ISSN: 2093-5374 / E-ISSN: 2093-5382

Drone Image Localization by Faster R-CNN Algorithm and Detection Accuracy

Maysoon Khazaal Abbas Maaroof[1*], and Med Salim Bouhlel[2]

[1*]National School of Electronics and Telecommunications, Information Technology, University of Sfax, Tunisia; Assistant Professor Information Technology, Basic Education College, University of Babylon, Babil, Iraq. basic.maysoon.maroof@uobabylon.edu.iq, https://orcid.org/0000-0002-4035-0537

[2]Smart Systems for Engineering & E-health based on Technologies of Image & Telecommunications (SETIT), ISBS, Sfax University, Tunisia. medsalim.bouhlel@isbs.usf.tn, https://orcid.org/0000-0003-2952-3967

Received: October 25, 2024; Revised: December 11, 2024; Accepted: January 15, 2025; Published: March 31, 2025

Abstract

Since unmanned aerial vehicles (UAVs) provide real-time monitoring of vast areas, their rapid development has been crucial to the advancement of surveillance applications. However, in the face of complex environments, present surveillance systems frequently suffer from an initial lack of efficiency, scalability, and adaptability. In order to detect and track any security threats in real time, this study aims to create a unique AI-based aerial surveillance framework that makes use of CNNs and Fast R-CNNs. It trains and validates object identification models using publicly accessible UAV datasets in relation to important parameters like robustness, processing speed, and accuracy. The suggested framework for object detection using augmented intelligence thus applies to contemporary surveillance systems, which are designed to be reliable, resilient, and able to effectively satisfy contemporary security requirements. This study presents a brand-new, incredibly effective Faster R-CNN created especially to tackle the difficult object placement issue in aerial photos. For pinpointing the precise location of things of interest, the algorithm works incredibly well. The average accuracy has increased significantly to above 70%, according to the results. With an F1-score of 92.7%, the Fast R-CNN model achieved precision and recall scores of 93.1% and 92.4%, respectively, while still performing within the average of 94.7%.

Keywords: Unmanned Aerial Vehicles, CNN, Faster R-CNN, Accurate Object Detection, and Image-Based Localization.

1 Introduction

The accuracy of the localizer is paramount to the accuracy of the application that utilizes it. This is used for monitoring, mapping, environmental analysis, and structural assessment. In these areas of research that are well-established, drones and machine-learning-based imaging have each contributed to the development of the capacity of UAVs to utilize high-resolution datasets for image processing (Quamar et al., 2023; Behera et al., 2023; Muzammul et al., 2024). The Faster R-CNN algorithm was intended

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 16, number: 1 (March), pp. 172-189. DOI*: 10.58346/JOWUA.2025.I1.010*

*Corresponding author: National School of Electronics and Telecommunications, Information Technology, University of Sfax, Tunisia; Assistant Professor Information Technology, Basic Education College, University of Babylon, Babil, Iraq.

primarily to have a high degree of accuracy with the added benefit of being simple to detect objects (Abbasi et al., 2023). The beneficial aspect of pre-trained transfer learning models is that they require a smaller training dataset than deep networks that are created from scratch, this provides a faster rate of convergence, (Xiao et al., 2020; Pazhani & Vasanthanayaki, 2022; Wan & Goudos, 2020; Ghazi et al., 2023; Qusef et al., 2023).

This research has proposed a use of the Faster R-CNN algorithm for the drone's image location. The Faster R-CNN framework for localization combines the different components of subtypes like feature extractors, RPNs, and classification, and regression tasks associated with R-CNN boxes. The primary goal of the endeavor is to increase the fidelity of the 2D predictions of bounding boxes in test images (Seo et al., 2022; Ali et al., 2021; Ali et al., 2021; Ghazi et al., 2021). It attempts to answer the question of how to effectively train and fine-tune a backbone-based Faster R-CNN model in order to improve the performance of localization, specifically for aerial images. Ultimately, in response to a primary goal, a powerful digital framework must be built that can incorporate artificial intelligence into multiple real-time, image-based applications (Devi et al., 2016). The ultimate goal is to ensure that it's possible to observe real-time outdoor activities via deep learning models (Yan et al., 2021; Du et al., 2022; Noori et al., 2019). Acknowledged that the current methods have a lack of optimal environmental awareness in real-world environments- especially in response to unexpected changes in the outdoors. Additionally, the existing models for feature extraction are having difficulty recognizing small objects because of the limited response capacity of the deeper networks (Placed et al., 2023; Tsintotas et al., 2022; Ullah et al., 2024; Álvarez et al., 2020; Mitra et al., 2021; Mahmuddin & Al-dawoodi, 2017; Al-Dawoodi & Mahmuddin, 2017). This paper addresses the existing shortfall in the literature by suggesting new Faster R-CNN models that are appropriate for the modern dataset.

1) Significance and Background

The rapid-evolving industrial technologies, especially UAVs that are unmanned, have helped to acquire high-resolution images of various areas of the subject. This capacity is utilized in multiple fields, including agriculture, environmental conservation, disaster management, and military surveillance. In the context of micro grids and their association with advanced measurement infrastructure (AMI), drones have a significant role in the constant collection of images that are important for scientific research and the evaluation of systems (Kleinschroth et al., 2022; García-Fernández et al., 2021; Papić et al., 2021). The accurate identification of a place in an image increases the effectiveness of these technologies. The primary goal of image localizing is to determine the spatial association between an observed object in an image and its counterpart in the real world (Liu et al., 2020; Liu et al., 2024). The importance of exact image placement is recognized across the board, this has been utilized in multiple areas including agricultural research that is advanced, human-computer interaction, manufacturing processes that are intelligent, traffic monitoring, and social media platforms (Zhuang et al., 2021; Ma et al., 2021).

The existing body of literature regarding the infrastructure associated with (IoT) is primarily limited to the study of how images are located using drones. This prohibition may come from the addition of multiple scholars of different methods in order to enhance the process of finding drones via images. These approaches typically employ increasingly complex models that explore the association between image-based localization and larger fields of computer vision and analysis (Gupta & Fernando, 2022; Chakravarthy et al., 2022). Recently, the development of various methods for locating objects has increased significantly, this is primarily due to the critical importance of the technology to enhance object-based technologies, including the detection and tracking of objects. However, a significant shortfall remains in the systematic expansion of information regarding object position to the more specific domain of image location within the context of drones (Panigrahi & Bisoy, 2022; Du et al., 2021).

2) Goals of Research

The advances in the detection and segmentation of objects have been primarily caused by the worldwide spread of convolutional neural networks. These principles are also applicable to the analysis of drone videos. However, certain methods of these disciplines require additional expansion in order to be relevant to the purposes of this research. This research aims to assess the effectiveness of Fast R-CNN algorithm in the context of drone image recognition, with a specific interest in a singular target that is of concern. The primary goals of this research are apparent from the numerous subordinate goals: - To create a model of a signal and training network that is dedicated to the location of the drone's image, this will enhance the accuracy of detection and the speed of processing via the Faster R-CNN framework (Shinan et al., 2024). To determine how effective the Faster R-CNN algorithm is at finding drones in comparison to other approaches in the field. To enhance and improve the Faster R-CNN method by including extra layers that are supplementary, this will lead to a higher degree of operational success. To assess the effects of the developed model's accuracy and efficiency in detection. To augment the field of drone-based photo analysis and propose alternative methods of imaging (Shiu et al., 2023).

The Faster R-CNN method of localization is employed. To authenticate the experimental findings through commonly embraced benchmarks as well as various performance metrics. Image localizing is the process of analyzing an image in order to determine the location of an object that is recognized. The information resulting from the location's data includes critical information, such as the object's designation, as well as its original and final locations.

2 Literature Review

Drone technology has experienced significant success in multiple disciplines because of its capacity to transport aerial images and other data. Numerous case studies have been conducted by scientists in multiple disciplines that utilize drones. The evolution of drones has led to a transition from simple aircraft to complex autonomous systems, this has presented a new opportunity for scientists to alter the traditional way of doing things (Elmeseiry et al., 2021). Also, drones are referred to as Extended Unmanned Aerial Vehicles (UAVs), and have been utilized in numerous scientific investigations since the 80's (Kavitha, 2024; Kesavaraj et al., 2023; Almusawi et al., 2024; Trivedi et al., 2023). The majority of the uses have been dedicated to military and agricultural fields. However, the utilization of drones and other aerial vehicles for the collection of data is still modest in the contemporary study of natural sciences and urbanism, (Labib et al., 2021; Dronova et al., 2021; Kim et al., 2022). The term 'drone' was first documented in scientific literature as far back as 1997. As smaller drones have become more frequented and budgeted, their capacity to conduct practical research has increased greatly. This evolution has helped researchers to reduce the necessary time and expense needed to research, this effectively lowers the cost and time needed (Jońca et al., 2022; Butcher et al., 2021; Schäffer et al., 2021).

Object detection is a significant problem in the field of visual perception, the goal is to locate specific objects within images with the greatest degree of precision. This obstacle can be divided into two primary categories: the difficulty associated with locating objects and the difficulty associated with recognizing objects (Madhan & Shanmugapriya, 2024; Zou et al., 2023; Wu et al., 2020). Object localization is concerned with finding the specific spatial regions that an object is located in within an image, whereas object recognition is concerned with categorizing the objects that are located in these specific regions as belonging to their respective classes. In this context, the focus is on recognizing objects, which is also called image recognition. Contrasted with the recognition of objects, the location of images provides additional information about the exact position of the object that is enhanced However, recent advances in deep learning have led to significant enhancements in this area. These enhancements facilitate a categorization of detection methods into three different classes based on their CNN configurations (Liu et al., 2020).

Recent research has also demonstrated that UAVs are increasingly being employed in environmental monitoring, disaster response, and agricultural surveying. This suggests that they can produce high quality data in a cost-effective manner (Maaroof & Bouhlel, 2024). Also, innovative methods of UAV image processing have increased the accuracy and reliability of surveillance data significantly (Alanazi, 2023). For instance, recent advances have led to the creation of advanced image enhancements that improve the clarity and resolution of images recorded by UAVs in various lighting and weather conditions. These enhancements are crucial to apps that require precise analysis of images, such as land mapping and crowd surveying (Hamouda et al., 2020).

1) The Technology of Drones and their Utilization in Agriculture, Business, and Science

Drones are increasingly viewed as a transformational technology due to their diverse applications in multiple disciplines, including environmental monitoring, disaster response, surveillance, agriculture, public safety, and weather forecasting. One unique characteristic of drones is their ability to function as data collectors that are airborne. armed with cutting-edge data acquisition systems that are multi-sensory, they can acquire complex, multi-dimensional data. This information is then evaluated and summarized through machine learning and artificial intelligence processes. These abilities have a significant capacity to evaluate data through deep learning methods, which can be employed to understand, predict, and identify various objects and phenomena that have an association with a variety of applications.

Additionally, the effectiveness of drones in operation is dependent on the on-board computers, which have components like central processors and graphics processors, these components affect the entire functionality of the drones.

The fidelity of the position is increased, and the capabilities of the robot are augmented, this results in an increase in machine learning. The proposed methodology is simple and direct, and would allow a rapid alteration of the position of a robotic machine during the setup or re-positioning of production. a deep learning approach that is employed to augment the fidelity of the articulated robotic system. Around 300 additional iterations later, the placement accuracy increased significantly (AbdAli & Maaroof, 2022).

The database server is located at the central control room. This configuration facilitates increased data processing capacity and a simplified approach to dealing with real time weather information and flight paths of drones (Hussein et al., 2024). The increase in prevalence of drones is primarily caused by the computational resources, particularly multiple cores in a single processor and artificial intelligence software that includes machine learning and computer vision. Several important focal points can be identified that are essential to the survival of humans, as a result of the numerous applications that have been created. Initially, these apps were primarily focused on collecting geographic and physical data, this was then altered to be more beneficial in the utilization and analysis of data. A crucial component of these apps is the ability to differentiate and describe the different shapes, textures, and colors that images have. To enhance the processing of images and the classification of images, a convolutional neural network-based machine learning method is employed. The rich, detailed description of the procedure's visual effects can be employed in various visual-based business and operational scenarios, including the task of image search and placement. The application of image-based localization is primarily defined by the process of determining the position of a specific query image, whether in an interior or exterior setting, from a large amount of previously-unknowledgeable information regarding the query image.

2) Object Recognition Algorithms

R-CNN is significant towards the traditional sliding window method, it first utilized convolutional neural networks (CNNs) to create hypotheses regarding areas. After this innovation, the number of object detection methods increased significantly. Many of the methods were based on deep learning principles. In 2016, the framework called "YOU only look once" (YOLO) was introduced, this framework facilitated the detection of objects in real time by executing a single forward propagation on an image. This procedure simultaneously generates boxes that are predicted with associated confidence values, although without explicitly considering the goals of object classification or placement. The DSSD framework contains the concept of anchors, along with documented scales and aspect ratios, this increases the fidelity of the small-scale object detection and improves the average precision of obstacles associated with the average. However, its effectiveness is diminished when dealing with objects that are densely located, this leads to a decrease in the accuracy of detection. This approach takes a different route by segmenting the object detection boxes into different portions and using the smaller portion to represent the division of the object. This model seems to augment the capabilities of object detection, it also grants the basis for additional investigation. When contrasted with the R-FCN architecture against the F-CNN architecture, it is apparent that the method for collecting data in the R-FCN is more extensive, this leads to a higher degree of success in the detection of objects.

The algorithm mentioned is concerned with calculating the average of two adjacent boxes based on the average production of the network for each cell in the grid. Within this context, the information regarding the boundaries of objects is exclusively transmitted through the adjacent cells in the grid; however, this results in the classification of object information near the boundary to be ambiguous, which in turn prevents the detection of smaller objects. To bypass these obstacles, the Single Shot Detector (SSD) method has been employed to augment the computational efficiency. This method operates by dividing the image size according to a sliding window step, thereby influencing the final output dimensions. The distance or confidence of object in its current position, this further enhances the speed of detection. It's crucial to recognize that the algorithm's capacity to deal with real-time information directly affects its legitimacy in the detection process. The choice of object detection methods is often based on the specific attributes and location of the objects in question. When encountering images that are based on content and have the same type of problems the SPP model has a marginal effect on the Faster R-CNN in terms of training and classification performance. The sliding window method appears to have a greater propensity to recognize regional characteristics. The proposed algorithm solves issues associated with both immediate performance and precision when considering the magnitude of the variations and the number of applications in diverse situations. This discussion concerns a synthesis of high accuracy and increased practicality via the utilization of various fine-tuning methods. Notably, the implementation of the Fast R-CNN reduces the amount of time needed to process images during the optimization of network models on GPUs. Recent studies have shifted towards a multitasking approach, this is demonstrated by the MRFCN C RNN method of detection for drones. Despite improvements in recognizing the value of top-down approaches, several questions that are still unanswered remain. These obstacles are primarily caused by two sources: the limited capabilities of the network in producing high-resolution images, and the intrinsic nature of the images themselves. A comprehensive examination of these aspects demonstrates the value and practical significance of this research for future endeavors. A significant shortcoming that was recognized in this research is that, in comparison to today's deep learning networks dedicated to object detection, there is a significant difference in performance. Additionally, the lack of comprehensive reporting on the results and effects of the study that should be the subject of subsequent studies has led to a lack of focus on the results and effects of the study that are of greatest importance.

3 Methodology

Recently, a new method, Fast R-CNN, was commended for its effective accuracy and wide generalizability regarding large datasets pertaining to the detection of objects, this is an important aspect of place recognition. In this section, we'll discuss the specifics of Fast R-CNN. The design of Faster R-CNN is primarily composed of two parts: RPN and Fast R-CNN. RPN is responsible for creating an N number of boxes that hold an image's anchor points at a particular position with regard to their scales and aspect ratios. A softmax layer then calculates the classification probability for these boxes with anchor. It also predicts the regression of the bounding box, which is the difference between the anchor boxes and the ground-truth boxes. The deep residual network receives input in a variety of dimensions and employs a set of convolutional layers. This is followed by a layer of regions of interest pooling that facilitates the preparation of features for the following stages. Within the architecture, the Fast R-CNN module is responsible for receiving the pre-designed feature maps that originate from the RoI layer, these maps are then passed through a fully connected layer. The RoI's features are finally transferred to two separate output layers that are intended for softmax classification and box regression. In the experiment, this deep learning platform was employed using graphics cards with 12GB of random access memory. Other methods, besides the current approaches, could be beneficial in evaluating the outputs of localization in comparison to the current method. The general consensus is that the primary factor responsible for the degree to which Faster R-CNN is adopted in any particular application is the quality and relevance of the data.

Accuracy is the frequency of correct classifications of all samples in the total. Mathematically, this is expressed as follows in equation 1:

$$Accuracy = \frac{TP+PN}{TP+PN+FP+FN} \tag{1}$$

TP is the number of true positive results, TN is the number of true negative results, FP is the number of false positive results, and FN is the number of false negative results in equation 2.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Precision is the percentage of true positive detections that are derived from all of the positive predictions made by the model. It's characterized by equation 3:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Recall is the percentage of true positive results that were revealed among all of the actual positive results. It is calculated as follows:

The F1-score is the average of the precision and recall values weighted by their respective weights. (As a result, this rating incorporates both false positives and false negatives.) The F1 rating is calculated as shows in equation 4:

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (4)$$

The IoU is a metric that calculates the degree to which an object is located. That is to say, it calculates the degree to which the predicted bounding box and the ground truth bounding box intersect, using the following equation 5:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad (5)$$

The area of overlap is the space between the predicted bounding box and the ground truth bounding box, while the area of union is the space between the two boxes.

A comprehensive collection of training data that is accurate describes the degree to which both the image's features and the label's location have an effect on the training and testing stages. A 1.25 kilogrammes aerial platform with four brushless rotors, a digital camera, and a light controller were all considered as references for the implementation of the system. Localization information was gathered over the course of the flight, the majority of it was obtained from the locations indicated by GPS signals. The labels were then assigned using validation methods that took into account the positional information. Images that were captured had an average of 243 objects per image, each image contained multiple objects that contributed to the total of 882 boxes that were bound to the system. Additionally, the positions in which the respective objects were observed were documented by the on-board cameras. Additionally, to assess the effectiveness of the detector in terms of efficiency, 261 images were captured that are valid. In the tested configuration, if the image is sized at 800x600 pixels, the ground resolution per pixel is 0.03 meters for the original. Other than the position data, the angles of yaw of the quadcopter and its distance from the ground were recorded on a regular basis. To assess the consistency, the time stamps of each camera's images were compared to the position data. The images with colors are then transferred to software that processes images, this software will make the background of the images identical.

1) Faster R-CNN: Introduction

An improved deep learning platform with an anchor in Region Proposal Networks is the fundamental component of faster Region-based Convolutional Neural Networks (R-CNN). A highly integrated mix of the Region of Interest and CNN capabilities is encouraged by the design. A key element of this strategy in a real-world system is how effectively the RPN can suggest anchor regions, which lowers the number of selections required. The state-of-the-art in picture localization and object detection has been achieved using this very dual-stage procedure. Functioning more quickly, R-CNN uses two main stages. During the first phase, it acquires ideas of potential targets, during the second phase, it processes the input to the network that aims to improve the accuracy of detection. The network is composed of three essential components. First, the CNN's backbone processes the input image to generate a feature set from it. This is followed by the application of RPN which collects and describes the boxes that are picked up and outlined by proposal clusters. Ultimately, a regression layer that processes all regions of concern over the dataset in order to define systematically the boxes of various classes. The methodology of this paper is, therefore, a hybrid of image classification and detection with the primary claim in the document being that Faster R-CNN is appropriate for fast search management. The procedure of deep learning for Faster R-CNN involves the use of aerial images that are captured by drones in order to ensure comprehensive oversight of the procedure, this procedure is characterized by its low cost, low complexity, low energy consumption, and also by its low risk management.

However, the process of collecting and preprocessing drone images should be conducted in a practical framework. As a result, this process necessitates the collection of both training and validation images in real-world conditions. The article focuses on the crucial importance of image datasets derived from real-world applications of drones. Numerous datasets are available that are pertinent to the use of drones. For instance, one dataset associated with this project is the Object Detection in Aerial Images dataset. Another well-known public dataset is a subset of a larger dataset that is considered the first large public image repository that is specifically designed to be used with remotely piloted drones in the outdoors. These datasets typically have a simple visual presentation and a limited number of scenes that are most common in regards to their target type. Additionally, this intended undershooting can enhance the machine's visual acuity, but it also serves as a benefit to the development of an algorithm that is practical in real-world conditions, which is of paramount importance in the rapidly evolving 360° scenario. For this, two publicly accessible datasets that detect drones and a private dataset are employed, all of them are tested in different conditions. These datasets are comprised of images gathered by high-end commercial drones that are capable of flying at a maximum speed of 40 km/h. Occupied a testing ground that represented a advanced landscape of similarity reference, honoring high-fidelity virtual travel rules. It employs a on-board RGB camera that captures images at a 1280 × 720 resolution. In this regard, the private datasets have different objects and poses that allow for the artificial variation of objects (e.g., different weather conditions) and changes in appearance (day and night). This is intended to attempt to create a rapid, accurate R-CNN model specifically for the task of estimating the umbral threshold. Each image in the dataset was scrutinized, this class was associated with the car's front end and possessed a well-formatted appearance. To circumvent the overfitting associated with the learning process of the model based on these labeled datasets, both normalization and image augmentations were employed. Additionally, to create the corresponding image boxes for the test that follows, the point cloud data was gathered and preprocessed in order to be useful for the test. Overfitting is typically caused by a small training dataset. To circumvent it, the data would be normally trained with a median interval of half the length of the window, then data augmentation would be implemented to increase the robustness of the learning process. In this study, methods including the normalized Laplacian of Gaussian's method of enhancement that is uniform along with various data augmentation methods like rotation, flipping, shearing, and zooming were employed. The datasets that are trained, tested, and validation should be selected with care in order to ensure that the neural network accurately represents the space of the problem and provides effective decisions. As a result, the drawing of results from test datasets to practical applications should be accompanied by a bias reduction and an awareness of the presence of bias. Additionally, the background and speed vectors should be perpendicular to the operational environments.

4 Experimental Results

The investigation utilized a large dataset that was entirely composed of drone-captured images, with GPS information associated with each image that was known. This dataset contains images that were captured from various perspectives and distances, they traversed different paths and elevations.

1) Overview of the Images

The images depict an outdoor setting, possibly a public area or an event space with a significant number of people present. The images are annotated with bounding boxes and labels identifying detected objects. The objects identified by faster R-CNN algorithm include:

- Persons (labeled as "person")

- Cars (labeled as "car")
- Umbrellas (labeled as "umbrella")

a. Detection Accuracy and Observations

Figure 1: Explain a Large Number of People in a Crowded Area, Particularly in Front of a Large Blue Structure

- The model has identified a large number of people in a crowded area, particularly in front of a large blue structure.
- Most detected persons seem to be accurately labeled, with bounding boxes fitting the individuals reasonably well.
- A fire hydrant is also detected in the bottom left corner, which appears to be correctly identified, as shown in figure 1.

Figure 2: Show Relaxed Outdoor Setting with Fewer People of Faster R-CNN Dataset Detection Effect

- This image shows a more relaxed outdoor setting with fewer people.
- The model accurately detects the individuals present, with bounding boxes properly aligned.
- Umbrellas are detected in the scene, and cars in the background are correctly labeled.

- The detections appear accurate with minimal overlap, and the bounding boxes are well-fitted to the objects, as shown in figure 2.

Figure 3: The Bounding Boxes Accurately Encapsulate the Objects

- Similar to Image 1, this image captures a crowded scene in front of a blue building.
- People are densely packed, and the model detects most individuals effectively.
- There are no visible false positives in this image, and the bounding boxes seem to accurately encapsulate the objects of interest, as shown in figure 3.

Figure 4: Umbrella, People and Car Image Bounding Boxes in Different Angle

- This image is similar to Image 2 but from a different angle.
- The detection of people is consistent, and the model also identifies other objects such as cars and umbrellas correctly.
- The bounding boxes and labels are correctly applied, showing that the model is effective even from different perspectives, as shown in figure 4.
- In figure 1 and figure 2 decreased the threshold of probability to 0.6 while in figure 3 and 4 put it 0.8 to show the boundary boxes around the objects.

An analysis of more than 50% overlap between anchors and objects in the dataset, where more than half of the cases consolidated over both dimensions equaled or were larger than 128 pixels for lengths and widths. Several of approximately 250,000 positive and 1,000,000 negative bounding boxes were identified for possible location within the dataset. While segmenting the data into training and testing, it was noted that the number of average training images changes, and this was over about 437 iterations. This saved the model weights for each epoch with an approximate CPU cost of 2 hours. The evaluation of the algorithm showed a quite appropriate success rate, especially for small objects. Although in flat areas and urban scenes the algorithm began to give less correct results, at higher degrees of truth in the data, the more truth was detected in the data, so the better it became. One of the key metrics for evaluating the model in the Localization-Drone learning was that of matching images to a coordinate class above a certain level of accuracy of 0.7. Performance metrics were that the Fast R-CNN model attained precisions of 94.7%, with precisions and recalls set at 93.1% and 92.4% to give an F-score of 92.7%. These are the details of the metrics, as shown in Table 1, bringing out how detection accuracy received a big boost with the faster region proposal and classification processes.

The IoU metric was calculated over the predicted bounding boxes by the Fast R-CNN model at 0.82, which is an average value and it showed high localization accuracy. In operational decision-making surveillance applications, such a high level of accuracy in localization is needed.

Table 1: Performance Comparison of CNN and Fast R-CNN Models Based on Accuracy, Precision, Recall, and F1-Score Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	91.2	89.4	88.1	88.7
Fast R-CNN	94.7	93.1	92.4	92.7

Significantly surpassing the performance of other metrics employed, which was advantageous due to its clear observability in the results.

2) Dataset Description

The two distinct datasets of annotated images were specially curated for the context of the drone operation experience. Figures next will show the distribution of these training and test images within the datasets and example views of both RGB and drone imagery when stitched into panoramas. Notably, the drone imagery is of much lower resolution than the RGB imagery; both, however, were standardized in their resolutions in the training and test sets to effectively evaluate the performance of the algorithms. The reductions in the size of the image had to be done because of limitations in the architecture of the Faster R-CNN algorithm. Images were acquired under varied weather conditions giving both the frontal and side views of the target zones, with the acquisition dates strategically placed immediately after the applications of treatments. This temporal relationship was key in constructing the two datasets. While training the Faster R-CNN model, images from various collection dates were used; in the testing phase, a selection of hold-out images from the same fields was ideally taken on the same date to correspond to any gaps in the annotated data. Furthermore, throughout the data acquisition process, in a scenario when an acceptable Average Precision (AP) score for the background was achieved, unchallenged images were included in the test set to reinforce the completeness of the test collection without excluding any images.

3) Evaluation Metrics

The assessment of machine learning algorithms involves several different metrics, each providing particular insights into performance. Key evaluation metrics for the Faster R-CNN algorithm are Precision, Recall, and F1-score. Particularly in the domain of drone imagery, these are pivotal metrics to evaluate how well object detection works shows in figure 6. Precision may be defined as the ratio of true positive outcomes over the sum of true positives plus false positives. Recall, on the contrary, quantifies the ratio of true positives over the sum of true positives and false negatives. It is the F1-score that serves as the harmonic mean of the two—consequently embodying a single metric that balances both. With respect to this analysis, IoU was implemented as a threshold metric in object detection shows in figure 5. The IoU is fundamentally important for measuring the degree of overlap between a bounding box that has been predicted and the actual ground truth. It is calculated as the shared area between predicted bounding box and ground truth in it, divided by area of union for them. This computation ensures that the performance of the algorithm is being tested properly—that is, with the right criterion of what an intersection should be.

Bound box level precision, recall, F1-score was calculated by matching predicted bound boxes against ground truth bound boxes with a threshold of IoU greater than 0.5. Two hyperparameters underpinned the analysis: the pre-trained backbone version and input resolution concerning the dataset. Indeed, it was revealed that the computation memory footprint of Faster R-CNN allows it to operate on much smaller computational resources while achieving substantially super performance. Further, VGG-FPN was observed to outdo that with much more representation in the structured datasets. For the record, this paper does not judge the Faster R-CNN performance with high-performance computational settings, as the main aim here is to develop a truly robust and cost-effective model that acts as a provisional benchmark of state-of-the-art performance. Special care will have to be applied to avoid the usual pitfalls in machine learning model evaluations. For example, a common strategy employed to reduce overfitting is applied: All through training, the Faster R-CNN model is not supposed to see any image from the test set. Also, all performance metrics were calculated consistently, non-overlapping to weed off any possible confusion in the results.

4) Performance Metrics

While the visual analysis indicates that the model is performing well, actual performance metrics (like a confusion matrix or accuracy score) would require the true labels (ground truth) to compare against the model's predictions. However, based on the visual assessment:

- True Positives (TP): Most persons, cars, umbrellas, and fire hydrants are correctly detected with appropriately fitted bounding boxes.
- False Positives (FP): There don’t appear to be obvious false positives in the images provided. All labeled objects seem relevant.
- False Negatives (FN): There might be a few individuals not detected in the crowded scenes, especially in areas where people are tightly grouped or partially occluded.
- True Negatives (TN): The background and non-relevant objects are mostly ignored, which is a positive indication as shown in table 2.

Table 2: Comparison between Model's Predictions of Faster R-CNN

Faster-RCNN			
TP	**FN**	**FP**	**TN**
57.5	40	2.5	-
62.5	35	2.5	-
62.5	37.5	0	-
60.8	37.5	1.7	-

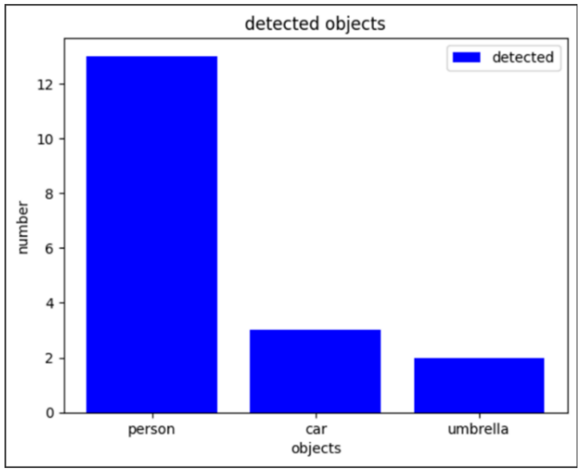

Figure 5: Illustrate Max-Min Value of Object Detection

Figure 6: Samples of the Drone Dataset

5 Conclusion and Future Directions

A new deep learning-centric framework that is ready to solve all image localization challenges associated with drone technology Concerning the drawbacks of traditional models, a new method for instant predictions in object detection was introduced. The efficacy of the framework was tested under various indexes of localization and classification errors on a specially prepared dataset. Most importantly, this dataset was used for the first evaluation of drone image localization performance. The quantitative results proved an increase in object localization precision as compared with the earlier traditional approaches. It is a very essential contribution to the fundamentals as well as the practical application in this field. IoU for the Fast R-CNN model Evaluation measured an average value of 0.82.

In forthcoming research endeavors, aim to broaden the application of various algorithms from multiple perspectives. A crucial prerequisite for assessing the localization accuracy of novel object detection algorithms is the establishment of an initial position. In instances where such a position is absent, it must be determined beforehand. focus is primarily on conducting relative evaluations; hence, concentrated exclusively on experiments involving drone imagery. Another avenue for future exploration involves the development of a comprehensive localization system that utilizes diverse sensor technologies within large and intricate environments. From this viewpoint, intend to combine inertial navigation systems with electro-optical sensors. The amalgamation of inertial sensors will enhance our capability for information fusion. The insights gained from this study are instrumental in the design of such systems, which are likely to facilitate the increasing integration of drones into everyday scenarios as they adopt various roles.

References

[1] Abbasi, G., Moridvaisi, H., & Ahmed, S. A. S. (2018). The algorithm for tracking non-rigid object in a sequence of images. *International Academic Journal of Science and Engineering*, *5*(2), 171–181. https://doi.org/10.9756/IAJSE/V5I1/1810034

[2] AbdAli, J. K., & Maaroof, M. K. A. (2022). Phase Change Analysis of Robot Laser Drilling with Accuracy Enhancement by Deep Learning. *International Journal of Analytical, Experimental and Finite Element Analysis*, *9*(3), 50-61. https://doi.org/10.26706/ijaefea.3.9.2211163

[3] Alanazi, J. M. (2023). Effective Machine-Learning Based Traffic Surveillance Moving Vehicle Detection. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *14*(1), 95-105. https://doi.org/10.58346/JOWUA.2023.I1.008

[4] Al-Dawoodi, A. G. M., & Mahmuddin, M. (2017). An empirical study of double-bridge search move on subset feature selection search of bees algorithm. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2-2), 11-15.

[5] Ali, R., Kang, D., Suh, G., & Cha, Y. J. (2021). Real-time multiple damage mapping using autonomous UAV and deep faster region-based neural networks for GPS-denied structures. *Automation in Construction*, *130*, 103831. https://doi.org/10.1016/j.autcon.2021.103831

[6] Almusawi, A., Albdairi, M., & Qadri, S. S. S. M. (2024). Integrating Autonomous Vehicles (AVs) into Urban Traffic: Simulating Driving and Signal Control [Article]. Applied Sciences (Switzerland), 14(19), Article 8851. https://doi.org/10.3390/app14198851

[7] Álvarez, R., Díez-González, J., Verde, P., & Perez, H. (2020). Comparative performance analysis of time local positioning architectures in NLOS urban scenarios. *IEEE Access*, *8*, 225258-225271. https://doi.org/10.1109/ACCESS.2020.3044389

[8] Behera, T. K., Bakshi, S., Sa, P. K., Nappi, M., Castiglione, A., Vijayakumar, P., & Gupta, B. B. (2023). The NITRDrone dataset to address the challenges for road extraction from aerial images. *Journal of Signal Processing Systems*, *95*(2), 197-209. https://doi.org/10.1007/s11265-022-01777-0

[9] Butcher, P. A., Colefax, A. P., Gorkin III, R. A., Kajiura, S. M., López, N. A., Mourier, J., ... & Raoult, V. (2021). The drone revolution of shark science: A review. *Drones*, *5*(1), 8. https://doi.org/10.3390/drones5010008

[10] Chakravarthy, A. S., Sinha, S., Narang, P., Mandal, M., Chamola, V., & Yu, F. R. (2022). Dronesegnet: robust aerial semantic segmentation for UAV-based IoT applications. *IEEE Transactions on Vehicular Technology*, *71*(4), 4277-4286. https://doi.org/10.1109/TVT.2022.3144358

[11] Devi, G., Kavitha, M., & Surendar, A. (2016). High Speed Image Searching for Human Gait Feature Selection. *International Journal of communication and computer Technologies*, *4*(2), 88-95.

[12] Dronova, I., Kislik, C., Dinh, Z., & Kelly, M. (2021). A review of unoccupied aerial vehicle use in wetland applications: Emerging opportunities in approach, technology, and data. *Drones*, *5*(2), 45. https://doi.org/10.3390/drones5020045v

[13] Du, G., Wang, K., Lian, S., & Zhao, K. (2021). Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. *Artificial Intelligence Review*, *54*(3), 1677-1734. https://doi.org/10.1007/s10462-020-09888-5

[14] Du, L., Sun, Y., Chen, S., Feng, J., Zhao, Y., Yan, Z., ... & Bian, Y. (2022). A novel object detection model based on faster R-CNN for spodoptera frugiperda according to feeding trace of corn leaves. *Agriculture*, *12*(2), 248. https://doi.org/10.3390/agriculture12020248

[15] Elmeseiry, N., Alshaer, N., & Ismail, T. (2021). A detailed survey and future directions of unmanned aerial vehicles (uavs) with potential applications. *Aerospace*, *8*(12), 363. https://doi.org/10.3390/aerospace8120363

[16] García-Fernández, M., Sanz-Ablanedo, E., & Rodríguez-Pérez, J. R. (2021). High-resolution drone-acquired RGB imagery to estimate spatial grape quality variability. *Agronomy*, *11*(4), 655. https://doi.org/10.3390/agronomy11040655

[17] Ghazi, A., Alisawi, M., Hammood, L., Abdullah, S. S., Al-Dawoodi, A., Ali, A. H., ... & Nawaf, A. Y. (2023, September). Data mining and machine learning techniques for coronavirus (COVID-19) pandemic: A review study. In *AIP Conference Proceedings* (Vol. 2839, No. 1). AIP Publishing. https://doi.org/10.1063/5.0167882

[18] Ghazi, A., Aljunid, S. A., Idrus, S. Z. S., Fareed, A., Al-dawoodi, A., Hasan, Z., ... & Abdullah, S. S. (2021). Hybrid Dy-NFIS & RLS equalization for ZCC code in optical-CDMA over multi-mode optical fiber. *Periodicals of Engineering and Natural Sciences (PEN)*, *9*(1), 253-276.

[19] Gupta, A., & Fernando, X. (2022). Simultaneous localization and mapping (slam) and data fusion in unmanned aerial vehicles: Recent advances and challenges. *Drones*, *6*(4), 85. https://doi.org/10.3390/drones6040085

[20] Hamouda, M., Ettabaa, K. S., & Bouhlel, M. S. (2020). Smart feature extraction and classification of hyperspectral images based on convolutional neural networks. *IET Image Processing*, *14*(10), 1999-2005. https://doi.org/10.1049/iet-ipr.2019.1282

[21] Hussein, M., Taher, A. Q., Mohammed, M. J., Ghazi, A., Awad, A. M., Hassan, S. I., ... & Azimov, B. A. Enhancing Drone Stability in Adverse Weather Conditions: A Novel MySQL-Based Approach for Balancing UAVs in Rain. *Radioelectronics. Nanosystems. Information Technologies.*, *16*(5). https://dx.doi.org/10.17725/j.rensit.2024.16.629

[22] Jońca, J., Pawnuk, M., Bezyk, Y., Arsen, A., & Sówka, I. (2022). Drone-assisted monitoring of atmospheric pollution—A comprehensive review. *Sustainability*, *14*(18), 11516. https://doi.org/10.3390/su141811516

[23] Kavitha, M. (2024). Embedded system architectures for autonomous vehicle navigation and control. *SCCTS Journal of Embedded Systems Design and Applications*, *1*(1), 25-28. https://doi.org/10.31838/ESA/01.01.06

[24] Kesavaraj, K., Mithran, N., Venkatesan, M., Vinoth, R., Dhamodharan, N. A., & Swetha, A. (2023, December). Security framework for net gun-equipped unmanned aerial vehicles. In *AIP Conference Proceedings* (Vol. 2914, No. 1). AIP Publishing.

[25] Khazaal Abbas Maaroof, M., & Bouhlel, M. S. (2024). A Survey of Deep Learning Techniques and Computer Vision in Robotic and Drone with Applications. In *BIO Web of Conferences* (Vol. 97, p. 00008). EDP Sciences. https://doi.org/10.1051/bioconf/20249700008

[26] Kim, Y. N., Ko, S. M., & Kim, T. (2022). Hidden Markov Model based Anomaly Detection Method for In-vehicle Network. *Journal of Internet Services and Information Security*, *12*(2), 115-125. https://doi.org/10.22667/JISIS.2022.05.31.115

[27] Kleinschroth, F., Banda, K., Zimba, H., Dondeyne, S., Nyambe, I., Spratley, S., & Winton, R. S. (2022). Drone imagery to create a common understanding of landscapes. *Landscape and Urban Planning*, *228*, 104571. https://doi.org/10.1016/j.landurbplan.2022.104571

[28] Labib, N. S., Brust, M. R., Danoy, G., & Bouvry, P. (2021). The rise of drones in internet of things: A survey on the evolution, prospects and challenges of unmanned aerial vehicles. *IEEE Access*, *9*, 115466-115487. https://doi.org/10.1109/ACCESS.2021.3104963

[29] Liu, G., Li, C., Zhang, S., & Yuan, Y. (2024). VL-MFL: UAV Visual Localization Based on Multi-Source Image Feature Learning. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2024.3383509

[30] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, *128*, 261-318. https://doi.org/10.1007/s11263-019-01247-4

[31] Liu, X., Dyke, S. J., Yeum, C. M., Bilionis, I., Lenjani, A., & Choi, J. (2020). Automated Indoor Image Localization to Support a Post-Event Building Assessment. *Sensors*, *20*(6), 1610. https://doi.org/10.3390/s20061610

[32] Ma, J., Jiang, X., Fan, A., Jiang, J., & Yan, J. (2021). Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision*, *129*(1), 23-79. https://doi.org/10.1007/s11263-020-01359-2

[33] Madhan, K., & Shanmugapriya, N. (2024). Efficient Object Detection and Classification Approach Using an Enhanced Moving Object Detection Algorithm in Motion Videos. *Indian Journal of Information Sources and Services*, *14*(1), 9-16. https://doi.org/10.51983/ijiss-2024.14.1.3895

[34] Mahmuddin, M., & Al-dawoodi, A. G. M. (2017). Experimental study of variation local search mechanism for bee algorithm feature selection. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2-2), 103-107.

[35] Mitra, R., Kaddoum, G., Dahman, G., & Poitau, G. (2021). Performance analysis of information theoretic learning-based cooperative localization. *IEEE Communications Letters*, *25*(7), 2196-2200. https://doi.org/10.1109/LCOMM.2021.3068258

[36] Muzammul, M., Algarni, A., Ghadi, Y. Y., & Assam, M. (2024). Enhancing UAV aerial image analysis: Integrating advanced SAHI techniques with real-time detection models on the VisDrone dataset. *IEEE Access*, *12*, 21621-21633. https://doi.org/10.1109/ACCESS.2024.3363413

[37] Noori, A., Amphawan, A., Ghazi, A., & Ghazi, S. A. (2019). Dynamic evolving neural fuzzy inference system equalization scheme in mode division multiplexer for optical fiber transmission. *Bulletin of Electrical Engineering and Informatics*, *8*(1), 127-135. https://doi.org/10.11591/eei.v8i1.1399

[38] Panigrahi, P. K., & Bisoy, S. K. (2022). Localization strategies for autonomous mobile robots: A review. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 6019-6039. https://doi.org/10.1016/j.jksuci.2021.02.015

[39] Papić, V., Šolić, P., Milan, A., Gotovac, S., & Polić, M. (2021). High-resolution image transmission from UAV to ground station for search and rescue missions planning. *Applied sciences*, *11*(5), 2105. https://doi.org/10.3390/app11052105

[40] Pazhani, A. A. J., & Vasanthanayaki, C. (2022). Object detection in satellite images by faster R-CNN incorporated with enhanced ROI pooling (FrRNet-ERoI) framework. *Earth Science Informatics*, *15*(1), 553-561. https://doi.org/10.1007/s12145-021-00746-8

[41] Placed, J. A., Strader, J., Carrillo, H., Atanasov, N., Indelman, V., Carlone, L., & Castellanos, J. A. (2023). A survey on active simultaneous localization and mapping: State of the art and new frontiers. *IEEE Transactions on Robotics*, *39*(3), 1686-1705. https://doi.org/10.1109/TRO.2023.3248510

[42] Quamar, M. M., Al-Ramadan, B., Khan, K., Shafiullah, M., & El Ferik, S. (2023). Advancements and applications of drone-integrated geographic information system technology—A review. *Remote Sensing*, *15*(20), 5039. https://doi.org/10.3390/rs15205039

[43] Qusef, A., Ghazi, A., Al-Dawoodi, A., Alsalhi, N. R., Shudayfat, E. A., Alqawasmi, A., ... & Murad, S. (2023). An energy management system using optimized hybrid artificial neural network for hybrid energy system in microgrid applications. *Operational Research in Engineering Sciences: Theory and Applications*, *6*(2).

[44] Schäffer, B., Pieren, R., Heutschi, K., Wunderli, J. M., & Becker, S. (2021). Drone noise emission characteristics and noise effects on humans—a systematic review. *International journal of environmental research and public health*, *18*(11), 5940. https://doi.org/10.3390/ijerph18115940

[45] Seo, D. M., Woo, H. J., Kim, M. S., Hong, W. H., Kim, I. H., & Baek, S. C. (2022). Identification of asbestos slates in buildings based on faster region-based convolutional neural network (faster R-CNN) and drone-based aerial imagery. *Drones*, *6*(8), 194. https://doi.org/10.3390/drones6080194

[46] Shinan, K., Alanazi, F., Alhazmi, H. E., Riaz, F., Fatima, N., Abid, A., ... & Ashraf, M. U. (2024). EUAC: An Advanced Privacy Protection for Location-based Service System. *Journal of Internet Services and Information Security, 14*(2), 202-225 https://doi.org/10.58346/JISIS.2024.I2.013

[47] Shiu, Y. S., Lee, R. Y., & Chang, Y. C. (2023). Pineapples' detection and segmentation based on faster and mask R-CNN in UAV imagery. *Remote Sensing*, *15*(3), 814. https://doi.org/10.3390/rs15030814

[48] Trivedi, J., Devi, M. S., & Solanki, B. (2023). Step Towards Intelligent Transportation System with Vehicle Classification and Recognition Using Speeded-up Robust Features. *Archives for Technical Sciences*, *1*(28), 39-56. https://doi.org/10.59456/afts.2023.1528.039J

[49] Tsintotas, K. A., Bampis, L., & Gasteratos, A. (2022). The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 19929-19953. https://doi.org/10.1109/TITS.2022.3175656

[50] Ullah, I., Adhikari, D., Khan, H., Anwar, M. S., Ahmad, S., & Bai, X. (2024). Mobile robot localization: Current challenges and future prospective. *Computer Science Review*, *53*, 100651. https://doi.org/10.1016/j.cosrev.2024.100651

[51] Wan, S., & Goudos, S. (2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks*, *168*, 107036. https://doi.org/10.1016/j.comnet.2019.107036

[52] Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*, 39-64. https://doi.org/10.1016/j.neucom.2020.01.085

[53] Xiao, Y., Wang, X., Zhang, P., Meng, F., & Shao, F. (2020). Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information. *Sensors*, *20*(19), 5490. https://doi.org/10.3390/s20195490

[54] Yan, D., Li, G., Li, X., Zhang, H., Lei, H., Lu, K., ... & Zhu, F. (2021). An improved faster R-CNN method to detect tailings ponds from high-resolution remote sensing images. *Remote Sensing*, *13*(11), 2052. https://doi.org/10.3390/rs13112052

[55] Zhuang, P., Li, H., Tan, S., Li, B., & Huang, J. (2021). Image tampering localization using a dense fully convolutional network. *IEEE Transactions on Information Forensics and Security*, *16*, 2986-2999. https://doi.org/10.1109/TIFS.2021.3070444

[56] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), 257-276. https://doi.org/10.1109/JPROC.2023.3238524

Authors Biography

Maysoon Khazaal Abbas Maaroof completed her Master's in Information Technology with a specialization in Computer Science and Engineering from Savitribai Phule Pune University, VIT College, India in 2015. She is currently working at the University of Babylon within the Basic Education College. She holds the position of departmental e-learning manager and administers the E-Learning Unit for the Computer and Mathematics Department. Playing a very active role in academic affairs, she manages the role of e-learning manager for her department's academics since 2015. Meanwhile, she is doing her PhD at SFAX University, Tunisia, which brings more exposure to that particular area. The passion for teaching and technology is reflected in her keen desire to promote digital learning and innovation in higher education. She has more than10 articles published and 2 book.

Med Salim Bouhlel Full Professor University of Sfax, Tunisia. Head of Research Lab SETIT since 2003 President — Tunisian association on HMI (since 2013) Editor-in-Chief “HMI” International Journals "MLHC and a dozen of special issues Chairman of many International conferences re-search interests: Image processing, Telecom, and HMI in which he has obtained more than 20 patents so far More than 500 articles published in IJ, IC & books.