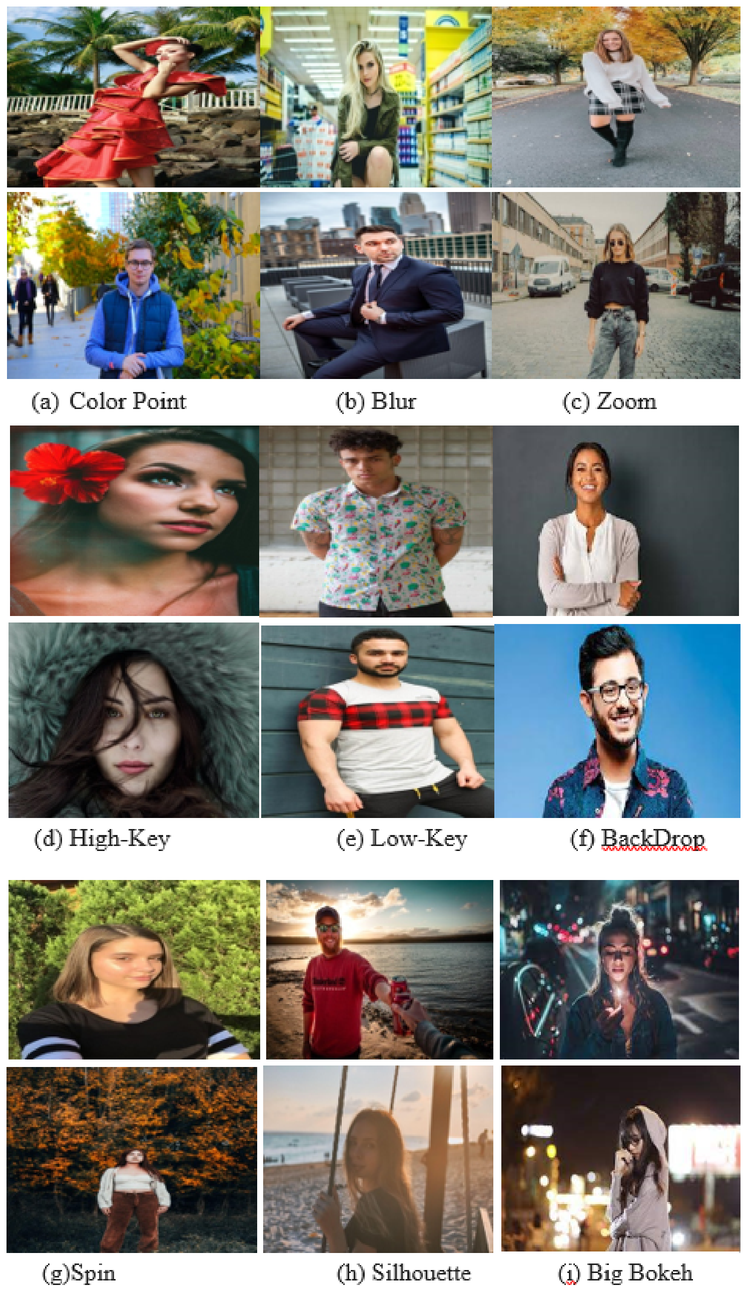


Figure 2: Example of dataset images collected for each category

Figure 2 shows the sample of collected images for each category, including Color point, Blur, Zoom, High key, Low Key, Backdrop, Spin, Silhouette, and Big Bokeh.

3.3 Proposed model formulation

The proposed model (MobileNetV2) is formulated and processed with the help of parameters such as static background/foreground, symmetry, and many more. Further, a softmax activation function is used with its various learning rates. A primitive type detector engine runs on the collected datasets, and the ground truth labels are received, i.e., called annotated datasets. The data is partitioned into training, validation, and testing with 90%, 5%, and 5%, respectively. The MobileNetV2 is applied to the collected dataset as it is an efficient feature extractor for object detection and segmentation. Data augmentation like rotation, zoom and Horizontal Flip are used to remove the overfitting of the model. Further, InceptionV3 and VGG16 are applied to the same dataset and compared with the proposed model in terms of attained accuracy. Figure 3 shows the Model formulation for Automatic filter recommendation using MobileNetV2.

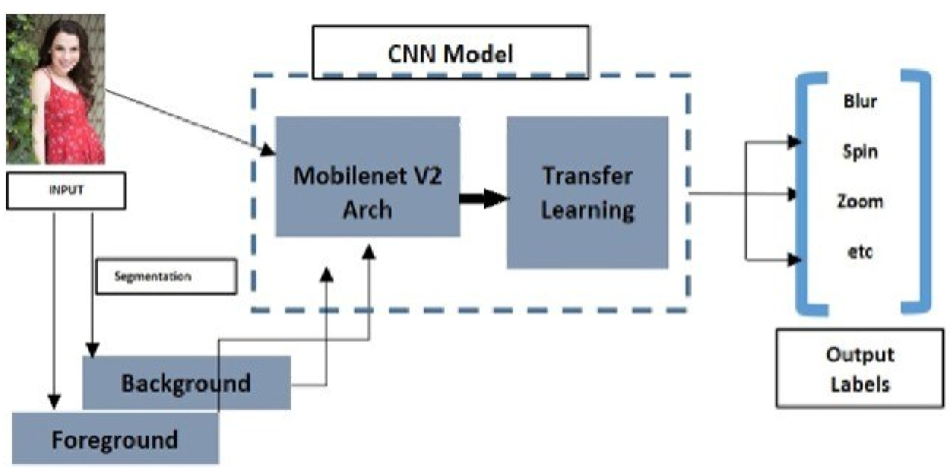


Figure 3: Model formulation for Automatic filter recommendation using MobileNetV2

3.4 Proposed algorithm

MobileNetV2 is a classification model which provides real-time classification capabilities under computing constraints in different devices like smartphones. Therefore, in this section, step by step model algorithm is explained.
Step 1: Initialize the size of Images to width, height, channel, 224, 224, and 3, respectively.
Step 2: Apply Data Augmentation
Custom Image Data Generator is utilized for further processing:
(a) Rescale=1/255
(b) Zoom Range=0.2
(c) Horizontal Flip=True
(d) Rotation Range=10
This phase significantly helps to increase the diversity of data.
Step 3: Call the MobileNetV2 model and pass the preprocessed images as input for feature extraction.
Step 4: Load the sequential model because we have exactly one input and one output tensor.
Step 5: Load the conventional base layer with the ImageNet weight, max pooling, including the top value set to "False," and input shape set to the image's shape, i.e. (224,224,3).
Step 6: Add a dense layer for interface with Softmax activation.

The output vector element is in the range (0,1) and sums to 1. The softmax of each vector x is computed using

$$\exp(x)/\mathsf{tf.reduce\_sum}(\exp(x)) \tag{1}$$

where x is the input tensor, this returns the tensor, an output of softmax transformations (non-negatives).
Step 7: Compile the model using the Adam optimizer with a learning rate of 1e-4 and epochs count of 52 and set the clipnorm value to 1.

$$w_t = (w_(t-1)) - \alpha \times (\vec{m}_t/(\sqrt{\vec{v_t}} + e)) \tag{2}$$

i) Where w_t =weight at time t
ii) W_{t-1} = weight at time $t-1$
iii) α = learning rate
iv) m_t = aggregate of the gradient at time t
v) v_t = variance at time t
vi) e = constant

Step 8: Apply categorical cross-entropy loss function LCE

$$L_{C_E} = -\sum_{i=1}^{n} t_i \log{(p_i)} \text{for } n \text{ classes} \tag{3}$$

where (t_i) is the truth label and (p_i) is the softmax probability for the (i^{th}) class.
Step 9: Train the model and save the results for further processing and calculations.
Step 10: Calculate the model's accuracy to check our proposed model's efficiency.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{4}$$

4 Experiment results and discussion

In the result formulation, the dataset images are filtered based on the different parameters and passed to the model for further analysis. InceptionV3, VGG 16, and MobileNetV2 are models applied to the annotated dataset to find the best accuracy.

4.1 Filter recommendation using MobileNetV2

MobileNetV2 (a CNN-based model) is applied in the annotated dataset as it is used for mobile devices. It is set up on an inverted residual formation, with residual connections between bottleneck levels. MobileNetV2 is a wide network of 53 layers, and it consists of a fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers. Table 3 depicts the precision, recall, f1-score, and support for all nine filters using the MobileNetV2 Model.

Table 4 shows the various accuracy levels based on the different learning rates. Mainly, three learning rates are used, i.e., 1e-3, 1e-4, and 1e-5. MobileNetV2 is trained on these learning rates to achieve accuracy.

Table 5 shows the accuracy of the MobileNetV2 model using the SoftMax activation function with 52 epochs. The Best accuracy achieved from the proposed model (MobileNetV2) is 85 % on learning rate 1e-4.

Figure 4 (a) depicts the training and validation loss of the MobileNetV2 model against the epoch on the x-axis and loss on the y-axis, as it lowers with subsequent epochs. Figure 4 (b) shows the accuracy for MobileNetV2 against the epoch on the x-axis and accuracy on the y-axis, as it enhances with the succeeding epochs.

	precision	recall	f1-score	support
backdrop	0.85	1.00	0.92	51
big bokeh	0.70	0.88	0.78	51
blur	1.00	0.62	0.77	50
color pop	0.77	0.83	0.80	52
high key	0.65	0.82	0.73	50
lowkey	0.84	0.64	0.73	50
silhouette	0.94	0.92	0.93	53
spin	0.77	0.71	0.73	51
zoom	0.90	0.85	0.87	52

Table 3: Testing data for MobileNetV2

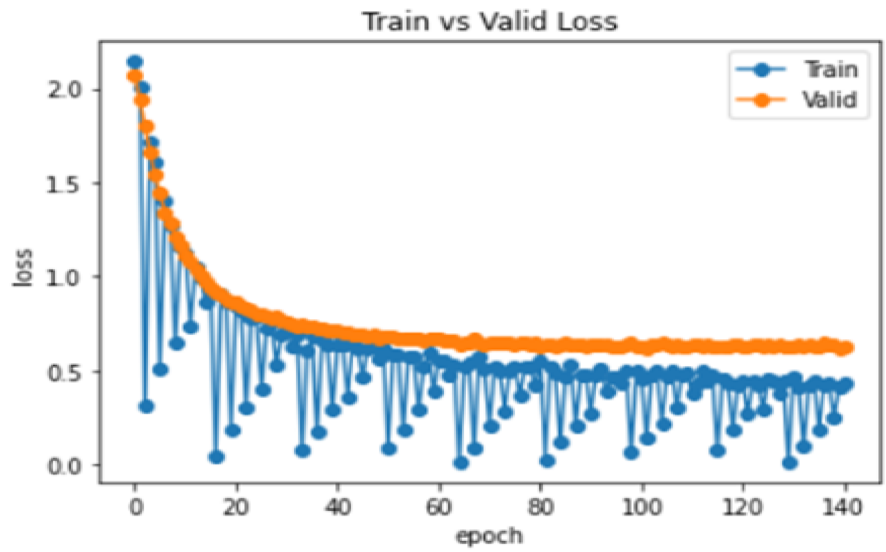
Learning Rate	Accuracy
1e-3(Without Dropout)	0.8423611521720886
1e-4(Without Dropout)	0.8479166626930237
1e-5(Without Dropout)	0.8034722208976746
1e-3(With Dropout)	0.8187500238418579
1e-4(With Dropout)	0.8513888716697693
1e-5(With Dropout)	0.7451388835906982

Table 4: Accuracy based on different learning rates of MobileNetV2

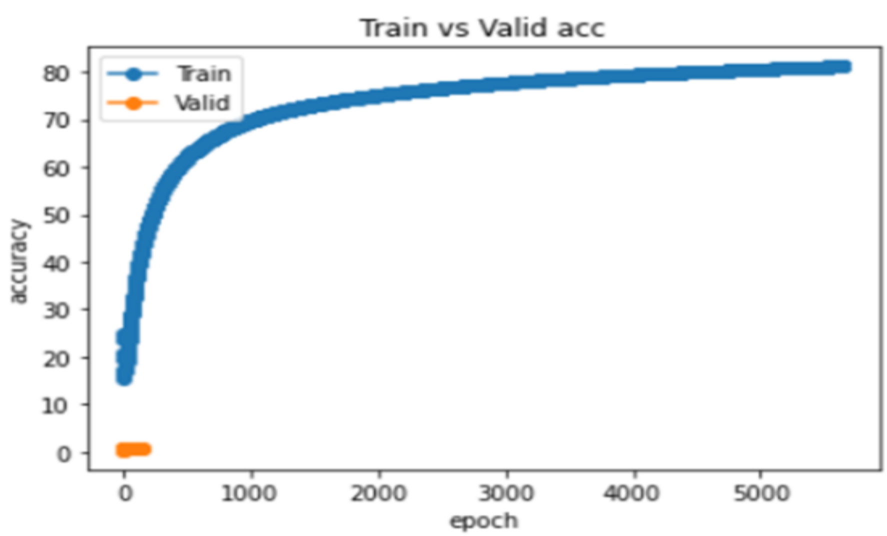
Figure 5 indicates the confusion matrix of MobileNetV2 for training and testing data, where the X-axis displays the predicted label, and the Y-axis shows the true label. The best accuracy of this model is 85%, with approximately 52 epochs and activation used as SoftMax.

Model	Activation	Epoch	Accuracy
MobileNetV2	SoftMax	52	0.8513888716697693

Table 5: Accuracy of proposed Model (MobileNetV2)



(a)



(b)

Figure 4: (a) Training and Validation loss of MobileNetV2 (b)Validation and Training Accuracy of MobileNetV2

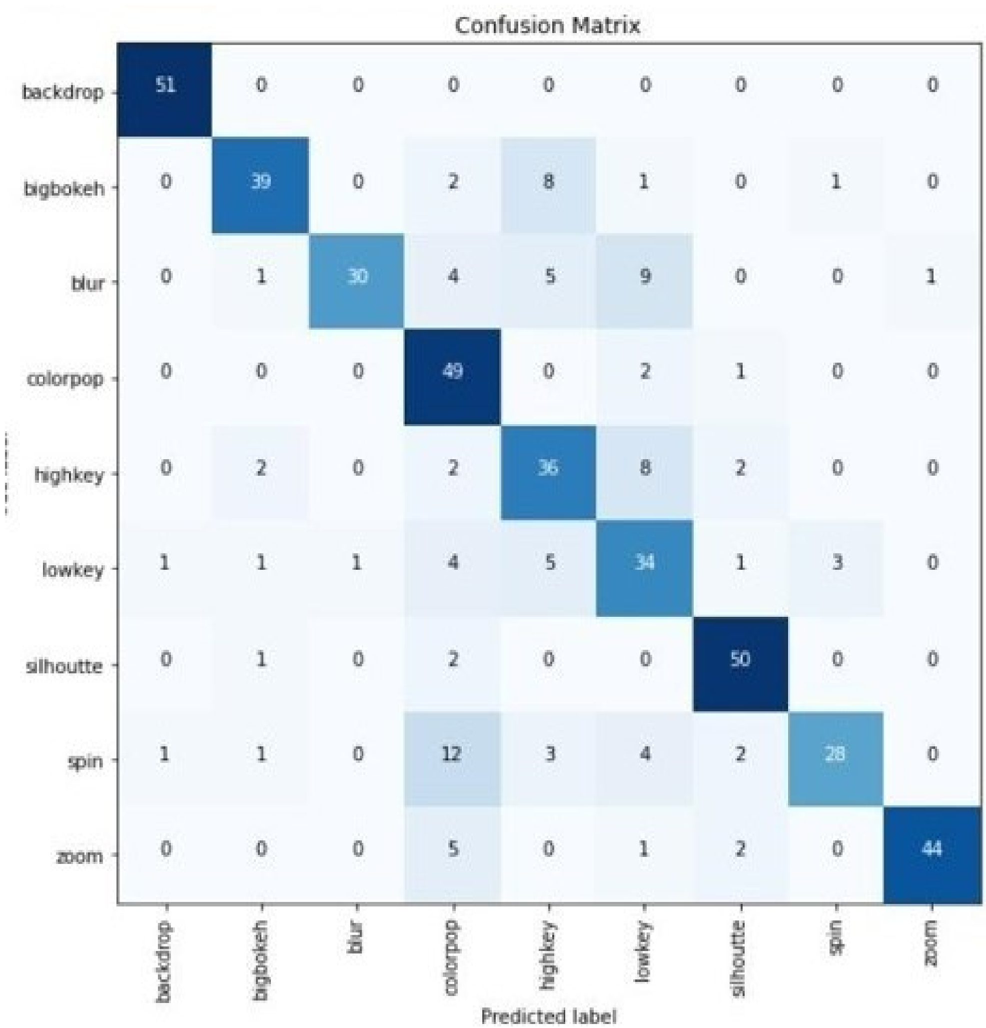
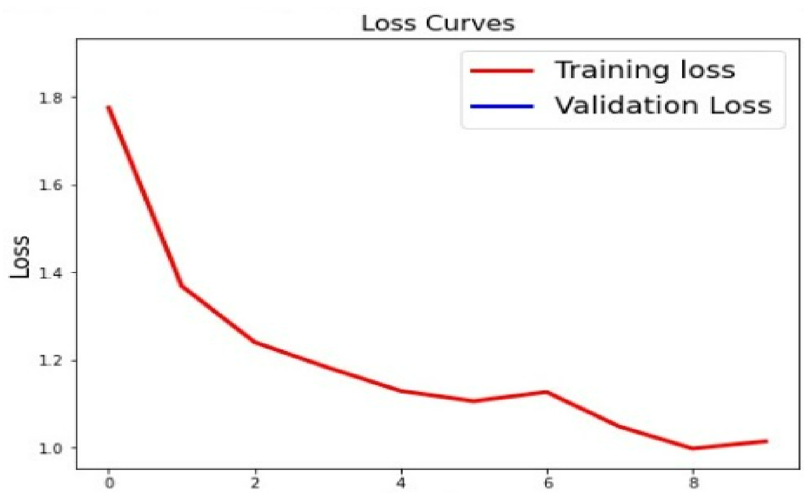
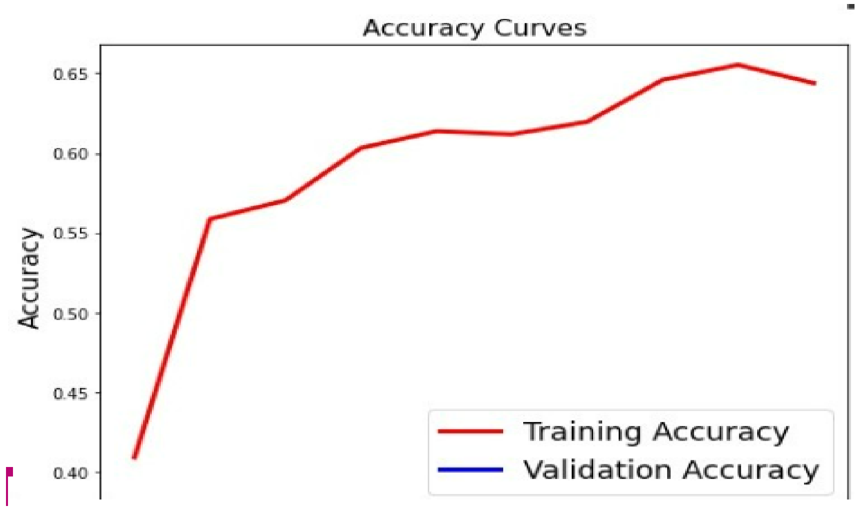


Figure 5: Confusion matrix of MobileNetV2 model for training data



(a)



(b)

Figure 6: (a) Training and Validation loss of InceptionV3 with subsequent epochs (b)Training and Validation Accuracy of InceptionV3 with subsequent epochs

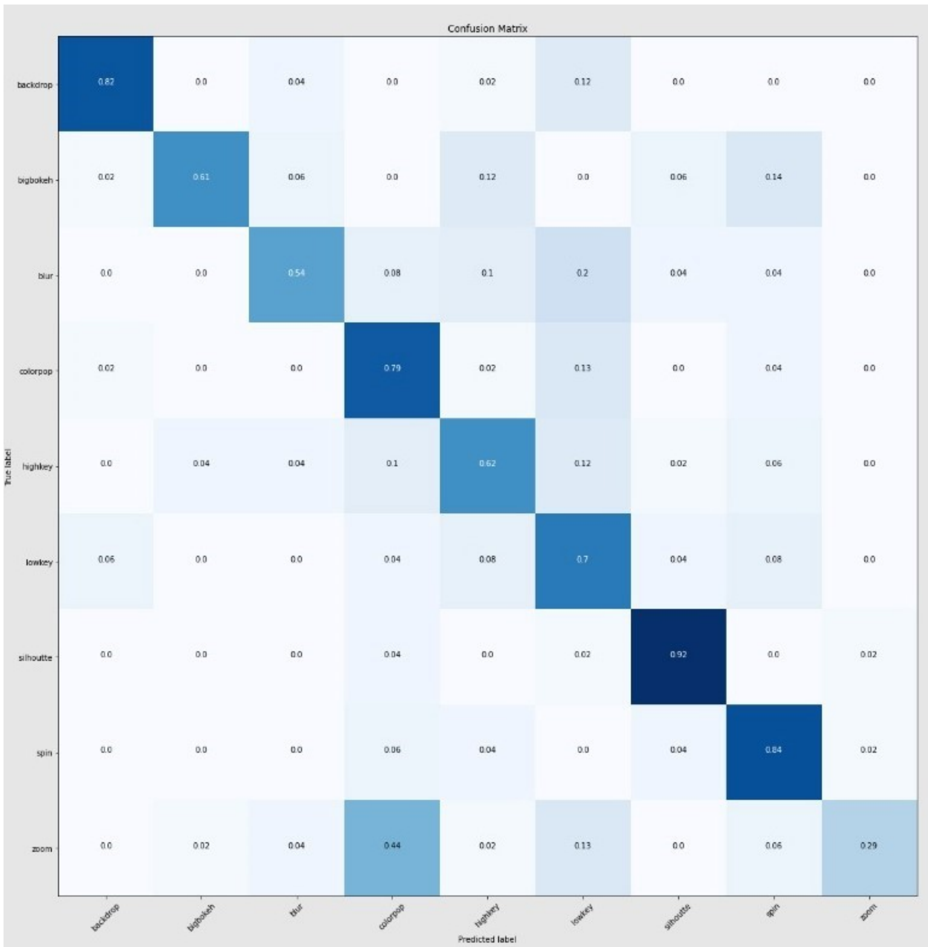
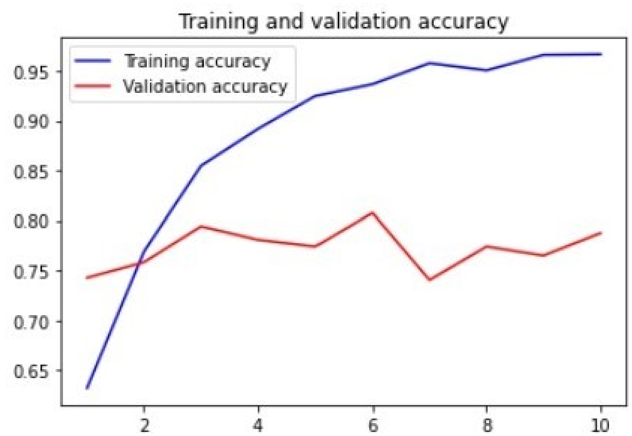


Figure 7: Confusion matrix of InceptionV3 model for training data

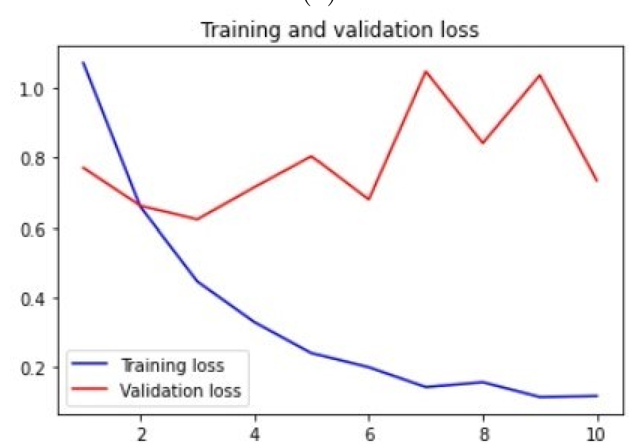
4.2 Filter recommendation using InceptionV3

The InceptionV3 model is used for categorization/grouping purposes. It has 48 layers and concatenated layers 1 x 1, 3 x 3, and 5 x 5 convolutions. By using this, training speed can be increased by reducing several variables. This model is suitable for training purposes as it gives good accuracy. Figure 6 (a) depicts the training loss for the InceptionV3 model where the x-axis indicates the epochs and the y-axis shows the Loss, as it lessens with the succeeding epochs. Figure 6 (b) depicts the model's accuracy for the InceptionV3 model, where the x-axis indicates the epochs and the y-axis indicates the accuracy as it progresses with the succeeding epochs.

Figure 7 depicts the confusion matrix of the InceptionV3 model for training and testing data. The accuracy of InceptionV3 is 80%.



(a)



(b)

Figure 8: (a) Training and Validation Accuracy of VGG16 with subsequent epochs (b)Training and Validation Loss of VGG16 with subsequent epochs

Model	Accuracy
MobileNetV2	85%
InceptionV3	80%
VGG 16	81%

Table 6: Comparison of different models

4.3 Filter recommendation using VGG16

The VGG16 comprises 16 layers and can assort photos into 1000 objects, like keyboards, animals, pencils, mice, etc. The model also supports an image input size of 224 X 224 pixels.

Figure 8 (a) depicts the training and validation accuracy for the VGG16 model. The x-axis indicates the epochs, and the y-axis indicates the accuracy, as it enhances with successive epochs. Figure 8 (b)depicts the training and validation loss for the VGG16 model, where the x-axis indicates the epochs and the y-axis indicates the loss, as it lessens with the succeeding epochs.

Figure 9 depicts the confusion matrices procure the test set where the x-axis indicates the predicted labels and the y-axis indicates the True labels for the VGG16 model. The accuracy of VGG 16 models is 81%.

4.4 Comparison of MobileNetV2, VGG16, and IncentionV3

The basic difference between MobileNetV2 and other models is that MobileNetV2 uses a depth-wise separable convolution, whereas InceptionV3 uses standard convolution and VGG uses a small reception field in the convolutional model. The fewer parameters in the MobileNetV2 model make it lighter and easier to use. MobileNetV2 is usually faster for the same accuracy across the entire latency spectrum. Table 6 shows the comparison of MobileNetV2, InceptionV3, and VGG16 models. As shown in table 6, the best accuracy is 85% for the MobileNetV2 model compared to the other models.

5 Conclusion and future scope

In this research work, we presented an approach for a bokeh-type recommendation engine based on nine filters. The dataset was collected through various online sites and categorized into nine filters. The images are annotated before training the model. The model is trained on multiple parameters, and three

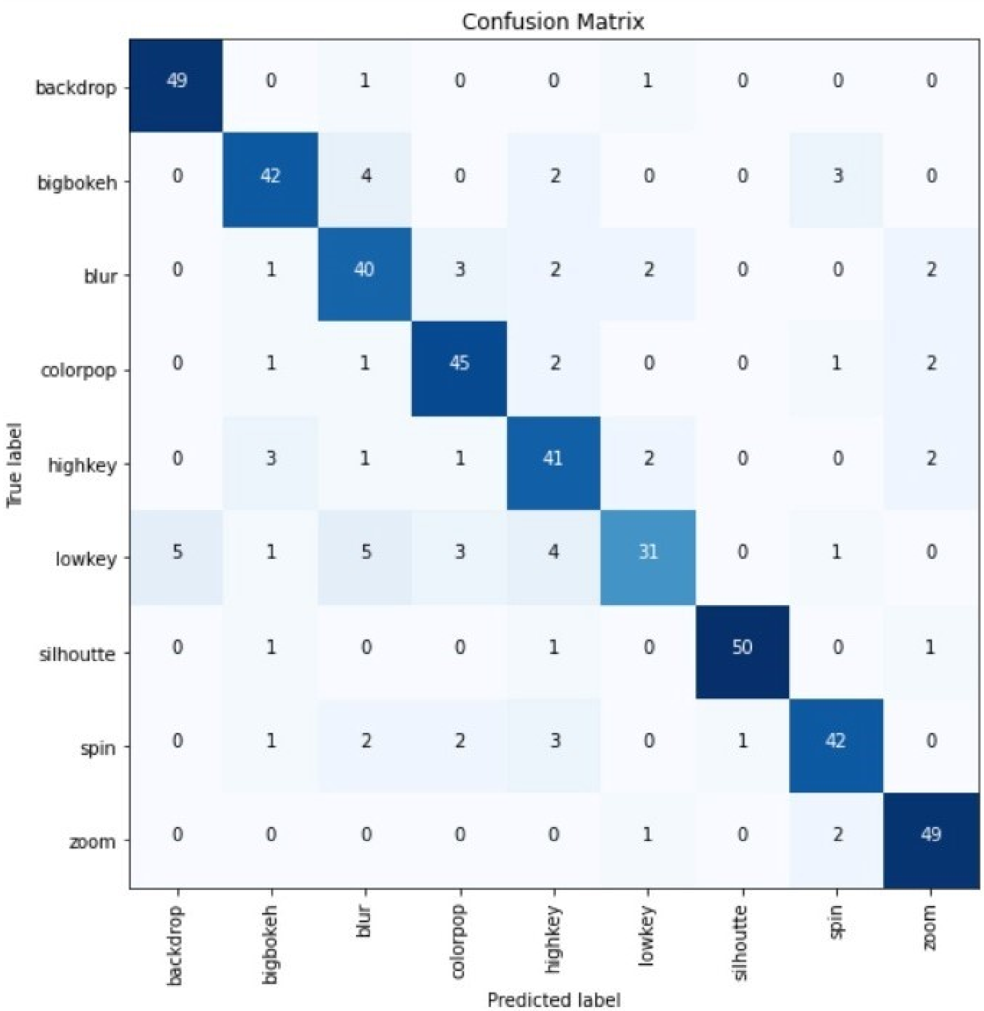


Figure 9: Confusion matrix of train data of VGG16 model

different types of models are used to get the best possible result. The best accuracy, i.e., 85%, is achieved with the MobileNetV2 (a proposed) model. The final analysis of the work concludes that a recommendation Bokeh type system will provide the best choice of filter for the provided image. The results show deep learning gives good results for the classification of emotions as well as in performing sentiment analysis. The results are better than among other existing models suitable for image visualization based on users' interests. In the future, these models can be applied on a larger scale to solve the real-life problem of Bokeh's

Acknowledgements

We are thankful to the Samsung Research Institute, Banglore, India, for supporting us in providing the dataset for model formulation.

Authors' contribution

Meenu Gupta and Rakesh Kumar performed writing- original draft and conceptualization. Jaismeen, Shreya Dhanta, Nishant Kumar Pathak, Yukti Vivek, Ayush Sharma, Deepak performed data cleaning and execution. Gaurav Ramola, Sudha Velusamy verified and authenticated the primary dataset. Meenu Gupta proofread, revised, restructured, and improved the whole manuscript substantially.

Funding statement

The author(s) received no funding for this study

Conflicts of interest

The authors declare they have no conflicts of interest to report regarding the present study.

References

[1] S. Bakhshi, D. Shamma, L. Kennedy, E. Gilbert, Why we filter our photos and how it impacts engagement, *Proc. 9th International AAAI Conference on Web and social media*, 2015, pp. 12–21. ⇒250

[2] S. Dutta, Depth-aware blending of smoothed images for bokeh effect generation. *J. Visual Comm. Image Represent.* **77** (2021) 103089. ⇒251

[3] V. Gajarla, A. Gupta, Emotion detection and sentiment analysis of images. *Georgia Institute of Technology*, **1** (2015) 1–4. ⇒251

[4] V. D. Gesu, M. Maccarone, An approach to random images analysis. *Proc. of the Springer II Conference In Image Analysis and Processing*, 1988, pp 111–118. ⇒250

[5] R. Guha. Improving the performance of an artificial intelligence recommendation engine with deep learning neural nets, *Proc. 6th EEE International Conference for Convergence in Technology (I2CT)*, 2001, pp 1–7. ⇒253

[6] A. Ignatov, J. Patel, and R. Timofte. Rendering natural camera bokeh effect with deep learning. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp 418–419. ⇒249, 252

[7] A. Ignatov, J. Patel, R. Timofte, B. Zheng, X. Ye, L. Huang, X. Tian, S. Dutta, K. Purohit, P. Kandula, et al., Aim 2019 challenge on bokeh effect synthesis: Methods and results. *Proc. of IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp 3591–3598. ⇒250

[8] P. Isola, J. Xiao, A. Torralba, A. Oliva, What makes an image memorable? , *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp 145–152. ⇒250

[9] R. Jain, M. Gupta, S. Taneja, D. J. Hemanth, Deep learning based detection and analysis of covid-19 on chest x-ray images. *Applied Intelligence*, **51,** 3 (2021) 1690–1700. ⇒251

[10] H.-H. Li, Y.-H. Liao, Y.-N. Huang, P.-J. Cheng, Based on machine learning for personalized skin care products recommendation engine. *Proc. of the IEEE International Symposium on Computer, Consumer and Control (IS3C)*, 2020, pp 460–462. ⇒252

[11] Y. Liu, C. Yuan, Automatic segmentation of background defocused nature image, *Proc. of the IEEE 2nd International Congress on Image and Signal Processing*, 2009, pp 1–5. ⇒249, 251

[12] W. Niu, J. Caverlee, H. Lu, Neural personalized ranking for image recommendation, *Proc. of the 11th ACM international conference on web search and data mining*, 2018, pp 423–431. ⇒249

[13] B. Ramzan, I. S. Bajwa, N. Jamil, R. U. Amin, S. Ramzan, F. Mirza, and N. Sarwar. An intelligent data analysis for recommendation systems using machine learning. *Scientific Programming*, **2019** (2019). ⇒251

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks. *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018, pp 4510–4520. ⇒253

[15] F. Saxen, P. Werner, S. Handrich, E. Othman, L. Dinges, A. AlHamadi, Face attribute detection with mobilenetv2 and nasnet-mobile, *Proc. of 11th the IEEE International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2019, pp 176–180. ⇒252

[16] X. Shen, A. Hertzmann, J. Jia, S. Paris, B. Price, E. Shechtman, I. Sachs, Automatic portrait segmentation for image stylization. *In Computer Graphics Forum*, **35,** 2 (2016) 93–102. ⇒249

[17] W.-T. Sun, T.-H. Chao, Y.-H. Kuo, W. H. Hsu, Photo filter recommendation by category-aware aesthetic learning. *IEEE Transactions on Multimedia*, **19,** 8 (2017) 1870–1880. ⇒250, 252

[18] N. Wadhwa, R. Garg, D. E. Jacobs, B. E. Feldman, N. Kanazawa, R. Carroll, Y. Movshovitz-Attias, J. T. Barron, Y. Pritch, M. Levoy, Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics (ToG)*, **37,** 4 (2018) 1–13. ⇒249

[19] F. Wang, Y. Zhang, Y. Ai, W. Zhang, Rendering natural bokeh effects based on depth estimation to improve the aesthetic ability of machine vision. *Machines*, **10,** 5 (2022) 286. ⇒250, 253

[20] X. Xu, D. Sun, S. Liu, W. Ren, Y.-J. Zhang, M.-H. Yang, J. Sun, Rendering portraitures from monocular camera and beyond, *Proc. of the European Conference on Computer Vision (ECCV)*, 2018, pp 35–50. ⇒249

Received: November 9, 2022 • Revised: December 17, 2022