

Rescue Bot

Aadya¹, Aryan Goyal², Himanshu Sachdeva³, Vaibhav Srivastava⁴, Parth Shukla⁵, Dr Harpreet Vohra⁶ and
Dr Gaganpreet Kaur⁷

¹⁻⁷Thapar Institute of Engineering & Technology, Patiala, India

Email: aaadya_be19@thapar.edu, agoyal9_be19@thapar.edu, hsachdeva_be19@thapar.edu, vsrivastava_be19@thapar.edu,
pshukla_be19@thapar.edu, hvohra@thapar.edu, gaganpreet.kaur@thapar.edu

Abstract—Natural disasters have been a major problem for humanity and cause of majority of deaths in the past few centuries. Just in the last 2 decades about 7348 disasters were recorded worldwide causing 1.23 million deaths. Governments across the globe have implemented various methods and strategies to solve this issue but all of them are limited by one way or another i.e. being costly or having low efficiency or being slow to implement. Search and Rescue (SAR) operations have primarily been conducted using the combined efforts of police and relief squads, but with recent development in Artificial intelligence (AI) and deep learning, disaster management have become more efficient and fast in saving lives. Use of unmanned aerial vehicle (UAV) has helped the operatives to scan the terrain and locate disaster struck people. This paper presents a solution to help SAR operatives by combining both the deep learning technology and bot technology such as buggy and UAVs. With the use of microprocessors such as raspberry pi or orange pi, the Machine Learning (ML) model is connected to the buggy via Wi-Fi. Using this technique a low cost solution has been achieved with a mean average precision (mAP) of 30.3 has been achieved. Combining this system with Thermal Imaging Camera (TIC), temperature readings are recorded allowing the operator to detect objects under rubble by looking at abnormal temperature readings.

Index Terms— Natural Disasters, Disaster Management, Thermal Imaging, AI, Deep Learning, Object detection.

I. INTRODUCTION

Natural disasters are increasing day by day and their effects are so long lasting that they leave a huge impact on the minds of people either due to losing their loved ones in such horrendous events or due to the effects that they make on the beauty of the Earth. On an average about 150 million of lives were affected by the natural calamities every year in the last 20 years. Though many steps have been taken to prevent such massive destruction like the use of drones and bots, implementation of computational programming and Deep Learning but all of these had one problem thought they were built to provide more information after the disaster but did not aid the operatives during the operation. This paper will be explaining the formation and implications of the bot along with the existing and newly evolved technologies used for forming the bot. For example “Guardian S”, a snake like robot developed by company named SARCOS is designed to traverse high terrain and analyzing the situation to report to the rescue team before conducting the rescue operations.

Figure 1 displays a flow chart of how the SAR operations are conducted. A search team is sent out to scan the terrain and look for survivors this is usually done via helicopters. If any survivor is found the search team contacts rescue squads which are then moved to their location to help the survivors. All the survivors are

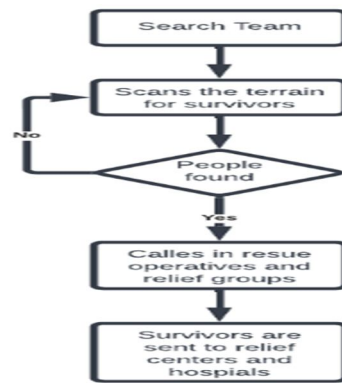


Figure 1. Human detection by rescue management system

transferred to a nearby hospital or relief centres for treatment. This process is extremely lengthy and slow, though many lives have been saved by this process just as many lives have been lost. With the use of AI the system will detect the presence of human. Using computational programming especially the ML/DL the system is able to surround the detected person with a bounding box. With the increase of natural disasters, embedding the technologies of AI and IOT in SAR operations have become a necessity in order to increase the survival rate. The paper is organised into various subsections the section II deals with literature review section III deals with the methodology describing the system architecture and flowchart of the model section IV deals with the working of the rescue model. section V deals with the results obtained and sections VI, VII deals with conclusions and references respectively.

II. LITERATURE REVIEW

Many technologies have been developed for object detection and face recognition. You Look Only Once (YOLO V1) is one of the fastest object detection system having a simple & unified architecture that trains on full images & directly optimises the detection performance, being highly precise for large sized objects it is less precise for small sized objects. Further improvements on YOLOV1 led to YOLO V3. which is more accurate & better performer and provides precise, easy & accurate measurement of small sized objects due to better utilisation of GPU but is large in size due to more convolutional layers ,whereas EfficientDet & BiFPN. are cheaper ways of object detection, give high accuracy & better efficiency. BiFPN is a solution to multi scale feature fusion & is formed by removing nodes at Input edge & adding an extra edge from input to output nodes. EfficientDet is a highly efficient, accurate & fast family of object detectors formed by the combination of efficient Netbackbone, BiFPN & compound scaling. It has the limitations of scaling and multi scale feature fusion. On the other hand MobileNetV2 was developed to detect mobile applications accurately without any special operators, being highly efficient in computation with less cost & more precise. At last there is SSD Lite which is a mobile friendly variant of regular SSD and is more computationally efficient as parameter count & computational cost is significantly low. In case of face recognition there are existing technologies such as Principle Component Analysis (PCA) that can pre-process the data by reducing the dimensions of data by removing redundant information., Linear Discriminate Analysis (LDA) is used mainly for face classification. It does dimensionality reduction and label info to evaluate different categories of data, Support Vector Machine (SVM) which is used for small samples & high dimensionality face recognition problem. Other one used is Adaptive Boosting (ADABOOST) that can resolve the problem of adjusting the training set& combining the weak classifier to a strong classifier. Deep Learning (DL) also plays an important role in face recognition. Rules used to learn face recognition features by DL are like Local Binary Patterns (LBP), Local Ternary Patterns (LTP), Gabor feature, Non Negative Matrix Factorization (NMF). It faces Pose Illumination Expression problem. There are two techniques called Arcface & Sub-center Arcface which encourages one dominant sub-class that contains different samples & enhances the discriminative features. DCNN trains itself by two ways by softmax loss based methods & Triplet loss based methods but softmax loss based methods are preferred for their high accuracy due to clean data sets. There is a technique named as LOCAL BINARY PATTERNS in which face is divided into small regions by considering the shape and texture information of face and then the LBP histograms are extracted & concatenated into a single histogram to represent the face image. It extracts features efficiently & simply at a faster pace. The operator uses several neighbourhoods of different sizes including the circular

neighbourhoods as well. In histograms there is a description of face in multiple levels, where the labels contain information about the patterns on a pixel level, this information is further used to get data on a regional level and the regional histograms are concatenated to get the global description of face, but the main problem is the classification of histograms on the basis of face textures. In case of imaging & sound detection there are existing technologies such as the first one is MOBNET System, which is used to gather the end user requirements to create a solution for supporting the traditional activities of Search and Rescue (SAR) teams. It uses Digital Cellular Technology (DCT)-European Global Navigation Satellite System (EGNSS) technology to track victims. It is a cost effective and a latest technology version, based on the fact that the shorter the time between the deployment and victim localization the more likely the victim can be rescued. Another is Thermal Imaging Surveillance System in which thermographic cameras are used in factories with heavy machines so that we can monitor the conditions of machines by analysing the temperature conditions. This system can also be used to detect trespassers in poor lightning condition with high accuracy. The temperatures emitted by objects are recorded by the cameras in the IR region which can be used for accurate geometric survey of deformations caused due to moisture, geometric calibration (focal length, position of principal point, radial, tangential distortions).

III. MOTIVATION

Natural disasters are increasing day by day and their effects are so long lasting that they leave a huge impact on the minds of people either due to losing their loved ones in such horrendous events or due to the effects that they make on the beauty of the Earth. Some of the examples are:

- Heat waves in Europe in 2003 which killed 72,200 living beings in many European countries.
- Earthquakes in Pakistan in 2005 which killed 73,300 living beings & many more people were left missing.
- Earthquakes in China in 2008 which killed 87,500 living beings.

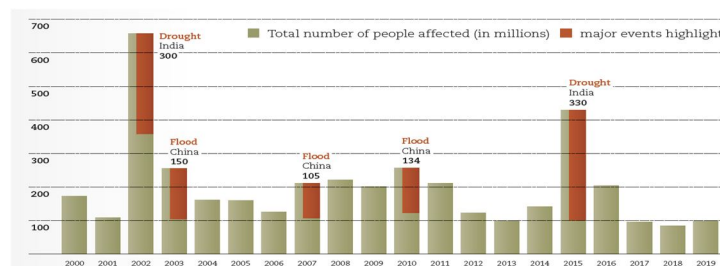


Figure 2. Number of people affected per year in millions

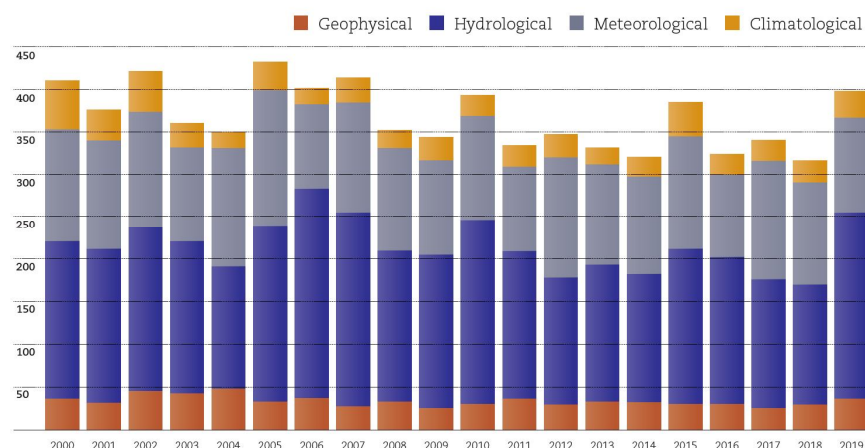


Figure 3. Number of disasters per year

The graphs represent the number of disasters and lives affected by them (in millions) every year.

Figure 2 shows that on an average about 150 million lives were affected by natural disaster every year from 2000-2019.

Figure 3 depicts the number of natural disasters every year during the same period, key thing to note is the fact that Hydrological and meteorology disaster like floods, accounts for about 70% of all natural disasters. After seeing these gruesome events occurring in front of us, we aimed to reduce deaths caused due to delays in SAR operations; we tried to implement a bot that could help in improving these situations, even minor improvements in the response time of the authority's makes massive difference. If the authorities can be equipped with additional technology that could estimate the location of disaster & identify the presence of individual at that location. After identification, we can rescue them by using appropriate manpower and cut down on search time effectively.

IV. PROBLEM DOMAIN

The scope of our project is based on solving some problems such as:

1. Property damage and money needed for operation isn't effective.
2. time needed for locating the victims still remains high thus may lead to not having an effect.
3. Detecting people struck under rubble or drowning remain ineffective.

And in the bot by using a TIC we eliminate the limitations of optics as it traces heat which is not limited by LOS. Since the Rescue Bot is built to accompany the operatives this allows it to be an aid and can result in saving both human livestock and property damages. By having a proper combination of hardware and software the project can be built upon both a land bot and an aerial UAV.

V. METHODOLOGY

The Rescue Bot is equipped with an infrared thermal imaging camera to help in detecting the living beings' presence, using the I2C connection on the SBC the TIC will work send its received data to the SBC. The co-equipped HD camera shall provide a stable feed to the base station. The whole system will be powered and connected by "Orange Pi," which shall serve as a connecting tissue between our bot and base station via its inbuilt Wi-Fi and Bluetooth. Using the Putty software, the SBC can be used as a host and create a local server an a given internet connection and any pc on the same network can act as a client via Putty. This implementation method allows for greater versatility in its usage and gives it high upgradeability and accessibility options.

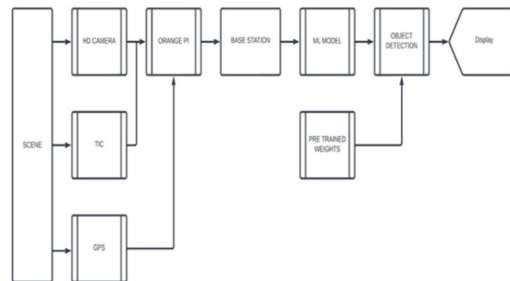


Figure 4. System Architecture

The above figure depicts the system architecture of the rescue bot. The system works in a combination of ML and IOT, the bot is moved to the scene to be investigated. The scene is captured by TIC and the HD camera both. While the GPS module tracks the longitude and latitude of the bot. The feed from both TIC and HD camera is sent to the microprocessor. The info is then sent to the base station which is connected to the same network as to the microprocessor. At the base station the info is processed, and the video stream is sliced into frames and fed to our ML model while the feed from TIC and GPS are displayed on the screen. ML model scans the frames and carries out object detection using the pre-trained weights. If the required object is found, the model the bounds it with a bounding box and displays the result. Finally, the screen has a stream of the scene with a bounding box surrounding the desired object, if found, alongside the feed from Thermal camera and GPS location displayed in separate windows.

The scene will be detected and captured via both the HD camera and Thermal Imaging Camera (TIC), the HD camera will capture the scene using visible light and TIC shall make use of IR rays to form a thermal map.

Both images will be sent to Orange-pi.

Next the Orange pi shall provide a live stream of the images it received to the connected display and feed it to our DL model via using the internet connection provided by our Wi-Fi module.

The pi shall also provide a stream of GPS location provided by our GPS module.



Figure 5. System Flowchart

The DL model will then make use of its provided tools to process the images and make them comparable to the database provided to it.

The images will be matched with the data base and the results will be based on if the incoming images matched the database images.

In case the images match, the result will be displayed in form of a bounding box upon the object matched and no bounding box if none detected.

VI. WORKING

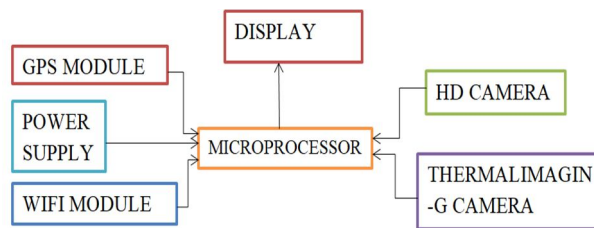


Figure 6. Circuit Diagram of Buggy

The Figure 6. depicts the basic working of the bot. HD camera will capture the feed and send it to the microprocessor using the USB interface. TIC will send its signal by using 26-pin GPIO, which will be used by GPS to send the coordinates of the object. The wi-fi module will be used for to and fro communication using the USB interface. The microprocessor will be powered by 10000 mAh power bank via the USB-C charging port.

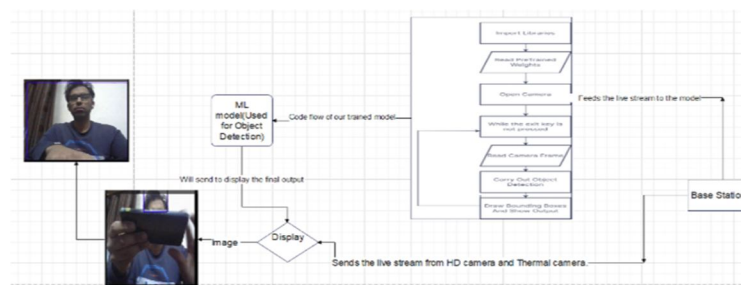


Figure 7. Flowchart of Object Detection

This Figure7 shows the procedure of functioning of code .Object Detection is done by using python and the efficientdet-lite0 model, which has a latency of ~146ms, giving the system about 7 frames per second. The pre-trained weights are taken from Tensor Flow lite. The camera's input is taken and executed the pre-trained neural network on the frames they received. Using the above mentioned pre-trained weights and confidence tuned, the program will draw rectangular bounding boxes around the objects along with an appropriate output label showing what they have detected.

VII. RESULTS

The final outcome of the project is a system that combines a bot with Object Detection technology using deep learning. This advanced bot has the capability to move autonomously across a field while remaining connected to a base station via a Wi-Fi connection facilitated by the Raspberry Pi, Orange Pi, or any other type of SoC board.

To achieve this level of sophistication, the model has been trained on the VisDrone-2019 dataset, which is a comprehensive collection of aerial images and videos captured by drones in various locations. These images and videos contain a wide range of objects such as pedestrians, vehicles, and other elements that can be encountered in a real-world environment. The dataset serves as an ideal source of information for training the bot to recognize and classify different objects.

The implementation of transfer learning on EfficienDet-D0, a highly efficient object detection model, has resulted in an mAP of 28.5, indicating the bot's ability to accurately detect and classify objects. This achievement is a testament to the effectiveness of using deep learning and transfer learning in developing advanced computer vision systems.

Overall, the integration of the bot and object detection technology using deep learning has the potential to revolutionize various industries, such as agriculture, construction, and security, by providing a highly efficient and cost-effective solution for automating tasks that require visual intelligence.



Figure 8. Sample Images of visDrone-2019 dataset

TABLE I. MODEL RESULT ON VISDRONE-2019 DATASET

Model	Test-dev			Val
	AP	AP50	AP75	AP
Proposed Model	30.5	48.5	29.7	28.5
YOLOV3	30.2	55.7	31.5	29

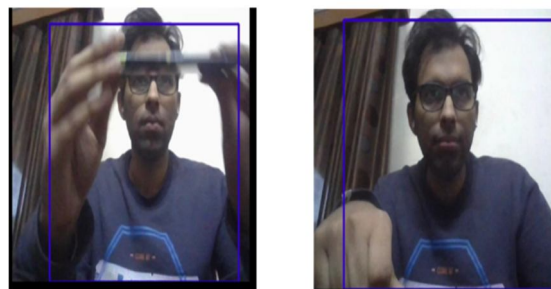


Figure 9 .Code Output with & Without Obstruction

In the system, the HD camera feed is initially captured and sent to the microprocessor. From there, the feed is streamed using HTTP Live Streaming to the connected system via WiFi, as shown in Figure 10. This feed can be viewed on the connected system and allows users to observe the bot's movements and the objects it detects in real-time.

The stream from the camera feed is then fed into the machine learning (ML) model, which has been trained on the VisDrone-2019 dataset using transfer learning on EfficienDet-D0, which is a state-of-the-art, very efficient due to its Bidirectional Feature Pyramid and highly scalable over different resolution and input image size. The ML model analyzes the stream and identifies any objects present in the field of view.

Once an object has been detected, the ML model draws bounding boxes around it, indicating its real time location in the frame. This information can be used by the user on the base station and mark their locations.

TABLE II. PERIMETER COMPARISON SHEET

Model	TEST-DEV			Val	Params	FLOPs
	AP	AP50	AP75	AP		
EfficientDet-D0	33.6	53.5	32.1	30.3	3.9M	2.5B
YOLOv3	33.0	57.9	34.4	-	-	71B
EfficientDet-D1	40.5	59.1	43.7	40.2	6.6M	6.1B
RetinaNet-R50	39.2	58.0	42.3	39.2	34M	97B
RetinaNet-R101 (640)	39.9	58.5	43.0	39.8	53M	127B
EfficientDet-D2	41.8	60.5	45.6	43.5	8.1M	11B
Detectron2 Mask R-CNN R101-FPN	-	-	-	42.9	63M	164B
Detectron2 Mask R-CNN X101-FPN	-	-	-	44.3	107M	277B
EfficientDet-D3	47.0	65.4	51.5	42.6	12M	25B
ResNet-50 + NAS-FPN	44.2	-	-	-	60M	360B
ResNet-50 + NAS-FPN	44.8	-	-	-	60M	563B
ResNet-50 + NAS-FPN (1280@384)	45.4	-	-	-	104M	1043B
EfficientDet-D4	46.2	65.8	53.4	45.5	21M	55B
AmoebaNet+ NAS-FPN +AA	-	-	-	48.6	185M	1317B
EfficientDet-D5	52.6	69.7	56.1	51.3	34M	135B
Detectron2 Mask R-CNN X152	-	-	-	50.2	-	-
EfficientDet-D6	52.6	71.5	57.2	52.2	52M	226B
AmoebaNet+ NAS-FPN +AA	-	-	-	50.7	209M	3045B
EfficientDet-D7	51.4	70.7	55.8	53.4	52M	325B

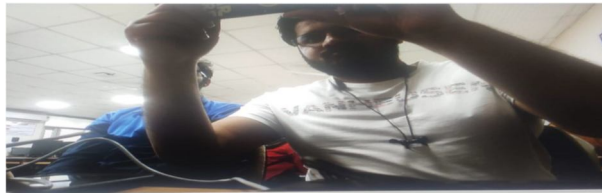


Figure 10 Live Streaming

The transmission of the thermal camera feed to the base station allows for real-time monitoring of the area under surveillance at the base station. Which can be used to view the thermal image feed and the temperature readings as it is displayed in Figure 11.

```

Ambient Temperature : 41.23
Object Temperature : 33.91
Ambient Temperature : 41.23
Object Temperature : 33.91
Ambient Temperature : 41.23
Object Temperature : 33.91
Ambient Temperature : 41.23
Object Temperature : 33.91
Ambient Temperature : 41.23
Object Temperature : 33.77
Ambient Temperature : 41.23
Object Temperature : 33.77
Ambient Temperature : 41.23
Object Temperature : 33.77
Ambient Temperature : 41.21
Object Temperature : 33.77
Ambient Temperature : 41.21
Object Temperature : 33.77
Ambient Temperature : 41.21
Object Temperature : 33.77
Ambient Temperature : 41.21
Object Temperature : 33.77
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 33.87
Ambient Temperature : 41.21
Object Temperature : 34.13
20_Sep_2020|20:48:55
20_Sep_2020|20:48:55

```

Figure 11.TIC Output

```

Latitude=12.9814865and Longitude=77.6683425
Latitude=12.9814848333and Longitude=77.6683436667
Latitude=12.9814841667and Longitude=77.6683451667
Latitude=12.9814818333and Longitude=77.6683461667
Latitude=12.9814853333and Longitude=77.6683491667
Latitude=12.9814783333and Longitude=77.6683485
Latitude=12.9814701667and Longitude=77.6683466667
Latitude=12.981464and Longitude=77.668345
Latitude=12.9814586667and Longitude=77.6683438333
Latitude=12.9814525and Longitude=77.6683428333
Latitude=12.9814458333and Longitude=77.6683421667
Latitude=12.9814395and Longitude=77.6683421667
Latitude=12.9814331667and Longitude=77.668342
Latitude=12.981428and Longitude=77.6683425
Latitude=12.981423and Longitude=77.6683428333
Latitude=12.9814185and Longitude=77.6683431667
Latitude=12.9814146667and Longitude=77.6683436667
Latitude=12.9814095and Longitude=77.6683443333
Latitude=12.9814056667and Longitude=77.6683456667
Latitude=12.981401and Longitude=77.668346
Latitude=12.9813966667and Longitude=77.66834

```

Fig 12.GPS Coordinates

The GPS coordinates are sent from GPS module to the microprocessor and will be displayed on the serial monitor as displayed by Figure12.

VIII. CONCLUSION

There is a need of a strong Object Detection System that can help in identifying the presence of human beings during rescue operations due to natural calamities like earthquakes, floods etc. With the above research, the concept of using a 'Rescue Bot' for object detection came up. The rescue bot equipped with an infrared thermal imaging camera is combined with a pre-trained ML model that will help in identifying human thermal maps. Based upon the surveys, Efficientdet-lite0 object detection model was selected because it is good with light weight computation devices, less power consumption, better performance and better streaming quality when it is being used in a wireless connection. Based upon IOT, the whole system will be powered and connected by Orange Pi 3lts which will be a connecting tissue between the bot and the base station. The model when tested concluded that the human trapped between debris is easily detectable.

REFERENCES

- [1] Półka, M., Ptak, S., & Kuziora, Ł. (2017). The use of UAV's for search and rescue operations. Procedia engineering, 192, 748-752.

- [2] Yastikli, N., & Guler, E. (2013). Performance evaluation of thermographic cameras for photogrammetric documentation of historical buildings. *Boletim de Ciências Geodésicas*, 19, 711-728.
- [3] Wong, W. K., Tan, P. N., Loo, C. K., & Lim, W. S. (2009, April). An effective surveillance system using thermal camera. In 2009 international conference on signal acquisition and processing (pp. 13-17). IEEE
- [4] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. ArXiv preprint arXiv:1804.02767.
- [5] Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790).
- [6] Li, L., Mu, X., Li, S., & Peng, H. (2020). A review of face recognition technology. *IEEE access*, 8, 139110-139120.
- [7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
- [8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
- [9] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699).
- [10] Ahonen, T., Hadid, A., & Pietikainen, M. (2004, May). Face recognition with local binary patterns. In European conference on computer vision (pp. 469-481). Springer, Berlin, Heidelberg.
- [11] Yaghmaei, N. (Ed.). (2020). Human Cost of Disasters: An Overview of the Last 20 Years, 2000-2019. UN Office for Disaster Risk Reduction.
- [12] Guardian, S. (2017). Guardian S.
- [13] Kylo v. United States. (n.d.). Oyez. Retrieved August 11, 2022, from <https://www.oyez.org/cases/2000/99-8508>
- [14] Naing, K. M., Zakeri, A., Iliev, O., & Venkateshaiah, N. (2018, September). Application of Deep Learning Technique in UAV's Search and Rescue Operations. In Proceedings of SAI Intelligent Systems Conference (pp. 893-901). Springer, Cham
- [15] Zuluaga, J. G. C., Leidig, J. P., Trefftz, C., & Wolffe, G. (2018, July). Deep reinforcement learning for autonomous search and rescue. In NAECON 2018-IEEE National Aerospace and Electronics Conference (pp. 521-524). IEEE.
- [16] Goyal, S., Shagill, M. D., Kaur, A., Vohra, H., & Singh, A. (2020). A yolo based technique for early forest fire detection. *Int. J. Innov. Technol. Explor. Eng.(IJITEE)* Vol, 9, 1357-1362.
- [17] "International union for conservation of nature", [online]. Available: <https://portals.iucn.org/library/sites/library/files/documents/2019-007-En.pdf>, last accessed 2019/08/09.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
- [19] Xiang, Q., Wang, X., Li, R., Zhang, G., Lai, J., & Hu, Q. (2019, October). Fruit image classification based on Mobilenetv2 with transfer learning technique. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering (pp. 1-7).
- [20] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2037-2041.
- [21] Tang, H., Yin, B., Sun, Y., & Hu, Y. (2013). 3D face recognition using local binary patterns. *Signal Processing*, 93(8), 2190-2198.