International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Small Object Detection in Recognizing Text from Multi Aspect Image Dataset Using Artificial Intelligence

Rishabh Dixit[1], Sushil Kumar Maurya[2]
Computer science and Engineering Department,
Bansal Institute of Engineering and Technology, Lucknow
rishabh.saurabh333@gmail.com

Abstract: In earlier years, in smart cities, object detection system has been improved by several folds due to many novel deep learning models. Deep learning has outperformed the existing traditional computer vision techniques. Recently, many automated segmentation models have used the concept of feature extraction; the model proposes various feature extraction on the images. The models generally use a texture classification model and color detection models to predict the label of neighboring pixels, and the classification is used to validate the features. The parameter tuning of these models is generally based on the signal processing schemes. Many researchers have used optimized parameter sets obtained for a specific dataset, due to which the accuracy increases drastically. Still, the model is not scalable on any other data set. We propose a new hybrid technique using wavelet and watershed transform for small object detection for text extraction. Our hybrid model is scalable over a variety of images, the model is used on different photos, and the result shows improvement in accuracy in the presence of varying noise.

Keyword: Object detection, Masking DWT, watershed, small object, Morphological.

1. Introduction:

An object identification system finds real-world pairings using photographs of the world and and a priori-defined object model [1, 2]. The demand for object recognition systems is provisos to the ever-growing volume of digital photographs in all public and private databases accessed in smart cities via CCTV. Humans can make object identification extremely simple and readily, while robots find it quite challenging. Object recognition technologies are critical for robots to achieve increased levels of autonomy [3]. It is a popular field of robotics study that uses computer vision (CV) with machine learning (ML). Drones are increasingly getting employed as robotic devices. This study aims to determine how to present object identification algorithms and methods may be applied to drone image information. Compared to certain other robots [4, such as a wheeled robot], another of the benefits of utilizing a drone to identify items in a picture is that the drone can travel closer to objects.

Nevertheless, using a top - down perspective angles [5] and the difficulty of combining using a deep learning system for compute-intensive processes [6] pose challenges regarding UAVs. Whenever a drone navigates an area in the discovery of items, it is advantageous for the drone to have as much visibility as feasible [7, 8]. On the other hand, images captured by UAVs or drones varied significantly from those charged by a conventional camera. As a result, object identification techniques employed on "normal" photos cannot be presumed to work effectively on photographs captured by drones. Previous research has found that images obtained by a drone are frequently different from those accessible for training that is commonly shot with a hand-held camera. The study helps to identify drone information could be tough owing to camera's placement [9, 10] compared to images captured by a human, based on sort of pictures the network is trained on.

Object recognition is an automated image identification procedure dependent on statistical and geometric aspects that also demands precise categorization of items in pictures and involves accurate object estimation. The efficiency of object categorization and positioning are critical measures of model identification performance so we can easily recognize the thing that make the. Object recognition is utilized in a variety of applications, including intelligent surveillance, military object identification, UAV navigation, unmanned vehicle navigation, as well as smart transportation. The existing model, moreover, needs to recognize objects due to the diversity of the identified items. The complexity of object recognition is made more complex by changing light and a complicated background, particularly for objects in a diverse context. The standard multiscale pyramid approach of classification tasks and localization requires extracting the picture's statistical multistate data and subsequently classifying the image with a classifier [1–3]. It is tough, but many characteristics describe things that don't create a robust classification model since various types of photos are defined by distinct factors. As a result, those models could not identify the items, mainly when many recognized objects were in a single picture [1, 2].

Deep learning is becoming the standard approach for object recognition because of its expertise in object identification. These approaches (e.g., RCNN [4], Fast-RCNN [5], Faster-RCNN [6], SPP-Net [7], as well as R-FCN [8]) are effective at detecting multiple objects in pictures. However, for training

and testing, most such object recognition algorithms use the PASCAL VOC dataset [9]. The PASCAL VOC dataset is the most extensively utilized benchmark dataset in object categorization and identification because it enables a consistent assessment framework for identification techniques and learning capability. The dataset contains 20 catalogs relevant to human life, covering animals (horse, sheep bird, dog, cat, as well Moscow) plus humans, vehicles (car, bus, aircraft, bicycle, ship, motorbike, as well as the train), as well as the indoor item (ship, aircraft, train, car, motorcycle, bus, plus bicycle) (sofa, t.v., plus bottle,). As humans may observe with the individual item mentioned, most items with a set of data are enormous things. Although when usually were a few exceptions, little things with image, and focal length, like bottles, cause these small objects to appear to be quite enormous. As a result, a dataset-based recognition strategy, mainly a sizeable item, will not be successful in detecting minor things [10,11,12].

2. Related Works

In the realm of machine intelligence, object recognition is constantly a hot issue. The traditional sliding window recognition approach necessitates decomposing pictures into multi-scale pictures. Typically, a single image is divided into several sub-windows with millions of distinct positions and sizes. The system then uses a classifier to identify whether the observed entity is present in every window. Unfortunately, the procedure is wasteful since this necessitates a thorough search. Furthermore, various classifiers have an impact on object identification efficiency. Classifiers are developed per different observed types in sequence to generate a robust classifier. For instance, the Harr feature paired usingAdaboosting classifier [14] makes face identification possible. Furthermore, we employ HOG features (Histogram of Gradients) paired using a support vector machine [15] for pedestrian recognition, as well as HOG features mixed through DPM (Deformable Part Model) [16, 17] for common object recognition. Furthermore, if a picture has many types of recognized items, such entities will be unnoticed by computers.

Since 2014, when Hinton utilizes deep learning to win the ImageNet competition with the highest success rate, deep understanding has become a popular method for detecting objects. The object recognition model dependent on deep learning is separated into two classes: the first one is focused on region suggestions [18–20], like RCNN [4], SPP-Net [7], Fast-RCNN [5], Faster-RCNN [6], and R-FCN [8], as well as is the most extensively utilized. The other technique, like YOLO [21] and SSD [22], does not employ area suggestions and instead identifies the objects immediately.

For the first technique, the model selects ROIs first before detecting them; that is, several ROIs are created using chosen searches [23], edge box [22], or RPN [21]. The system subsequently uses CNN to extract features for every ROI, identifies items using classifiers, and determines the position of discovered objects. RCNN [4] generates around 2000 ROIs for every image using selective search [23] and then isolates and classifies convolution characteristics of 2000 ROIs correspondingly. Since there are a lot of people in such ROIs overlapping portions, identification is inefficient due to the enormous number of consecutive computations. For such a situation, SSP-net [7] and Fast- RCNN [5] provide mutualRoIsfunction.The approaches recover just a CNN characteristic from the entire source picture, while so every ROI's characteristic is retrieved separately from CNN feature using the ROI pooling process. As a result, the work required to retrieve every ROI feature is distributed. This approach lowers the 2000-times CNN operations in RCNN to single CNN operations, considerably improving computational efficiency.

YOLO doesn't employ region proposals and instead does straight convolution operations on the entire picture, making it quicker than Faster-RCNN in speed but less accurate. SSD additionally convolutions the image with a singular convolution neural network and predicts a series of boundary boxes with varying widths and length-to-width ratios at every item. During the test phase, networks expect each kind of item for every bounding box and modify the boundary box to fit the object's form. According to G-CNN [24], object detection challenges transforming the identification box into a tangible box from a preset grid. The model splits the whole picture into multiple scales to generate the starting bounding box and then uses the convolution technique to extract features from the full image. Then, using the Fast-RCNN approach, feature picture A limited-sized characteristic appearance is surrounded with an original structuring element. Finally, we may use regression to get a more precise bounding box. After multiple repetitions, the bounding box will become the outcome. In brief, two object identification algorithms exist for the present mainstream; the first is more accurate but slow. The second one is slightly less accurate, but it is faster. Regardless of how the item is detected, the feature extraction approach employs a multilayer convolution technique to produce a comprehensive abstract object property for the target object. However, because features collected by the system are limited and can adequately capture the item's properties, this technique reduces detection performance for tiny target objects.

Furthermore, the PASCAL VOC dataset is the primary object's data identification, as well as it contains 20 different item types, such as livestock, buses, and pedestrians. However, most of the things in photographs are enormous. Though in PASCAL VOC are some small items, such as a cup, however, because of the focus length, such minor entities appear to be quite enormous in the picture. As a result, PASCAL VOC is ineffective in detecting small objects.

The Microsoft COCO dataset [24] is a standardized dataset for object identification and picture segmentation, as well as other domains created by the Microsoft team. The dataset contains various small items with varying background complexities, making it excellent for tiny object recognition. The SUN dataset has a total of 131067 photos with 908 scene categories

as well as 4479 object classifications and a massive number of small items.

The study [13] used two standards to create the rich tiny object dataset. The first is that the things are no more than 30 cm in size. Another condition is that the area filled by the items in the image does not exceed 0.58 percent. The author similarly provides the map of RCNN depending on the dataset, which has a recognition performance of just 23.5 percent.

3. Methodology:

The utilization of shading and surface data has solid connections with human observation. In numerous pragmatic situations, the shading alone or surface alone image data needs to be adequately strong enough to portray the substance of the picture precisely. An illustration is given through a division of stock photos that show shading and surface qualities. This instinctive psychophysical perception incited PC visualization specialists to examine an extensive range of scientific models by inspecting nearby and worldwide characteristics of such two major picture characteristics. In any case, vigorous reconciliation of shading and surface properties is a long way from a minor target. This is inspired, to some extent, by the trouble in separating exact shading and surface models, which may be used locally to adjust toward varieties in picture content. Specifically, the division of normal pictures ended up being a testing undertaking, since these pictures show critical inhomogeneities in shading and surface.

Moreover, they are frequently described by a high level of unpredictability, arbitrariness, and inconsistency. In addition, the quality of surface and shading traits can differ extensively from picture to picture, and intricacies included by the uneven brightening, picture commotion, point of view, and scale contortions make the way toward distinguishing the homogenous picture locales amazingly troublesome. Every one of these difficulties pulled in considerable enthusiasm from the vision analysts, as the hearty reconciliation of the shading and surface descriptors in the division procedure has significant ramifications in the advancement of larger amount picture examination undertakings, for example, question acknowledgment, scene understanding, picture ordering and recovery, and so forth.

Over the past few decades, the field of image division based on the union of shading and surface descriptors has overgrown, reaching a peak in 2007 and 2009 with a high number of calculations. The fact that more than 1000 papers were published between 1984 and 2009 is significant because it shows how color-surface research has become one of the most investigated areas in computer vision and image preparation. A specific field of study has advanced, according to measurements that consider how many calculations on color-surface assessment were performed during the last three years (2007–2009). As a result, it is feasible to pinpoint specific examples or classes of techniques that demonstrate the concepts of component retrieval procedures or contain coordination strategies. This essay's goal is to fictitiously analyses some of the most significant subheadings in the discipline of color science. To develop effective picture division, a lot of research is being done on surface evaluation and auditing concepts and methods that have been studied over time, including color surfaces. We are not aware of any written work that was concerned with the effective investigation of the concepts and methods that were used in the development of color-surface division calculations, even though a few reviews have concentrated on the evaluation of shading alone division calculations [1-3] or surface alone division calculations [4–9]. We would like to underline that in this audit, the examination and classification of the dispersed works concerning the synchronization of shading and surface data in the division method are of special importance to us. This, in our opinion, is the most clever strategy that can result in a substantial knowledge of this crucial topic. Two elements significantly lend credence to the established methodology. The main problem with creating color-surface division plans is the information combination (highlight reconciliation) procedure. Thus, this kind of inquiry supports a precise classification of the dispersed calculation as a first step. It will also be feasible to discern evidence of nonspecific color-surface mix designs independent of the application setting, which is the de facto norm for the shading and surface element extraction methods, thanks to this inquiry. Along these lines, the chief targets of this paper are: (a) to arrange the principle slants in color–surface incorporation, (b) to test the application setting of suggested executions (at whatever point such exchange is suitable), (c) to talk about the assessment measurements that are at present used to evaluate the execution of the division procedures, (d) to audit the freely accessible information accumulations (picture databases) and (e) to break down the execution of entrenched best in class usage. It is helpful to note that this audit is essentially worried about the investigation of calculations intended for the division of still computerized pictures, and we will demonstrate when the assessed approaches have been connected to the division of video information. To give an exhaustive understanding of entering the realm of labor of color–surface division, we broke down numerous papers distributed in diaries and gathering procedures. To widen the extent of this paper, we won't limit ourselves just to the specialized appraisal of the researched calculations. Yet, we will likewise endeavor to give a great dialog where the thoughts that developed in the field of color–surface mix in recent decades are deliberately sorted. We will also look at the viable setting of the explored techniques at whatever point such exchange is conceivable.

Likewise, we will put an essential accentuation on the quantitative assessment of the cutting-edge usage in color–surface examination. In such a manner, we will exhibit numerical outcomes by breaking down best-in-class techniques. We will show the conditions and the information utilized as a part of the assessment procedure. This is contained three layers: picture handling (bring down layer), picture examination (center layer), and picture seeing (high layer), as appeared in Fig 1. Picture division is deemed the

initial step and a standout amongst the essential undertakings of picture investigation. Its goal is to separate data (spoken to by information) from a picture through picture division, protest portrayal, and highlight estimation, as shown in Fig 2. The aftereffect of division; has an impressive impact on the exactness of highlight estimation [2]. The computerization of the medicinal picture division is imperative in therapeutic imaging applications. It has discovered wide applications in various regions, for example, determination, restriction of pathology, investigation of anatomical structure, treatment arranging, and PC-coordinated medical procedure. Be that as it may, the inconstancy and the many-sided quality of The human body's anatomical elements brought about medicinal picture division remaining a problematic issue [3].

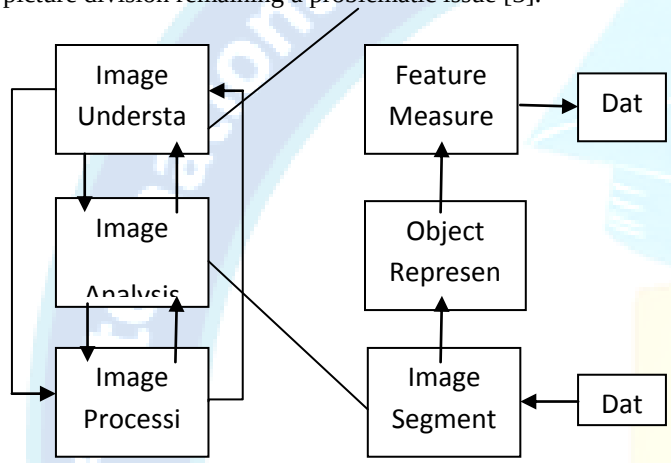

Fig 1: Image designing and image division [2].

Proposed Work:

The target of the research is to accomplish the application of image segmentation by recognizing text as a small object into disjoint groups with performance matching human impressions. This includes executing an unsupervised segmentation of text items on noisy pictures to meet the need for prior ROI assumptions. We need to combine two different approaches. Images are segmented using a technique that connects many processes based on color and texture attributes. First, using group coefficients obtained from the wavelet transform of the picture, arrange determines the textured highlights. Middle sifting is then linked to lessen ambiguities in surface regions close to image object limits from that point on. Finally, computed surface components are related to the space-built slope transition and the ensuing watershed shift to gain underlying segmentation.

The weighted average strategy for local features is used to combine sectioned areas obtained from watershed change with similar highlights.

Image DWT2:

2-D wavelet transform with a single level of discretization

Syntax: [cA,cH,cV,cD]=dwt2(X,'wname')

[cA,cH,cV,cD]=dwt2(X,LoD,Hi D)

[cA] = dwt2(...,'mode',MODE);

Description

The dwt2 function performs the two-dimensional wavelet decomposition over a single level. If you compare the method to wavedec2, that might be more suitable for your requirements. A specific wavelet ('wname'; see filters for more information) or particular wavelet decomposition filters (Lo D and Hi D) that you supply are used to conduct the decomposition.

dwt2(X,'wname') = [cA, cH, cV, CD] The wavelet decomposition of the input matrix X is used to generate the approximations coefficients matrix cA as well as the details coefficients matrices cH, cV, and CD (horizontal, vertical, and diagonal, respectively). Wavelet identification is part of the character vector for "name."

dwt2(X, LoD, Hi D) = [cA, cH, cV, CD] Calculates the two-dimensional wavelet decomposition as displayed above using the wavelet decomposition filters you supply.

The decomposition low-pass filter (Lo D) and high-pass filter (Hi D) are used.

Both Lo Das and Hi D ought to be of the same size.

Let sx = size(X) and if = length of filters. Then, size(cA) = size(cH) = size(cV) = size(cD) = sa, where sa = ceiling(sx/2) and sa = floor((sx+lf-1)/2) for different forms of extension if DWT extensions mode is periodization.

[cA, cH, cV, and CD] = dwt2('mode'', MODE) Using the MODE extensions mode that you choose, wavelet decomposition is calculated.

The extension mode you intend to utilize is included in the character vector MODE.

[cA,cH,cV,cD] = dwt2(x,'db1','mode','sym'); is a suitable application example.

Algorithm:

1. Obtaining Image Data: Analyze natural picture (I) and resize it before selecting its feature. The digital image is saved in binary format. The image is read as the 8-bit integer data and kept in the matrix format. In this code, the pictures are saved by different names, and on inserting a specific term, the image is imported and saved in matrix data.

2. 1 level Image DWT: Accomplish principal level 2D DWT on the image and acquire the estimation segment (A) as the transformed information.

[A DH DV DD]= DWT (I)

The I is the image matrix and passed through 2 dimensional filters acting along x and y direction along the rows and columns. The DWT applies the filtering to create the high pass and low pass frequency components of image. The low pass component are taken as approximated components saved as A and the high frequency components are DH known as horizontal details ,DV are vertical details and DD are the diagonal details.

3. 2 level Image DWT: Perform second level 2d DWT on image as well asacquireestimation segment (A1) of changed information.

[A1 DH1 DV1 DD1]= DWT(A)

The first level DWT gives Components with high as well as low frequencies. Low frequency data are large in amount as per energy density. They are further broken into low and high frequency. To accomplish this task the low frequency part known as approximation component (A) are further passed through the DWT again and the resultant matrix are called as A1 (approx.),DH1,DV1 and DD1 are details part.

4. Approximated Image Reconstruction by 2level IDWT: The resultant approximated part obtained by 2nd level DWT is in double precision format and called as wavelet coefficients. These approximation coefficients are further added with null value of details component to reconstruct the approximated image in the 8 bit integer format. It is taken as the performing the reverse DWT as well as reproduce image by considering just approx. segment as well as stifling the DH, DV and DD point by point segment. The details are removed by suppressing them to zero. In this way, the noisy edges and fine lines are removed the image become clear and the morphology become homogenous.

5. Entropy separating: The entropy values represent the information content at different location of image as per the density of pixel intensity variation. The image entropy based separation of high and low entropy regions are segregated to know the information content. Apply entropy based separation to remake approximated image. These entropy values are taken as features to segment the image in different regions.

$$H = -\sum_{i=0}^{255} p_i log_2 p_i$$

H=entropy of image

p_i=probability of occurrence of a symbol i.

6. Remove minuteunwanted objects: The image consists of dots or circular shapes as openings. The small openings are removed out to minimize ambiguity during the segmentation process. The removal of the undesirable openings having size lower than 100 pixels is performed in this code for making higher accuracy in segmentation.

7. Morphological Processing: The morphological process is performed to bring changes in the shape of image objects. This is performed by two different schemes known as erosion and dialation. The erosion removes the boundary pixels of image objects. It helps in separating two objects, which are closer. The dilation process adds on pixel at the boundary of object. It helps in aggregating object having very thin border lines.

E is a Euclidean space or an integer grid, B is a structural element that can be seen as a subset of Rd, and I is a binary image in E.The equation for dividing A by B is

$$A \oplus B = \bigcup_{b \in B} A_b$$

where A is translated by b to form A_b.

Commutative dilation is also provided by

$$A \oplus B = B \oplus A = \bigcup_{a \in A} B_a$$

8. Texture Masking: Textures are the patterns of similar type associated with variation of color and shades and geometrical shape. The natural imagesconsist of objects of specific color, shade and pattern and they may be easily separated out as different textures. The masking of the texture 1 as well as texture 2 is applied to create texture dependent divided image. It helped in separating out the background and the outside boundary of image object in one texture..

9. Edge Detection: Edges are the boundaries, line, borders associated with the borders of any object in the image. The edges are taken as the abrupt changes in the shades of the image objects. The Sobel filtering are applied first highlight edge borders, and then compute gradient magnitude to provide a picture with one at edges as well as zero for inside areas.

$$f(x) = \frac{I_r - I_l}{2}\left(\text{erf}\left(\frac{x}{\sqrt{2\sigma}}\right) + 1\right) + I_l$$

I_L and I_r are image intensity at left and right side. σ is the blur scale of image.

10. Edge Erosion: The unwanted image objects are small lines in the image and creates problem in separating of major segments. Hence, the erosion of such objects is having less than disk size 4 pixel done.Let A be a binary image in E, and let E be an integer grid or a Euclidean space. The structural element B defines the binary image A's degradation by:

$$A \ominus B = \{z \in E | B_z \subseteq A\}$$

where Bz is the vector z's translation of B, i.e..

$$B_z = \{b + z | b \in B\}, \forall z \in E$$

11. Edge widening: After removal of small and thin boundaries. The leftover boundaries are widen to clearly segment out the important image objects. Implement dilation to objects with a disc size of less than 4 pixels as well as rebuild the image.

12.Thresholding: Thresholding operation is applied to choose segmented bounds of strong luminous intensity on edge elements

13. Watershed Transform: The image consists of different intensities in the gray levels. The higher and lower intensity regions are considered as the maxima and minima location. It is called as peaks and valleys. The watershed transform

envisage the distribution of local minima with respect to intensity. It helps in the image objects with separated boundaries.
Let f ∈(D) have $\{mk\}_{\epsilon}$ minimums for some I index set. The collection of sites x∈D that are topographically closer to m_j than to any other regional minimum is known as the catchment basin (mi) of a minimum mi.

$$CB(m_i) = \{x \in D\} \forall jl\{i\}: f(m_i) + T_f(x, m_i) < f(m_i) + T_f(x, m_i) \text{ (i)}$$

The collection of sites in f's watershed that are not part of any catchment basin:

$$Watershed = (f) = D \cap (\bigcup_{i \in l} CB(m_i))^c$$

Let W stand for any label. A mapping is the watershed transform of f,λ: $D \rightarrow I \cup \{W\}$, such that $\lambda p = iifp \in (mi)$, and $\lambda p = Wifp \in Watershed$.
The watershed transformof *f*assigns labels to the points of D in such a way that *I* distinct catchment basins are each given a unique label and (ii) a special label W is given to every point of the watershed of *f*.

14. Superimposing of segmented image objects: All the segmented image objects obtained by entropy features, edge detection and watershed transforms are finally superimposed to account the final segmented image. Overlay a watershed-change connected, divided image made up of surface-based pieces.

$$Isuperimposed = I1 \cup I_2 \cup I_3$$

15. The segments are termed as the objects representing different kind of features. Extraction of the image features using entropy, wavelet coefficient and textures data for all the images.

16. Development of the training and testing data using the features of segmented image to develop a classification model using the ANFIS algorithm.

17. The clustering scheme is applied to generate a fuzzy inference system consisting of membership functions and fuzzy rules. The ANN learning scheme is applied to reform the developed fuzzy system using clustering.
A neuron with label j that receives input pj(t) from preceding neurons is made up of the following components:
An activation function that, given an activation aj, calculates the new activation at a particular point in time t+1 (t)., a discrete-time parameter, and an optional threshold j that is fixed unless learning changes it (t). the net input of the resulting relation, pj (t),

$$a_j(t+1) = f\left(a_j(t), p_j(t), \theta_j\right),$$

and a function for calculating the output of activation

$$o_j(t) = f_{out}(a_j(t))$$

The *propagation function* computes the *input* to the neuron from the outputs and typically has the form:

$$p_j(t) = \sum_i o_i(t) w_{ij}$$

A bias term can be added, changing the form to the following:

$$p_j(t) = \sum_i o_i(t) w_{ij} + w_{oj}$$

ω_{0j} is the bias.

4. Result and Discussion:
To test our text extraction technique from complex degraded images, we have used random images from various Internet sources containing a wide range of datasets with different background illumination, color, image quality, font sizes, orientation, etc. The proposed algorithm is implemented on the MATLAB R2018a platform. All the experiments are performed on a standard computer (Intel Core i7- 4770, 3.40 GHz CPU, 4 GB RAM, and Windows 8.1 Pro, 64-bit OS). Some of the complex degraded images and the extracted texts are shown in Fig 3 to Fig 6.

Fig. 3. Original Image

Fig. 4. Image with the Gaussian noise

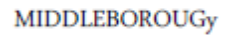

Fig. 5. Detected Object

Fig. 6. Image with salt and pepper noise

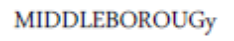

Fig. 7. Detected Object

The proposed method has its advantages and disadvantages. It works appropriately on many datasets, such as screenshots, invoices, documents, hoarding banners, etc. However, it might not function accurately on complex degraded images with poor image quality and uneven edges. A comparative analysis of accuracy for various embodiments is illustrated in figures 7, 8.

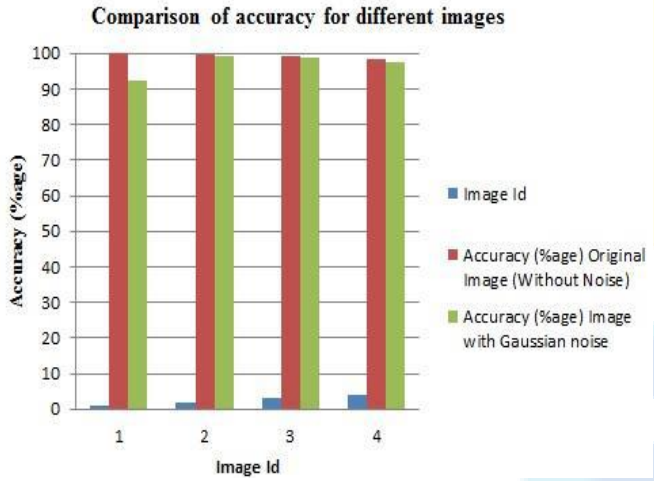

Fig. 8. Comparison of accuracy for different images with Gaussian noise.

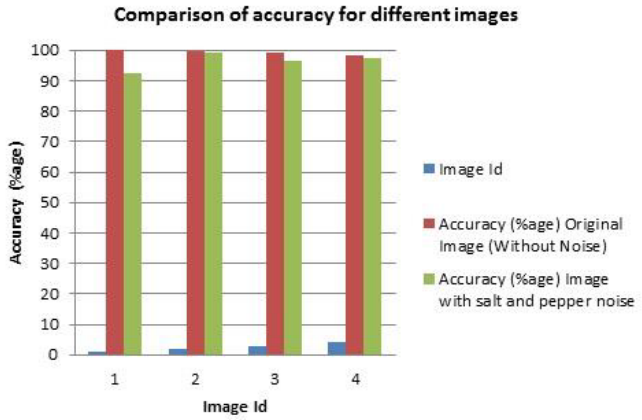

Fig. 9. Comparison of accuracy for different images with salt and pepper noise

5. Conclusion:

The text object detection model's performance also depends on the texture and color intensity in images. Previously the researchers had used a single hyper-tuned configuration to develop the model, but it impacted the model's scalability on various datasets. Our proposed hybrid technique has shown promising accuracy improvement, especially in the text under small object detection. The watershed and entropy features configuration can be obtained over the fitted dataset using the wavelet. Further, we can apply better morphological methods to enhance the model's efficiency. The proposed model has achieved improvement in accuracy score.

References

[1] P. Viola and M. Jones, "Rapid object detection using a boostedcascade of simple features," in Proceedings of the IEEE ComputerSociety Conference on Computer Vision and Pattern Recognition,pp. I511–I518, Kauai, Hawaii, USA, December 2001.

[2] R. Lienhart and J.Maydt, "An extended set of Haar-like featuresfor rapid object detection," in Proceedings of the InternationalConference on Image Processing (ICIP '02), pp. I/900–I/903,Rochester, NY, USA, September 2002.

[3] P. Viola, J. C. Platt, and C. Zhang, "Multiple Instance boostingfor object detection," in Proceedings of the Annual Conferenceon Neural Information Processing Systems (NIPS '05), vol. 18,pp. 1417–1424, Vancouver, BritishColumbia, Canada,December2005.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich featurehierarchies for accurate object detection and semanticsegmentation," in Proceedings of the 27th IEEE Conference onComputer Vision and Pattern Recognition (CVPR '14), pp. 580–587, Columbus, Ohio, USA, June 2014.

[5] R. Girshick, "Fast R-CNN," in Proceedings of the 15th IEEEInternational Conference on Computer Vision (ICCV '15), pp.1440–1448, Santiago, Chile, December 2015.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "FasterR-CNN: towardsreal-time object detectionwith region proposal

networks," IEEETransactions on Pattern Analysis and Machine Intelligence, vol.39, no. 6, pp. 1137–1149, 2017.
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid poolingin deep convolutional networks for visual recognition," IEEETransactions on Pattern Analysis and Machine Intelligence, vol.37, no. 9, pp. 1904–1916, 2015.
[8] J. F. Dai, Y. Li, K. M. He et al., "R-FCN: Object Detection viaRegion-based Fully," in Proceedings of the 30th Conference onNeural Information Processing Systems (NIPS 2016), Barcelona,Spain, 2016.
[9] M. Everingham, L. van Gool, C. K. I.Williams, J.Winn, and A.Zisserman, "The pascal visual object classes (VOC) challenge,"International Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010.
[10] Y. Ren, C. Zhu, and S. Xiao, "Small object detection in opticalremote sensing images via modified faster R-CNN," AppliedSciences, vol. 8, no. 5, article 813, 2018.
[11] L. Itti, C.Koch, and E.Niebur, "Amodel of saliency-based visualattention for rapid scene analysis," IEEE Transactions on PatternAnalysisandMachine Intelligence, vol. 20, no. 11, pp. 1254–1259,1998.
[12] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," IEEETransactions on Pattern Analysis and Machine Intelligence, vol.37, no. 3, pp. 569–582, 2015.
[13] C. Chen, M. Y. Liu, O. Tuzel et al., "R-CNN for small objectdetection," in Asian Conference on Computer Vision, vol. 10115of Lecture Notes in Computer Science, pp. 214–230, 2016.
[14] J. S. Lim and W. H. Kim, "Detection of multiple humans usingmotion information and adaboost algorithm based onHarr-likefeatures," International Journal of Hybrid Information Technology,vol. 5, no. 2, pp. 243–248, 2012.
[15] R. P.Yadav, V. Senthamilarasu, K. Kutty, and S. P. Ugale,"Implementation of Robust HOG-SVM based Pedestrian Classification,"International Journal of Computer Applications, vol.114, no. 19, pp. 10–16, 2015.
[16] L. Hou, W. Wan, K.-H. Lee, J.-N. Hwang, G. Okopal, and J.Pitton, "Robust Human Tracking Based on DPM ConstrainedMultiple-Kernel from a Moving Camera," Journal of SignalProcessing Systems, vol. 86, no. 1, pp. 27–39, 2017.
[17] A. Ali and M. A. Bayoumi, "Towards real-time DPM objectdetector for driver assistance," in Proceedings of the 23rd IEEEInternational Conference on Image Processing, ICIP 2016, pp.3842–3846, Phoenix,Ariz, USA, September 2016.
[18] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Insideoutsidenet: Detecting objects in context with skip pooling andrecurrent neural networks," in Proceedings of the 2016 IEEEConference on Computer Vision and Pattern Recognition, CVPR2016, pp. 2874–2883, LasVegas,Nev, USA, July 2016.
[19] T.Kong, A.Yao, Y. Chen, and F. Sun, "HyperNet: towards accurateregion proposal generation and joint object detection," in Proceedings of the 2016 IEEE Conference on ComputerVisionandPattern Recognition (CVPR), pp. 845–853, LasVegas,Nev,USA,June 2016.
[20] F. Yang, W. Choi, and Y. Lin, "Exploit all the layers: fast andaccurateCNNobjectdetectorwith scale dependent pooling and cascaded rejection classifiers," in Proceedings of the 2016 IEEEConference on Computer Vision and Pattern Recognition, CVPR2016, pp. 2129–2137, Las Vegas, Nev, USA, July 2016.
[21] Liu, X.; Liang, D.; Yan, S.; Chen, D.; Qiao, Y.; and Yan,J. 2018. FOTS: Fast oriented text spotting with a unified network. In IEEE Conf. Comp. Vis. Patt. Recognit. (CVPR),5676–5685.
[22] Xing, L.; Tian, Z.; Huang, W.; and Scott, M. R. 2019. Convolutional character networks. In Proceedings of the IEEEInternational Conference on Computer Vision, 9126–9136.
[23] Wang, Z.; Zheng, L.; Li, Y.; and Wang, S. 2019c. LinkagebasedFace Clustering via Graph Convolution Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[24] Rijcken Emil;KaymakUzay ;ScheepersFloortje,;Mosteiro Pablo;ZervanouKalliopi; Spruit Marco."Topic Modeling for Interpretable Text Classification From EHRs "Frontiers in Big Data vol 5 ,2022.