ISSN: 0974-5823 Vol. 7 No. 7 July, 2022

International Journal of Mechanical Engineering

Human Detection for Search and Rescue Operations Using Convolutional Neural Networks

[1]Nayanthara Prathap, [2]Vishwajeet Raj, [3]Abhishek Tiwari, [4]Amrit Nand, [5]Madhura R

[1]Final year VIII Sem BE UG Student, USN: 1DS18EC057, Dept. of Electronics & Communication Engg., Dayananda Sagar College of Engg., Bangalore, Karnataka, India

[2]Final year VIII Sem BE UG Student, USN: 1DS18EC0101, Dept. of Electronics & Communication Engg., Dayananda Sagar College of Engg., Bangalore, Karnataka, India

[3]Final year VIII Sem BE UG Student, USN: 1DS18EC107, Dept. of Electronics & Communication Engg., Dayananda Sagar College of Engg., Bangalore, Karnataka, India

[4]Final year VIII Sem BE UG Student, USN: 1DS18EC109, Dept. of Electronics & Communication Engg., Dayananda Sagar College of Engg., Bangalore, Karnataka, India

[5]Associate Professor, Dept. of Electronics & Communication Engg., Dayananda Sagar College of Engg., Bangalore, Karnataka, India

I.Abstract

Due to various natural disasters and calamities, there is an increasing need to organize an efficient and timely search and rescue operation (SAR) to assist the stranded/injured persons. SAR operation aims to search the largest territory area in the least or rather shortest possible time and detects a lost or an injured human. At present, drones (UAVs or drones) are increasingly involved in search operations, as they can quickly capture a large, controlled area. However, a detailed examination of many recorded materials remains a problem. It is not easy for an expert to search relatively small people considering the area, often covered by vegetation or merged with the ground and in unusual positions due to falls, injuries, or exhaustion. Thus, the detection of persons in images captured by aerial vehicles is critical. In this, the reliability of the existing state-of-the-art detectors and motors or devices or algorithms like Faster R-CNN, YOLOv4, RetinaNet, and Cascade R-CNN on a VisDrone benchmark and are based on a custom-made dataset SARD built to simulate rescue scenes was studied and investigated. Post-training, the models on selected datasets and detection results are compared to increase the efficiency and rate of detection based on attributes. We chose TensorFlow along with YOLO because of its high speed, accuracy, and a small number of false detections. The project proposes a prototype model which can be used in search and rescue operations because of its excellent results in detecting people in search and rescue scenarios.

II.Introduction

Many people are stranded in vast, non-viable areas due to physical and natural calamities such as floods, and landslides, among the other disastrous effects. Emergencies arise, for example, due to incorrect assessment of the distance of the destination, incorrect assessment and prediction of the difficulty of the road, due to changes in weather/climate conditions, inadequate, minimalistic clothing or equipment, non-compliance with information and warnings, or undue preparation and overestimation of one's abilities or knowledge.

Search and rescue operations (SAR) require magnanimous human potential and material resources because they usually involve a large number of members of the rescue service team, search dogs, police and air forces, and, more recently, crewless aerial vehicles (drones). Drones are used for various purposes and have become a standard in all search and rescue services at a global scale. Apart from searches in urban and non-urban areas, drones are used for searching on water bodies (including sea, rivers, etc.) or from avalanches. In a country where we lose scores of people every year to floods and other natural disasters, we expect this kind of project to enable the enhancement of search operations and save lives with timely intervention. Drones have increased the probability of finding a person, and the search time is shortened due to "scanning" a vast area in one flight.

In search and rescue

operations, the key object is the person. However, recorded from a bird's eye view, such recordings are not contained in the large data sets on which these state-of-the-art detectors are trained and models function accordingly. To achieve the best or highest possible accuracy of the detection model, the dataset on which our model is trained must have similar conditions to those that appear while testing the model, so it is necessary to train the model keeping in mind a bird's eye view of the data. Recently, datasets include images taken by drones; those images are collected for various purposes, like detecting objects in images and videos, moving images, tracking one or more persons, detecting an action, predicting a person's movement, or recognizing other events in images.

On the other hand, each dataset is tailored and structured to a specific purpose and does not include scenes and rescue operations cases. The closest scenarios shot by a drone to those in search and rescue involve people present in a field/park while walking or running, standing in a square, walking down a street, or lying on a beach.

In this project work, we use our dataset for transfer learning of the selected state-of-the-art person detectors: Faster R-CNN, YOLOv4, RetinaNet, and Cascade R-CNN and for re-tuning for person detection in search and rescue scenes. To improve the detection results of the TensorFlow and YOLOv4 libraries for modeling, we will be analyzing the influence of different network resolutions, detection accuracy, and transfer learning settings on detection performance. The robustness of the machine learning model to weather conditions and motion blur are also elements to be tested through this model. Finally, after a complete and comprehensive testing and analysis of the results, we propose a prototype model for detection of persons in search and rescue (SAR) scenarios that can significantly help SAR operations.

III.Literature Survey

Due to various natural calamities like floods, etc., various issues related to the rescue system are faced, such as finding people stuck in floods or somewhere difficult to find and can save their lives. As a result, many works have been done regarding the rescue system using various machine learning, openCV, CNN, and other algorithms.

Authors of Detection of bodies in maritime rescue operations using Unmanned Aerial Vehicles with multispectral cameras[1] provided the method of detection of the bodies in maritime rescue operations using aerial vehicles with cameras such as drones, etc.

Zia Uddin and Mojaharul Islam, the publishers of Search and Rescue System for Alive Human Detection by Semi-autonomous Mobile Rescue Robot[2], implemented another camera method: A robot to get the actual footage to detect the human body needed to be rescued.

Jong Hyun Kim, Hyung Gil Hong, and Kang Ryoung Park in Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light Camera Sensors[3] wrecked on human detection during the night or in the dark, using CNN with the help of any Image.

The aim of Detection of Human Bodies using Computer Analysis of a Sequence of Stereo Images[4] by Jurij Leskovec, Sentjost, Slovenia Jure, Leskovec was to integrate different areas of Artificial Intelligence into a working system for detection and tracking of human bodies and recognition of its position in 3D space using ML.

Keni Bernardin, Florian van de Camp, Rainer Stiefelhagen in Automatic Person Detection and Tracking using Fuzzy Controlled Active Cameras[5] presents an automatic system for monitoring indoor environments using pan-tilt-zoomable cameras and detecting persons even from the fuzzy images.

IV.Motivation & problem statement of the project

In India, we have witnessed various places where floods and torrential rain have led to the loss of lives due to landfall, avalanches, and other harsh calamities. The reason behind this is the inability to locate persons stuck beneath the debris. There has lately been an increasing need to organize search and rescue operations (SAR) to provide assistance and health care to the injured and find them in time. SAR operation aims to search the largest area of a given territory in the shortest time possible and find a lost or injured person within the "golden hour". Today, drones (UAVs or aerial vehicles) are increasingly involved in search operations, as they can quickly capture a large, controlled area. However, a detailed examination of many recorded materials remains a standing problem faced. Even for an expert, it is considerably difficult to find relatively small people considering the area where they are, often covered by vegetation or merged with the ground and in unusual positions due to falls, injuries, or exhaustion.

V.Scope of the Project

The most significant aspect in searching for a missing person is that the detector locates that person, and it is equally important how accurate the detection is. As time passes, the probability of a missing person's survival decreases and the search area grows exponentially. As the searching for persons is relatively small compared to the environment or the aerial view of the image captured, they often take up only a few pixels on the screen. Maintaining long-term concentration and attention is challenging, even for people trained to locate people in large and vast mountainous areas or an area covered with vegetation. Therefore, we must have a proper model to detect the stranded humans with maximum efficiency so rescue operations can be carried out within the 'golden hour.'

VI. Objectives of the Project

Our project has four main objectives:

1. Human Detection using a drone in case of humans stranded in harsh conditions
2. Implementation for multiple domains, OS, and hardware types.
3. In Real-time Implementation, the model can also be used during bad weather.
4. The device can be used by disaster relief squads in real-time SAR operations.

VII. Proposed methodology with block diagrams

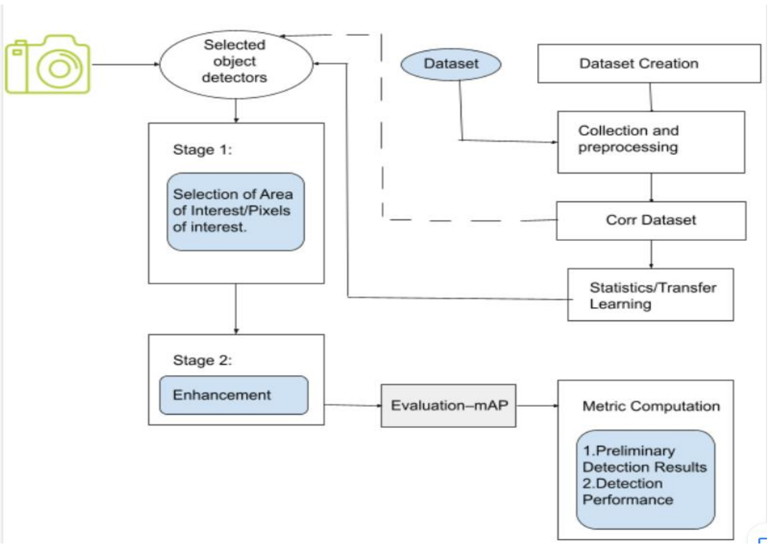

Fig.1: Block Diagram of our Working Model

DATASET CREATION

1) COLLECTION AND PREPROCESSING OF SARD DATASET

2) GENERATION OF CORR DATASET

3) STATISTICS OF DATASETS USED FOR TRANSFER LEARNING

SELECTED OBJECT DETECTORS

STAGE 1:

1) FASTER R-CNN

2) YOLOV4

3) RETINANET

4) CASCADE R-CNN

STAGE 2:

1) FEATURE PYRAMID NETWORK
2) EVALUATION METRICS
3) MEAN AVERAGE PRECISION (mAP)METRIC COMPUTATION

A. PRELIMINARY DETECTION RESULTS

B. DETECTION PERFORMANCE/ACCURACY AFTER TRAINING ON DOMAIN IMAGES

ROBUSTNESS TO WEATHER CONDITIONS AND MOTION BLUR

VIII. Tools that are going to be used for the project work

Software: language python and using Deep learning/ML/CNN. CNN-based object detectors, Cascade R-CNN, Faster R-CNN, RetinaNet, and YOLOv4 SARD dataset ROpti METRIC Pycharm, VS-code or Jupyter notebook IDE

Hardware: Drones/integrated hardware with in-built camera/UAV.

IX. Implementation process (h/w & s/w)

VGG, ResNet, ResNeXt, and MobileNet are pre-trained on the ImageNet or OpenImages dataset/SARD.

YOLO provides high speed and accuracy and a small number of false detections.

The SARD dataset will be used for transfer learning of the selected state-of-the-art person detectors: Faster R-CNN, YOLOv4, RetinaNet, and Cascade R-CNN and for re-tuning for person detection in search and rescue scenes.

A detector head can be divided into two types: one-stage and two-stage detectors. YOLO, SSD, and RetinaNet are examples of the one-stage detector. Two-stage detectors are R-CNN detectors, including Fast R-CNN, Faster R-CNN, and, R-FCN.

Feature Pyramid Network (FPN) is typically used to collect multiple feature maps, each with a different resolution, which helps recognize objects at different scales.

Image segmentation and contrast enhancement have been applied, followed by an SSD detector to detect persons in drone

and aerial images.

X. Results (experimentation, simulation)

Our project provides affordable, fast access to aerial data of a large area. This allows them to map the entire search zone and pinpoint possible places where the missing person might be trapped, detect and identify. The images here show the model detecting the humans in the images captured by our drone.

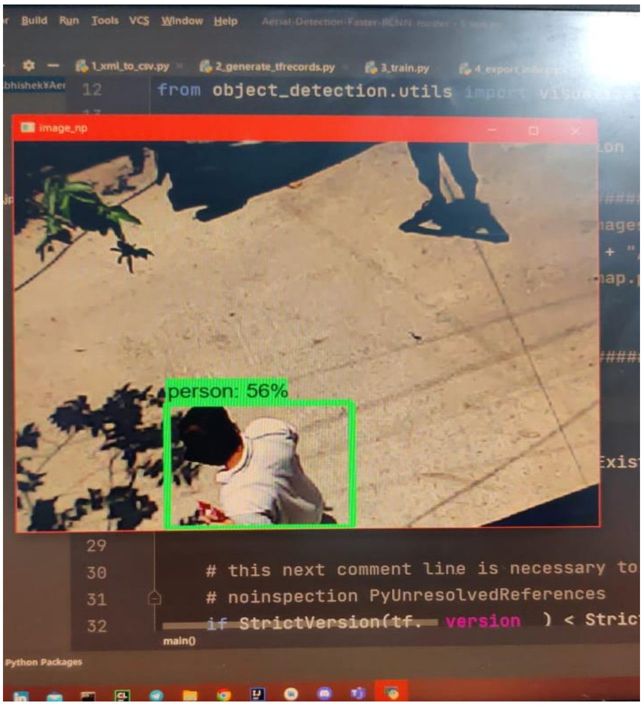

Fig.2(a): Human detection using images captured by drone along with probability/efficiency of detection

Fig.2(a): Human detection using images captured by drone along with probability/efficiency of detection

XI. Advantages / Applications

1) In search and rescue operations, teams race against the clock. Vision Aerial drones take less than two minutes to deploy, whereas helicopters or planes take much longer.
2) Our project protects human life in hard-to-reach areas such as mountains, forests, canyons, caves, and bodies of water.
3) Drones (UAVs or drones) are increasingly involved in search operations, as they can quickly capture a large, controlled area.
4) The automatic detection of persons and objects in images/videos taken by drones in these operations is very significant.
5) A drone outfitted with a thermal and infrared camera provides rescuers with invisible information to the naked eye.
6) Helicopters and airplanes can cost several thousand dollars per mission; a drone is a fraction of the cost when amortized over its lifetime.

XII. Conclusions

The ability to automatically detect people on drone images using computer vision methods is a significant help in SAR operations. This project will explore the state-of-the-art person detectors in drone images and propose a model for detecting persons in SAR actions.

Our model has achieved the best detection performances on the formed dataset in terms of average precision (AP) considering IoU precision and the object size, as well as the least false detection (FP), so it was further used in the experiment, called as YOLOv4. It is used along with TensorFlow, imutils, FuncTools, and Object Detection libraries for maximum efficient output of the detection in the image generated. In searching for missing people, the most critical thing is that the detector locates that person, and it is also essential how accurate or precise the corresponding detection is. We have experimentally selected parameters as a trade-off between accuracy and recall so that the model can be helpful in SAR actions.

The model's robustness will be tested on images with artificially generated bad, undesired weather conditions, image blur, and the results show a severe decrease in AP by more than 35%. After the model was also trained on the part of the images with bad weather effects, the model achieved significantly better results. To improve model training, we want to investigate how different transfer learning strategies regarding different combinations of datasets affect the detection result. After training the model on the selected dataset, each model is referred to in the text as a model (dataset) to make it easier to compare the models' performances and efficiency scores.

The results achieved by training the models in different training sets will be used one after another in a different order or mixed. In addition to the accuracy values, the improvement (Imp) of the model concerning the initial weights (original model) and ROpti values will also be shown. We mentioned earlier that the crucial factors in search and rescue operations

are detection accuracy and the speed of finding the missing person. Therefore, building a model with as few false detections (FP) as possible is essential because they consume or take into consideration human resources and take required time. For this reason, we introduced additional metrics that we called ROpti, computed as the ratio of the difference between true (TP) and false positives (FP) detections and possible detections (TPCFN) in the dataset.

$$ROpti = (TP - FP) / (TP + FN)$$

For perfect precision (no false positive), ROpti is equal to recall, and with perfect recall (no false negative), ROpti is equal to 1, which is a perfect score. As the number of FPs grows, ROpti decreases. In case TP is equal to FP, then ROpti is equal to zero, ROpti becomes negative, while TP is less than FP. The detection results consider ROpti measure, e.g., true and false-positive detections out of a total of 2500+ objects for different network resolutions with a default threshold of 0.25.

In future work, the plan is to use a thermal camera to increase detection performance and develop a prototype model for recognizing human activity (running, walking, standing, sitting, lying down) and tracking people in SAR scenes.

XIII. Possible outcome of the project

a) a novel dataset (SARD) of drone imagery in search and rescue operations, with statistics of the occurrence of a small, medium, and large object, annotated and prepared for supervised machine learning,

b) comparison of the performance of selected CNN detectors (Cascade R-CNN, Faster R-CNN, RetinaNet, YOLOv4) for use in search and rescue operations,

c) analyses of the influence of different network resolutions, detection accuracies, and condense values on YOLOv4 person detection performance, and analysis of different transfer learning strategies considering the impact on detection results,

d) proposal of ROpti metrics for evaluating detector performances for SAR operations, by taking into account that there are as many positive detections as possible and as few false detections as possible,

e) proposal of YOLOv4 model to be used for person detection in serch and rescue actions taking care to achieve the highest possible accuracy, with as few false detections that can be possible, with a network configuration which allows a human's online location and a configuration for off-line analysis, robust to numerous weather conditions.

XIV. References

[1] Detection of bodies in maritime rescue operations using Unmanned Aerial Vehicles with multispectral cameras- Antonio-Javier Gallego, Antonio Pertusa, Pablo Gil, and Robert B. Fisher.

[2] Search and Rescue System for Alive Human Detection by Semi-autonomous Mobile Rescue Robot- Zia Uddin, Mojaharul Islam.

[3] Convolutional Neural Network-Based Human Detection in Nighttime Images Using Visible Light Camera Sensors- Jong Hyun Kim, Hyung Gil Hong, and Kang Ryoung Park.

[4]. Detection of Human Bodies using Computer Analysis of a Sequence of Stereo Images- Jurij Leskovec, Sentjost, Slovenia Jure, Leskovec.

[5] Automatic Person Detection and Tracking using Fuzzy Controlled Active Cameras- Keni Bernardin, Florian van de Camp, Rainer Stiefelhagen.

[6] Object Detection using Convolutional Neural Networks with the help of a webcam, etc.- https://medium.datadriveninvestor.com/object-detection-with-convolutional-neural-networks-dde190eb7180

[7] Comprehensive tutorial to learn Convolutional Neural Networks (CNN)- https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/

[8] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[9] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53 (2021). https://doi.org/10.1186/s40537-021-00444-8

[10] X. Liu, Y. Su, and B. Xu, "The Application of Graph Neural Network in Natural Language Processing and Computer Vision," 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), 2021, pp. 708-714, doi: 10.1109/MLBDBI54094.2021.00140.

[11] S. Kim and S. Yoo, "Multimodal Surprise Adequacy Analysis of Inputs for Natural Language Processing DNN Models," 2021 IEEE/ACM International Conference on Automation of Software Test (AST), 2021, pp. 80-89, doi: 10.1109/AST52587.2021.00017.

[12] A. R. Sharma and P. Kaushik, "Literature survey of statistical, deep and reinforcement learning in natural language processing," 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 350-354, doi: 10.1109/CCAA.2017.8229841.