Vol-07 Issue 01, January-2023

ISSN: 2456-9348
Impact Factor: 5.004

IJETRM
International Journal of Engineering Technology Research & Management

BOKEH RENDERING USING A DATA DRIVEN MODEL

Ashish Chopra, Anshul Subramanian, Rajiv Murali, Sahil Tuli, Shabi Ul Hussan
Samsung Research Institute Noida

Abstract –
A Bokeh effect enhances the focal areas in a photograph by blurring out the background's uninteresting features while focusing on the foreground of the image. This method can be improved to create images that appear more realistic and organic. We propose a data driven solution using Local Interpretable Model-Agnostic Explanation unit which works along with an object classification model for better classification results. We also use a GAN based model to create a depth map of the image in order to analyze the number of layers in the image to provide a variable blur.

I.INTRODUCTION

The Bokeh Effect is a type of background blur caused by the design of camera lenses, the size of the aperture, the separation between the foreground and the background, and the various light and shadow patterns that emerge from these aspects. To produce this image, lenses with a shallow depth of field and a bigger size are frequently employed. As a result, replicating this look with a smartphone camera is quite challenging.

The proposed architecture uses a GAN based network to generate a depth map of the image [1]. Using this depth map the number of layers are analyzed and based on this, a variable blue can be produced. We use YOLOv3 [2] for object classification along with local Interpretable Model-Agnostic explanation model [3] while training in order to improve the quality of the prediction. Based on the object detected in the foreground of the image, we enhance its colors using the color spectrum enhancing module [4]. Finally, a weight is assigned to each layer detected in the image based on which a bokeh can be rendered. As the algorithm makes use of a GAN based network to generate a depth map, it is dynamic in nature while also being computationally light thus making it suitable for mobile computing.

II.PROPOSED ALGORITHM

Our algorithm consists of the following steps:

- Depth Map Generation – A GAN based architecture is used to generate a depth map which is then used to estimate the number of layers present in the image.
- Foreground and background separation – We use DeepLab v3+ in order to achieve it.
- Object Classification – We use YOLO V3 for object classification along with a Local Interpretable Model-Agnostic Explanation unit in order to improve the quality of the prediction.
- Subject Color Enhancement – The object in focus is then color-enhanced based on the object type which is predicted by the previous iteration of the algorithm.
- Bokeh Rendering – Based on the number of layers estimated from the first step, we use kernels of different sizes to vary the strength of the bokeh at each level.

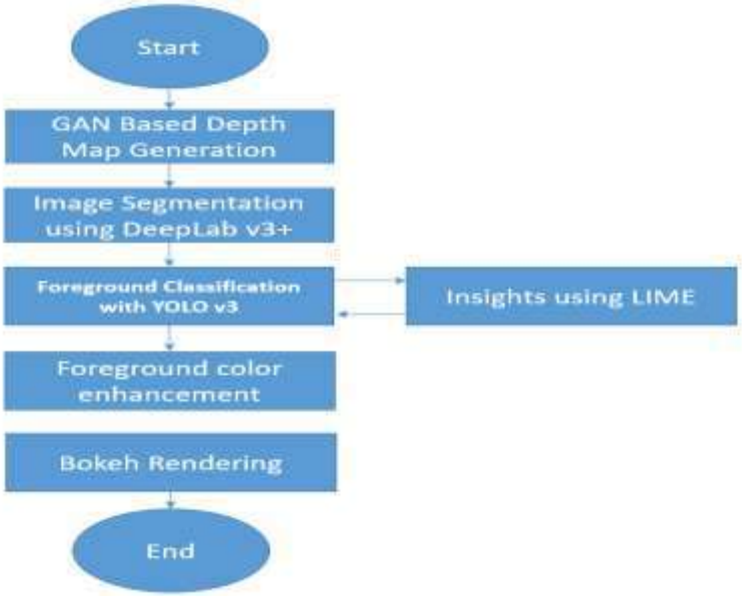

Start
GAN Based Depth Map Generation
Image Segmentation using DeepLab v3+
Foreground Classification with YOLO v3
Insights using LIME
Foreground color enhancement
Bokeh Rendering
End

Fig. 1 Algorithm Flowchart

A. Depth Map Generation

Recent studies have showed great improvement in depth map generation. [5] uses an encoder-decoder architecture which regresses the depth map from a sparse depth map of an image which is significantly better from methods such as [6, 7].
The image is passed to the encoder which is responsible for extracting the feature map (F) from the input image. The decoder then uses this output in order to produce ground truths to generate the depth map.

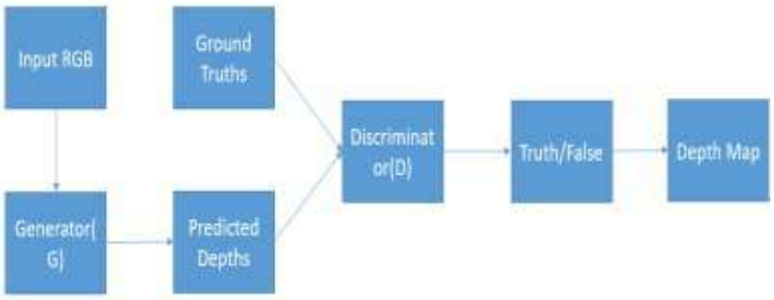


Fig. 2 Encoder-Decoder Architecture

Encoder
The encoder consists of 6 internal layers. At each processing layer, number of features extracted from the image increases. At the 4th layer, an inception net [9] is connected which helps in widening the network instead of deepening it. Max pooling is done to reduce dimensionality of the image and finally, aggregates of the features is done and sent to the decoder.

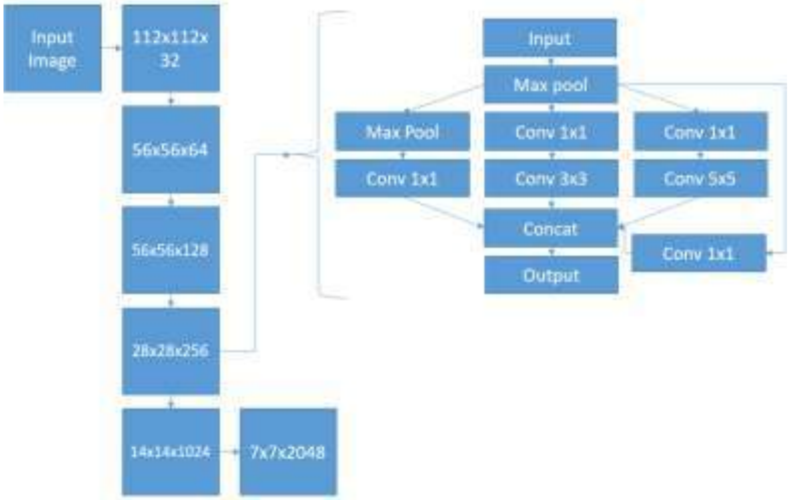


Fig. 3 Encoder Architecture

Decoder

It takes the output from the encoder and generates a semantic segmentation mask. It follows a 6 layer architecture. It takes learnt representations from the encoder and generates ground truths in order produce the correct output sequence. Finally it up-samples the image to match the desired image resolution.

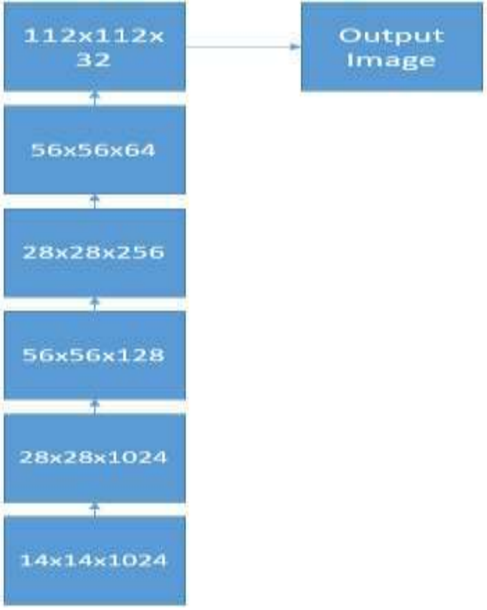


Fig. 4 Decoder Architecture

The function for this architecture is given as:

$$min_G max_D \ = E_{x\sim P_G}[\log(D(x)] + \ E_{\hat{x}\sim P_G}[\log\,(1 - D(\hat{x}))]$$

Where,
x is the ground truth and,
$\hat{x}$ is the depth map predicted by the generator.

For our input image, the depth map was generated as shown in figure 6.

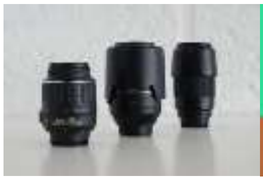


Fig. 5 Input Image

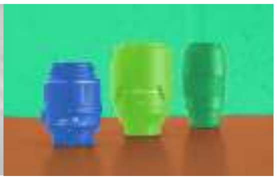


Fig. 6 Depth Map

B. YOLO v3

In order to separate the foreground and background, we use Google's DeepLab v3+ [10] model.

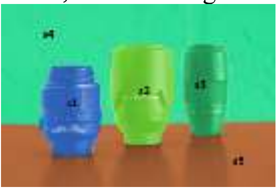


Fig. 7 Foreground and Background Separation

Subjects' s1, s2 and s3 are part of the foreground whereas s4 and s5 are part of the background.
YOLO is a Convolutional Neural Network (CNN) that can easily recognize objects. When compared to other networks, YOLO has the benefit of being significantly faster while yet maintaining accuracy. In YOLO, predictions are produced using a global context. In terms of mean average precision (mAP) and intersection over union (IOU) values, YOLOv3 is quick and precise. This enables us to preserve the temporal constraints associated with mobile computing while properly detecting objects in an image. We change the output layer of our classifier to classify 10 classes based on the most commonly photographed subjects for bokeh. These are as follows: Dogs, Cats, Humans, Flowers, Food, Cars, Bikes, Birds, Insects and Drinks.
Our model successfully recognizes these objects and color correction is performed on them in the next stage.

C. Local Interpretable Model-Agnostic Explanations Unit

The main objective of the classifier in the proposed algorithm is to recognize objects and separate the foreground from the background. We use YOLOv3 as a base classifier in order to achieve this. However, recognizing all possible objects within an image with 100% accuracy is difficult. Furthermore, YOLOv3 is a global approximator and hence requires improvement on local fidelity. Thus, we re-train our classifier to recognize an object in general based on features such as focus, intensity of light, the detailing in an object etc.
To improve this object recognition, we use the Local Interpretable Model-Agnostic Explanations (LIME) unit. By changing the input and observing how the predictions change, LIME is used to understand how the underlying model behaves. We alter the input by modifying visual components of an image, which helps us interpret the reason for a particular prediction [8]. By replacing the underlying model with an interpretable one through perturbations of the original instance, we produce an explanation for a prediction.

iJETRM
International Journal of Engineering Technology Research & Management

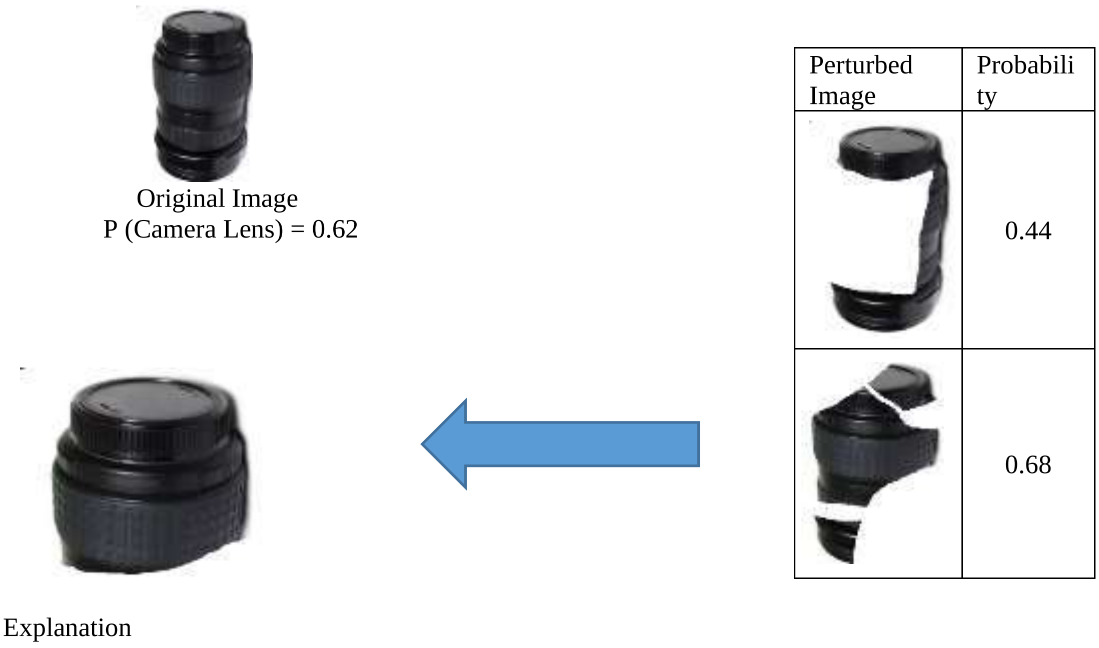


Fig.8 Working of LIME

LIME aids our model to ensure 100% recognition of the object, background and the foreground. It returns a list of explanations which help us understand how the model interprets above mentioned parameters.

D. Subject Color Enhancement

Using the classification obtained from the YOLO v3 unit, we then color enhance the image by firstly converting the RGB domain to hue, saturation and value (HSV) domain. From this we refer to the color correction table that we prepared which consists of the amount of color boosting required as per the input image luminance value for each class type. For example, if the object is classified as a human, saturation is modified to 1.1x and hue is boosted by a factor of 1.05x. This is done as facial features and color tones play an important role in human subject photography. Similarly, we have tested and tried out different ratios for each class.

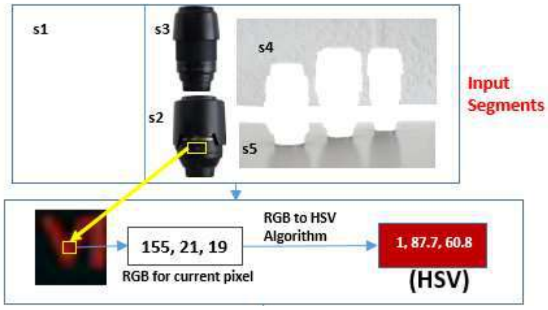


Fig. 9 RGB to HSV Conversion

In figure 9, the values are then boosted by a factor of 1.12 resulting in the color code (1.12, 98.22, 21.28).

E. Bokeh Rendering

After estimating the number of layers and enhancing the subject, the final step is to render the bokeh. We use different kernel sizes, the size of the kernel for each layer is calculated as:

$$kernel_{size} = (2\ X\ N_{layer}) + 1$$

The kernel then is created of the size [kernel$_{size}$, kernel$_{size}$].
Referring to figure 4, the image consists of 4 number of layers. Layer 1 has a kernel size of [3, 3], layer 2 has a kernel size of [5, 5] and so on. We apply a Gaussian Blur [11] using this kernel at each layer. The resultant image is as followed:

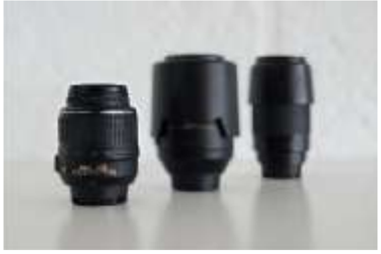


Fig. 10 Our Bokeh Rendering

III.RESULTS

We have compared the results of our algorithm with similar works. Main differences that we noted were:

- Most methods simply blur the background uniformly instead of varying the blur level.
- Methods do not enhance the subject
- The foreground stands out more in our model compared to other models due to the gradual increase in blur level applied.

In figure 11, we see that details in the foreground are much sharper and better visibly separated from the background in our bokeh effect. Unwanted Contrasting details are diminished in our bokeh effect while even preserving the general background object details. Segmentation along with ROI separation helps to minimize the Unwanted Details in the Non ROI background regions.

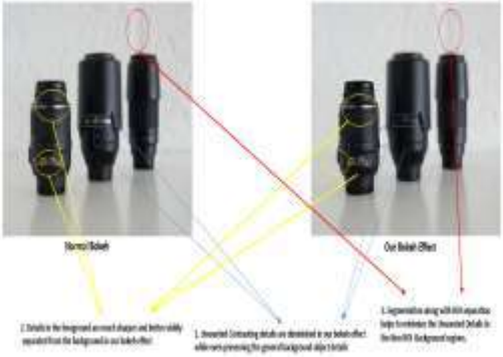
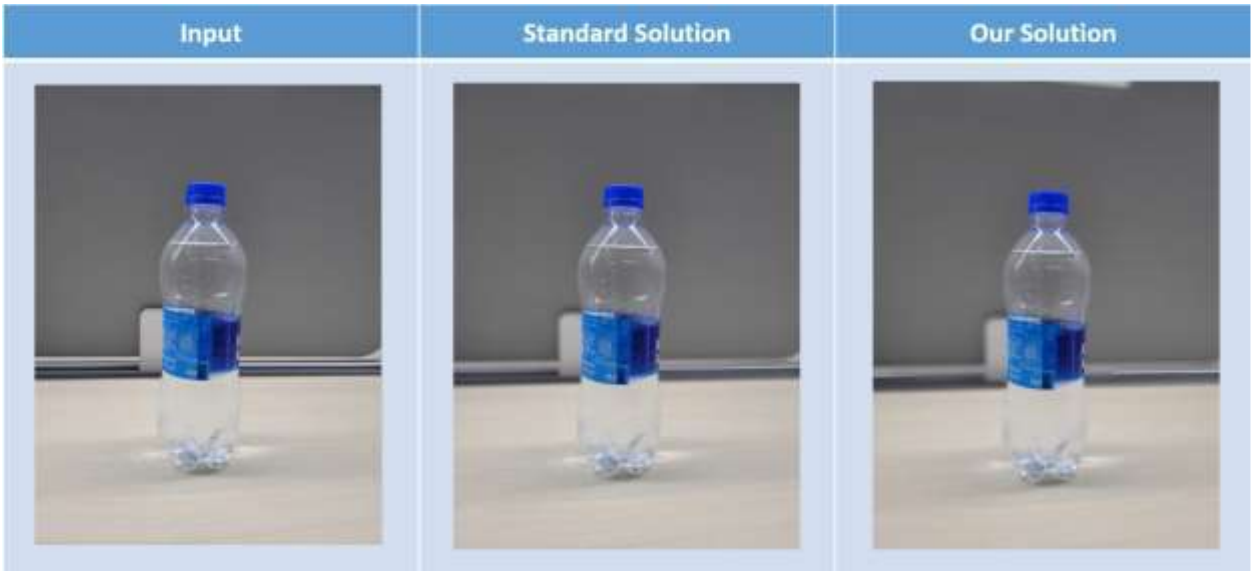
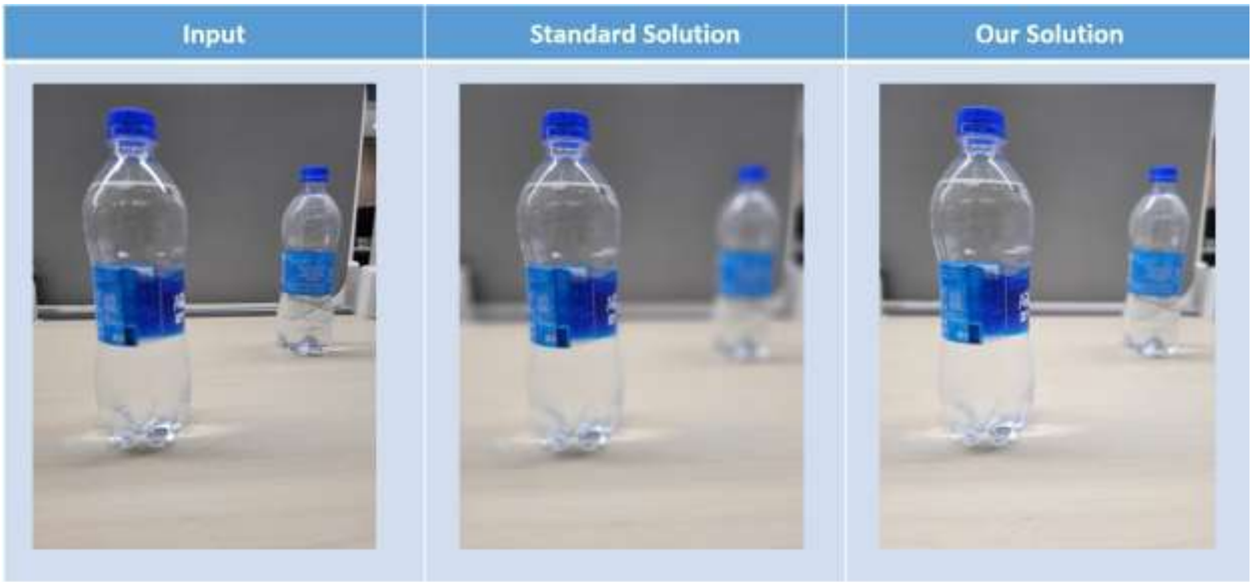

Normal Bokeh
Our Bokeh Effect
2. Details in the foreground are much sharper and better visibly separated from the background in our bokeh effect
1. Unwanted Contrasting details are diminished in our bokeh effect while even preserving the general background object details
3. Segmentation along with ROI separation helps to minimize the Unwanted Details in the Non ROI background regions.

Fig. 11 Normal Bokeh vs Our Bokeh Effect



Result 1



Result 2

In result 1, the subject of the image is a single bottle. The bokeh effect generated using the standard algorithm results in unnatural blurring of the background. This blurring is non-uniform and relatively low. However, the bokeh generated by the proposed algorithm generates a much more evident blur effect of the background while also performing visible edge sharpening on the subject of the image. This highlights the foreground hence adding to the distinction effect of the bokeh.

In result 2, the input image consists of two bottles positioned at different depths. One bottle being closer to the lens while the other being slightly distant. In the standard algorithm, there is a uniform blurring effect throughout the background hence reducing the focus from the second bottle altogether. This also reduces the information regarding the depth of the second bottle. We also observe that the edges of the primary subject are relatively soft.

The proposed algorithm applies a less significant blurring effect on the second bottle according to the distance from the lens. This ensures retention of the depth information in the second bottle. The algorithm also performs color enhancement on the primary subject which makes it appear much more distinctly as compared to the standard algorithm.

IV.CONCLUSION

In this paper we discussed about a GAN based network to generate a depth map which is then used along with a data driven object classification method in order to generate a variable bokeh effect on an image to highlight different layers with different kernel strengths. We also discussed their working and compared its results with other models. Upon comparing, we note that our model performs better when there are multiple objects at multiple lengths from the camera. This creates a more prominent bokeh effect than the generic method.

Further improvements can be done by using different types of blurring kernels, which provides the user with a variety of blur options that can be applied on the image; optimizing the GAN to detect more layers with more accuracy hence increasing the sharpness of the effect; and experimenting with other color formats.

V.REFERNCES

[1] Li, Y., Qian, K., Huang, T., & Zhou, J. (2018). Depth estimation from monocular image and coarse depth points based on conditional gan. In MATEC Web of Conferences (Vol. 175, p. 03055). EDP Sciences.

[2] Viraktamath, S. V., Yavagal, M., & Byahatti, R. (2021). Object Detection and Classification using YOLOv3. International Journal of Engineering Research & Technology (IJERT), 10(02).

[3] Zhao, X., Huang, W., Huang, X., Robu, V., & Flynn, D. (2021, December). Baylime: Bayesian local interpretable model-agnostic explanations. In Uncertainty in Artificial Intelligence (pp. 887-896). PMLR.

[4] Vishwakarma, A. K., & Mishra, A. (2012). Color image enhancement techniques: a critical review. Indian J. Comput. Sci. Eng, 3(1), 39-45.

[5] Fangchang Ma and Sertac Karaman. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In 2018 IEEE international conference on robotics and automation (ICRA), pages 4796–4803. IEEE, 2018.

[6] Miaomiao Liu, Mathieu Salzmann, and Xuming He. Discrete-continuous depth estimation from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 716–723, 2014.

[7] Qingxiong Yang. Stereo matching using tree filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 37(4):834–846, 2014.l

[8] Rebeiro M., Singh S., Guestrin C. "Local Interpretable Model-Agnostic Explanations (LIME): An Introduction", 2016 URL: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

[9] Hao, P., Zhai, J. H., & Zhang, S. F. (2017, July). A simple and effective method for image classification. In 2017 International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 1, pp. 230-235). IEEE.

[10] Si, Y., Gong, D., Guo, Y., Zhu, X., Huang, Q., Evans, J. & Sun, Y. (2021). An Advanced Spectral–Spatial Classification Framework for Hyperspectral Imagery Based on DeepLab v3+. Applied Sciences, 11(12), 5703.

[11] Gaussian Blur: Digital Image Processing, 2005 -2022. (n.d.). Science Direct. Retrieved September 15, 2022, from https://www.sciencedirect.com/topics/engineering/gaussian-blur