The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 22, Pages 142-151

ICBASET 2023: International Conference on Basic Sciences, Engineering and Technology

Spotting the Differences between Two Images

Raghunadh M V
National Institute of Technology

Srikanth KOTAKONDA
National Institute of Technology

Abstract: This paper presents a generalized solution to the classical problem of spotting the differences between two images. In this digital era, the authenticity of an image has become a big challenge to the researchers and engineers in the field of computer vision and image processing. Due to the rapid developments in digital technology, creation of photographic fakes and image manipulation has become easily accessible to everyone. With the availability of open-source editing software tools, the possibility of various image manipulations like image forgery, image tampering and image splicing have become almost inevitable. This paper addresses the problem by using classical image processing techniques along with the state-of-the-art YOLOv8 deep learning object detection algorithm. The results obtained are very promising when the model is trained on a synthetic dataset of 700 pairs of images. The uniqueness of the dataset is that each pair of images is different from any other pair of images and the number of differences between any two images may vary from 1 to 50.

Keywords: ORB, SIFT, SSIM, CNN, YOLO

Introduction

Due to the exponential growth in digital technology, the viewer almost lost the trust on the visual content and the situation can become even worse with the increase of more advanced processing tools. Many industries are automating their processes and hence it is very important to have a more robust and reliable tool which can spot the differences between the reference image and the test image.

The problem is approached in a quest to find a reliable solution by exploring the areas of document image analysis, quality inspection and robotics, remote sensing applications, medical imaging, biometrics etc. The work is mainly divided into two parts. The first part focuses on applying traditional image processing techniques like image alignment, ORB (Orient Fast Rotate Brief), SIFT (Scale Invariant Feature Transform) with SSIM (Structural Similarity) and the second part comprises of applying the state-of-the-art deep learning YOLOv8 (You only look once) object detection algorithm developed by ultralytics.

Method-I (Image processing with ORB/SIFT and SSIM):

In this method, FAST (Feature accelerated segment test) is performed on the given two images to extract the key/corner points and then feature point screening is done to get more useful corner points. Later, an intensity centroid method is used to find the dominant directions of key points and then feature descriptors are generated using BRIEF (Binary Robust Independent Elementary Feature) vector. Finally, BRIEF is improved by using steered BRIEF with patch orientation before applying homography for aligning the images.

Once the images are aligned, a method called SSIM (Structure Similarity) is applied by passing the difference between the two images as a threshold to get the desired result.

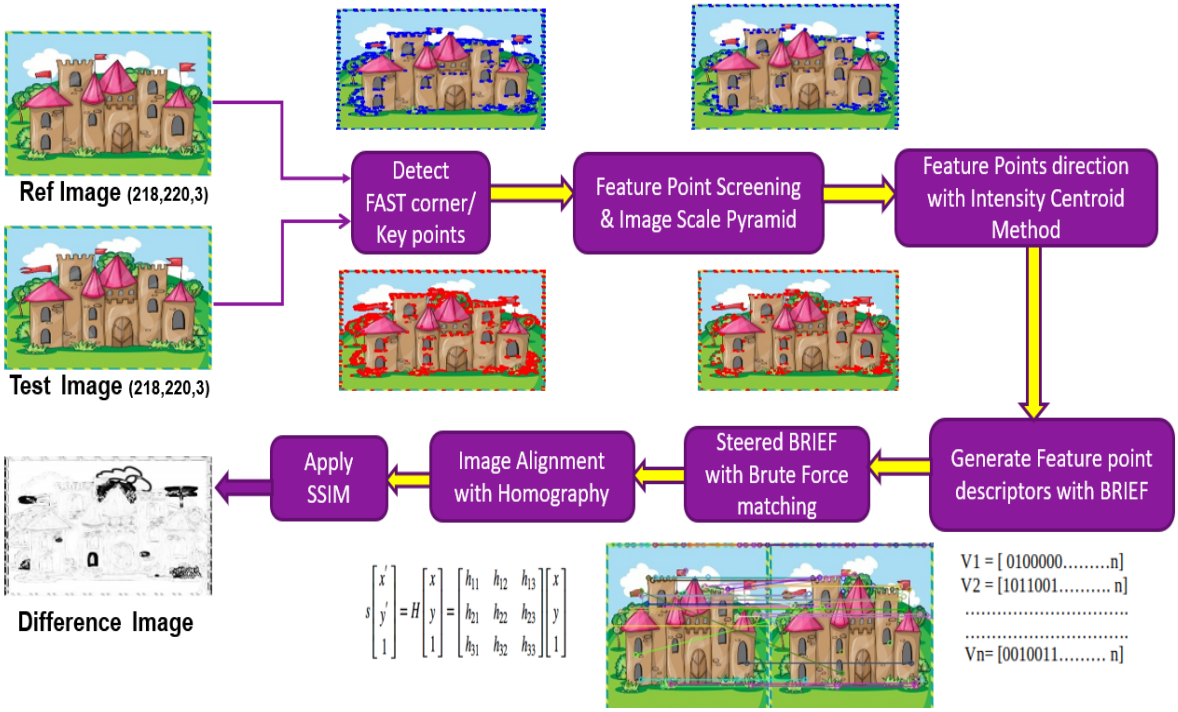


Figure 1. Block diagram describing the workflow of method-I

Mathematical Analysis

Using FAST (Features from Accelerated Segment test) and Breshnam Circle concept, it chooses 16 pixels at random and categorizes into 3 classes (Brighter, Darker & Similar). If more than 8 pixels are brighter than I_p (x, y), then it will be considered as a key point or feature point.

$$R = det(M) - k(trace(M))^2 \quad (1)$$

$$M = \sum w(x,y)\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \quad (2)$$

Figure 2. Feature point screening using harris response values

After locating key points or feature points, for detecting intensity change, a method called intensity centroid. First, a small image block B is considered, the moment of the image block is defined as

$$m_{pq} = \sum_{x,y\in B} x^p y^q I(x,y), \quad p,q = \{0,1\} \quad (3)$$

where x and y are pixel coordinates, and $I(x,y)$ is the gray value of the corresponding pixel. Then, find the centroid of the image block by the moment:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \quad (4)$$

Figure 3. Moment of image block and centroid

$$\theta = \arctan(m_{01}/m_{10}) \quad (5)$$

BRIEF takes all points found by FAST algorithm and converts it into a binary feature vector so that together they can represent an object. Binary feature vector is also known as Binary feature descriptor that only contains 1's and 0's. In BRIEF, each key point is described by a feature vector which can be either 128 bits or 512 bits string.

For example:

V1 = [0100000……n]

V2 = [1011001……. n]

…….

…….

V_N = [0010011…… n]

BRIEF starts by smoothing image using a gaussian kernel to prevent the descriptor from being sensitive to high frequency noise. It selects a random pair of pixels in a defined neighborhood around that key point. This defined neighborhood around the pixel is also known patch which is a square of some pixel width and height.

The first pixel in the random pair is drawn from a gaussian distribution centered around the key point with a standard deviation or spread of sigma. The second pixel in the random pair is drawn from a gaussian distribution centered around the pixel with a standard deviation or spread of sigma by two. Now if the first pixel is brighter than the second, it assigns the value of 0 to the corresponding bit else 1 is assigned. This process is repeated for 128 times for each key point, in this way BRIEF creates a vector for each key point in an image. This is called Binary test.

$$\tau(p;x,y) = \begin{cases} 1, p(x) < p(y) \\ 0, p(x) \geq p(y) \end{cases}$$

Figure 4. Performing binary test

Where p(x) is the grey value at the field x around the image feature point and p(y) is the grey value at the field y around the image feature point.

Finally, an N-dimensional vector $f_n(p)$ is obtained as

$$f_n(p) = \sum 2^{(i-1)} \tau(p;x,y) \qquad \text{where } 1<i<n \qquad (7)$$

But still BRIEF isn't invariant to rotation, so we use rBrief (rotation aware brief or steered brief) using Rotation matrix as shown below

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \qquad (8)$$

Matrix R_θ and N pairs of pixel points form a matrix Q:

$$Q = \begin{bmatrix} x_1, x_2, \cdots, x_N \\ y_1, y_2, \cdots, y_N \end{bmatrix} \qquad (9)$$

Then a rotation correction is performed to get Q_θ:

$$Q_\theta = R_\theta Q \quad (10)$$

Finally, we can get a directional descriptor:

$$g_{N(p,\theta)} = f_N(p) | (x_i, y_i) \in Q_\theta \qquad (11)$$

Figure 5. Rotational correction and directional descriptor

Then we use Brute force matching using RANSAC method and apply Homography for image alignment as shown below:

$$s\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Figure 6. Applying homography for image alignment

Finally passing the obtained H on to cv2.warpperspective(source, destination, H, size) we get a completely aligned image for further processing as shown below:

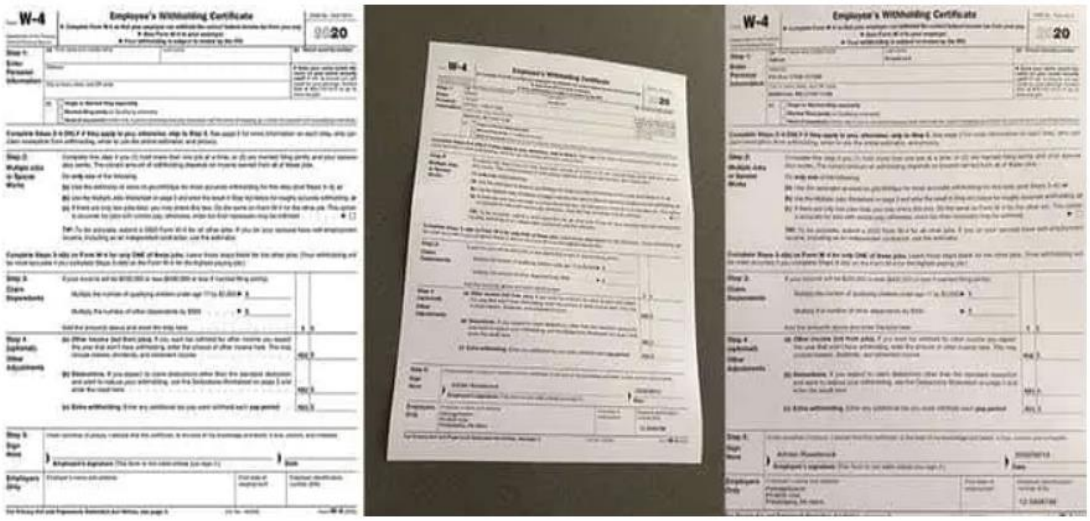


Figure 7. An example of image alignment

Then, applying a method called compare SSIM and using difference as threshold we will get the result image as a difference image between the given pair of images as shown below:

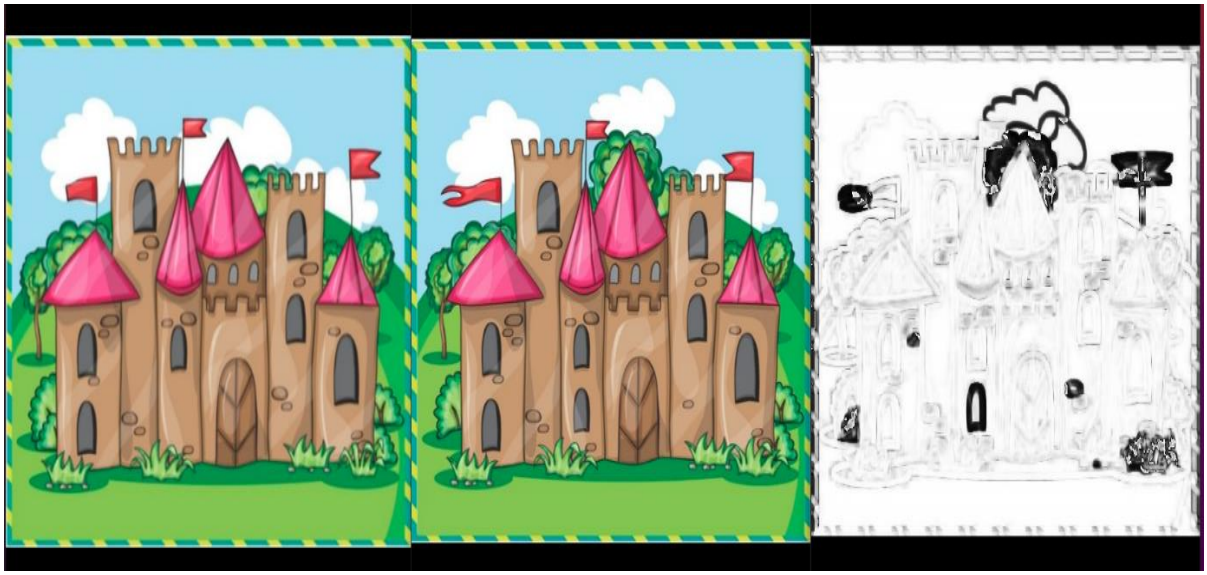


Figure 8. An example of spotting differences between two images

Method-II (Applying CNN and Deep Learning):

From Figure 9, It is clear that up to Image alignment all the steps are common as in method-I, the only change is that once the images are aligned, a new method called Concatenation is applied on the third dimension and then a reshape is done by expanding in horizontal direction, It can be clearly seen that because of this operation we can clearly see that the areas where the images differ have some vertical lines giving us a hint that the images differ in this part of the image.

Taking this as a fundamental building block, we take our work to the next level by applying some edge filter on the concatenated and reshaped image to get the edges and then apply a binary inversion with a high threshold resulting us the difference areas between the two images. As we can see clearly in figure 10, that the operation of concatenation and reshaping is giving us a good insight to spot the differences between the images. So, before applying the deep learning model, we tried to apply a basic two layered customized CNN (Convolutional Neural Network) with 'X' (Concatenated + reshaped image) as the input and 'Y' (Edge filtered + Binary inversion) as the target for the model to see the viability of deep learning approach to this problem.

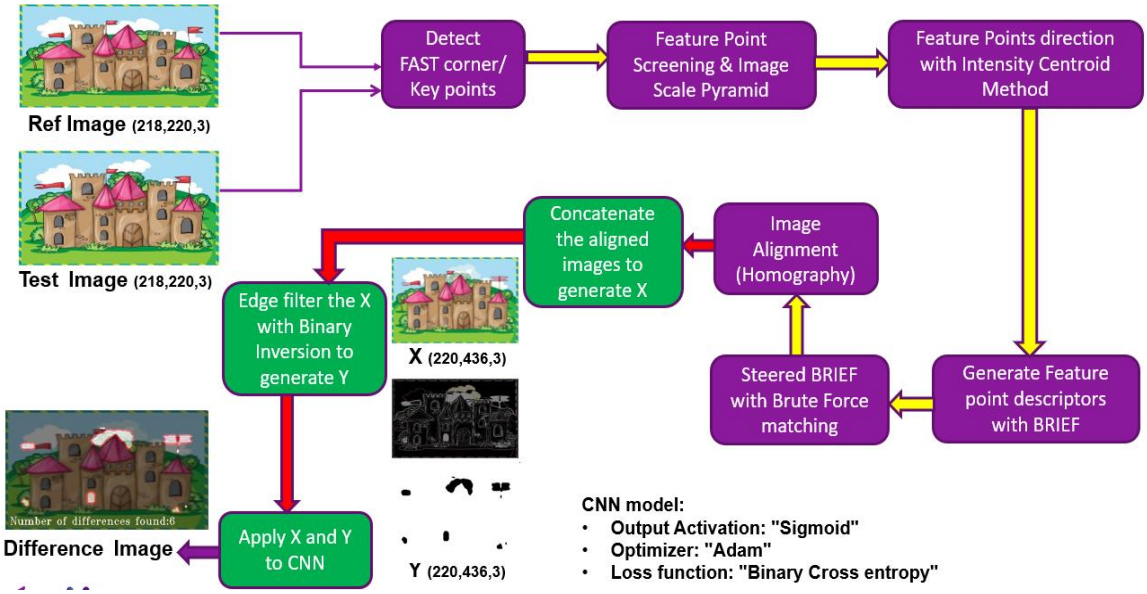
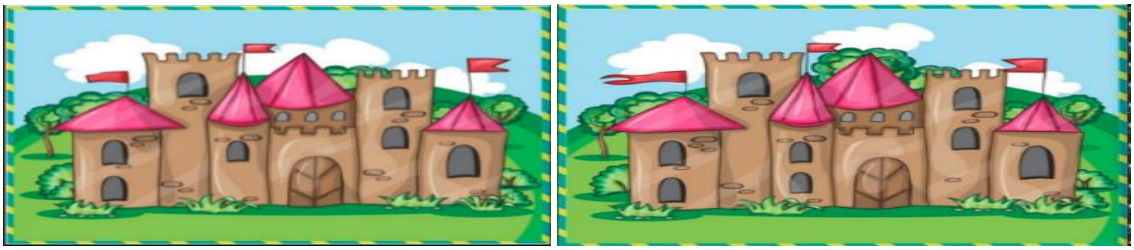


Figure 9. Method-II applying a customized cnn model to the problem



Reference image Test image

Figure 10. An example of concatenated and reshaped image

When this customized CNN model with 'Sigmoid' as the output activation function, the optimizer as 'Adam' and loss function as 'Binary cross entropy' is trained on a synthetic dataset of 700 images, the trained model when used to predict on a pair of unseen images, the results obtained are very promising as shown below:

Reference image Test image Difference image

Figure 11. Customized cnn model results.

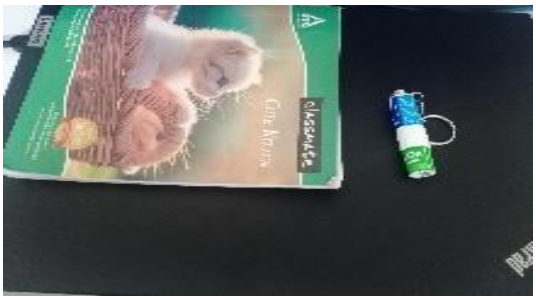
Drawbacks and Improvements

Although the above methods of Image processing and use of a customized CNN provided us some promising results, but there are some failure cases that can be seen from figure 11 itself where the cloth color in reference image is green, whereas the cloth color in test image is red, but the model failed to spot the difference. Apart from this, there are also some failure cases when the objects inside the images are displaced resulting in too many false positives.So, to overcome these limitations, we took our work further to apply deep learning algorithm.

Deep Learning with YOLOv8:

In this section, we annotate 700 images by labelling the areas of differences in the concatenated + reshaped images with a label called 'diff' converting the problem into a kind of object detection but we have only single class to detect that is 'diff'. Now as we have converted our problem to an object detection case, considering the large-scale availability for state-of-the-art object detection algorithms, we preferred to choose the latest YOLOv8 object detection algorithm developed by ultralytics. When we fine-tuned YOLOv8 pretrained model on our customized dataset the results are as follows:

Reference image



Test image

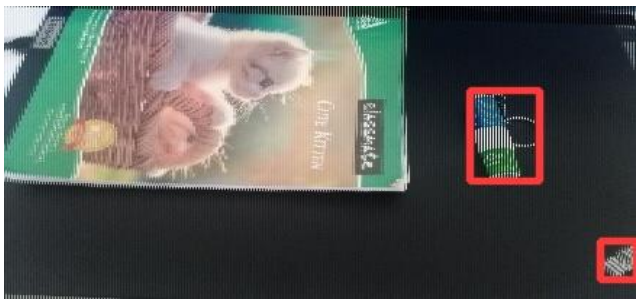


Figure 12. Yolov8n pre-trained model result.

Results and Discussion

Fine-tuned '**YOLOv8n**' pre-trained model result graphs:

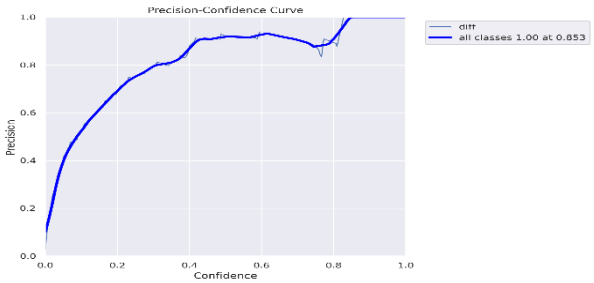


Figure 13. Precision vs confidence

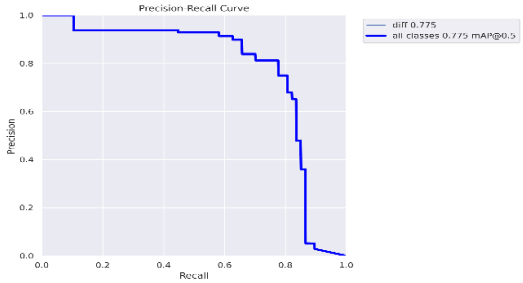


Figure 14. Precision vs recall

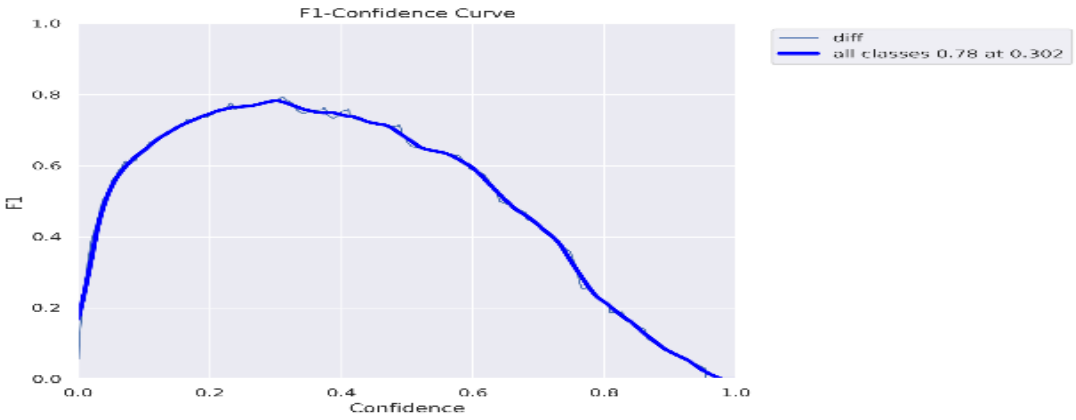


Figure 15.F1-score vs confidence

The following are the results when the model is fine-tuned with '**YOLOv8m**' pre-trained model:

Reference image Test image

Figure 16. Yolov8m pre-trained model result.

Fine-tuned '**YOLOv8m**' pre-trained model result graphs:

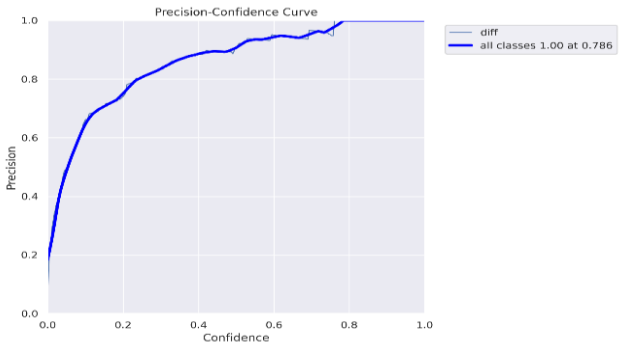


Figure 17. Precision vs confidence

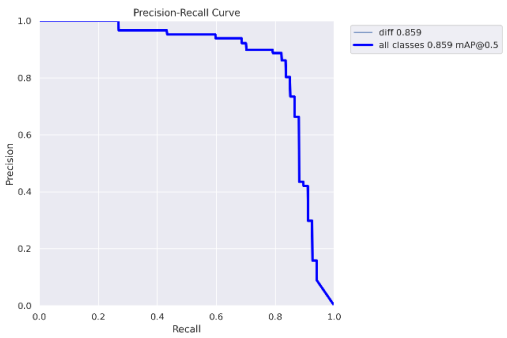


Figure 18.Precision vs recall

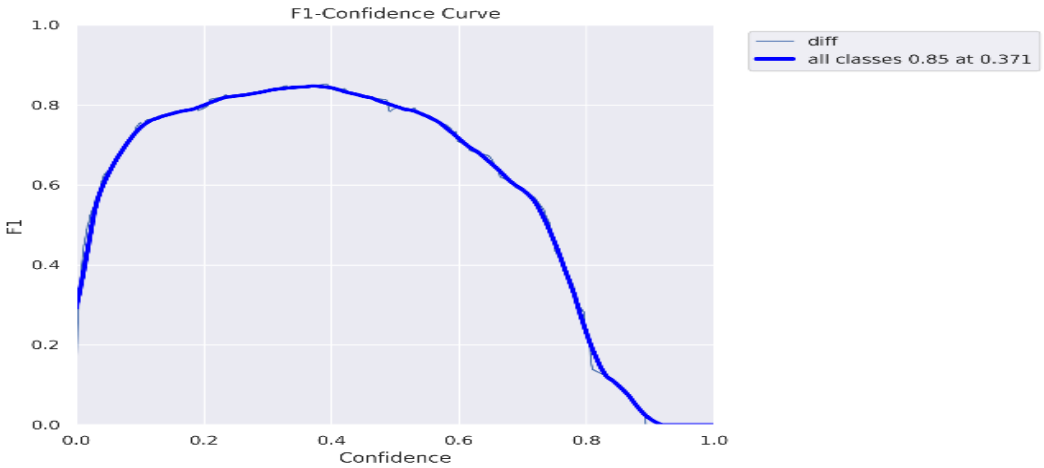


Figure 19.F1-score vs confidence

Conclusion

Although the problem of 'spotting the differences between two images' appears to be simple, but still it continues to exist in the research of computer vision and image processing. Hence considering the gravity of the problem, we have started from the classical image processing and explored till the latest state of the art deep learning algorithm('YOLOv8').The results after fine-tuning the YOLOv8 pre-trained models are convincing and if the model is trained on a huge dataset, it might produce very good results with good confidence irrespective of any internal variations like object displacement or external variations like rotational and luminance changes.

Recommendations

The use of more advanced image alignment techniques may give us improved results and reduces the false positives. There are totally five YOLOv8 pretrained models and in this paper, we used the light weight models YOLOv8n and YOLOv8m, for better results further research can be done using the other pre-trained variants of YOLOv8 and also improvise with the optimizer and loss function to improve the confidence in the results. This paper may give a new dimension to the problem of spotting the differences between two images and if explored further can even produce good results.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

*This work could not have been possible without the support and facilities provided by National Institute of Technology, Warangal, India. Our sincere thanks to all those who contributed to this work either directly or indirectly.

* This article was presented as oral presentation at the International Conference on Basic Sciences, Engineering and Technology (www.icbaset.net) held in Marmaris/Turkey on April 27-30, 2023.

References

Bianco,S., Ciocca, G., & Schettini, R.(2015). How far can you get by combining change detection algorithms? *ArXiv*.

Girshick, R., Donahue,J., Darrell,T.,& Malik, J.(2014) Rich feature hierarchies for accurate object detection and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 587.

Jain,R., & Doermann,D.S.(2013). Visualdiff: Document image verification and change detection. *2013 12th International Conference on Document Analysis and Recognition*, 40–44.

Qin, H., Yan, J., Li, X., & Hu,X.(2016). Joint training of cascaded cnn for face detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3456– 3465.

Wu,J., Ye,Y., Chen,Y., Weng, Z.(2018).Spotting the difference by object detection. *ArXiv*.

Author Information

Raghunadh M V
National Institute of Technology
Warangal,India
Contact e-mail: raghu@nitw.ac.in

Srikanth Kotakonda
National Institute of Technology
Warangal, India

To cite this article:

M V, R.. & Kotakonda, S. (2023). Spotting the differences between two images. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 22*, 142-151.