JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Clip-Based Method for Efficient Building Damage Assessment Using UAV Images

Amir Azizi
Research Associate
CYENS CoE, Nicosia, Cyprus

Abstract: Building damage assessment after natural disasters is a critical task in the recovery process. In this paper, we present a machine learning-based approach for multilabel classifying damaged buildings using UAV imagery and image processing techniques. Our method utilizes the RescueNet image dataset collected with DJI Mavic Pro quadcopters after Hurricane Michael. The dataset contains 4494 images divided into training, validation, and test sets. We employ a pre-trained deep neural network based on the CLIP (Contrastive Language-Image Pre-Training) method and fine-tune it on the RescueNet dataset to achieve accurate building damage assessments. Our experimental results, compared with state-of-the-art methods such as YOLOV8, EfficientNet, and MobileNetV2, indicate that our proposed method achieves the best performance in terms of accuracy and speed, demonstrating its superiority for object detection tasks. The final accuracy of our method is 92%, which demonstrates its effectiveness in real-world scenarios. Results show that the proposed approach provides a fast and reliable solution for building damage assessment and has the potential to be widely applied in disaster management.

Index Terms **- building damage assessment-UAV imagery-Image processing-Multi label classification-Machine learning.**

I. INTRODUCTION

Building damage assessment is a crucial task in the aftermath of natural disasters. It helps to determine the extent of the damage and prioritize the rebuilding process. In recent years, unmanned aerial vehicles (UAVs) have become a popular tool for collecting high-resolution imagery of disaster-affected areas. The images captured by UAVs provide valuable information for building damage assessment, but the manual process of analyzing this information can be time-consuming and prone to errors. To address this challenge, this paper presents a machine learning-based approach for the multilabel classification of damaged buildings using UAV imagery and image processing techniques. UAV technology has revolutionized the field of building damage assessment by providing high-resolution imagery of disaster-affected areas. UAVs equipped with cameras can capture images of the affected area from various angles, providing a comprehensive view of the damage. The use of UAVs in disaster management is particularly advantageous, as they can provide images in real-time, even in hazardous or inaccessible areas. In recent years, there has been an increased interest in using machine learning techniques for building damage assessment using UAV imagery. Machine learning algorithms can automate the process of analyzing the images, reducing the time and effort required for manual analysis. The use of machine learning techniques can also improve the accuracy and reliability of building damage assessments, as the algorithms can detect and classify the damage with high precision. In this paper, we present a machine learning-based approach for the multilabel classification of damaged buildings using UAV imagery and image processing techniques, to automate the building damage assessment process and improve its accuracy and reliability. The contribution of this paper is twofold. Firstly, it provides a comprehensive overview of the current state of the art in building damage assessment using UAV imagery and image processing techniques. Secondly, it demonstrates the effectiveness of a machine learning-based approach in automating the building damage assessment process, reducing the time and effort required for manual analysis. The proposed approach has the potential to be widely applied in disaster management, providing fast and reliable solutions for building damage assessment in Real-time.

II. RELATED WORKS

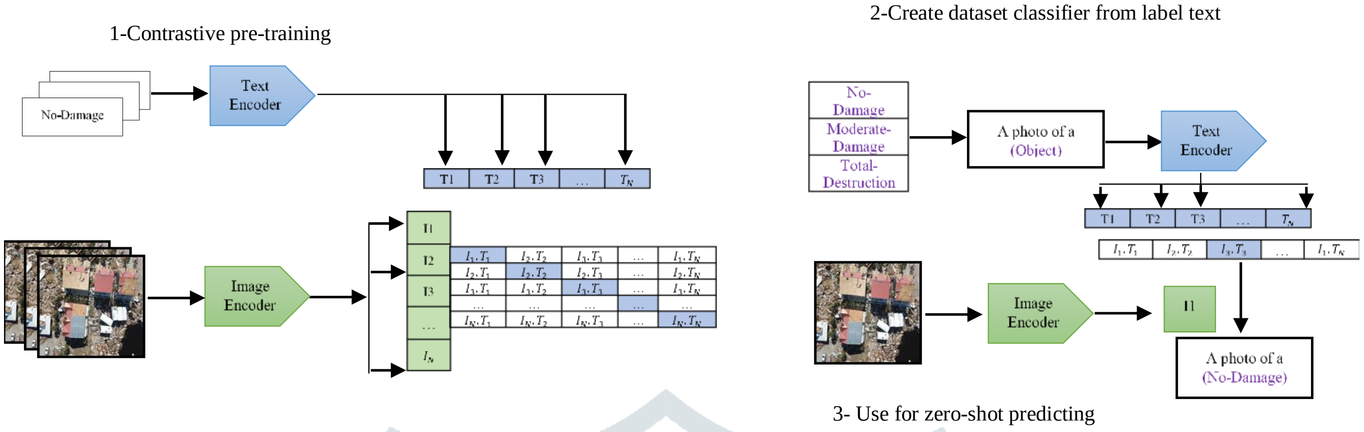
Over the past decade, the utilization of drones has undergone significant advancement, encompassing a broad spectrum of domains, including agriculture, commerce, humanitarian aid, and disaster management[1]. The versatility and efficacy of this technology have made it a widely sought-after tool, both for professionals and hobbyists alike. With the increasing popularity of drones, it's crucial to explore their full potential and implications, making further advancements in this field. The use of drones in disaster management and building damage assessment has become increasingly prevalent in recent years, due to their unique capabilities and advantages over traditional methods. Drones equipped with high-resolution cameras and other sensors can quickly and efficiently survey large areas, capturing images and data that can be used to assess the extent of damage and inform response efforts. This not only saves time but also increases the accuracy of damage assessments and reduces the risk to human life. In disaster zones, drones can also be used for search and rescue operations, providing aerial views of the affected areas and assisting in locating survivors. With the ability to quickly and safely survey disaster zones, drones have become an essential tool for disaster management and building damage assessment.

The objective of [2] was to identify the influencing factors of typhoons and assess their relative importance using the disaster theory and the Analytic Hierarchy Process (AHP). The effectiveness of this method was demonstrated through the collection and analysis of field data at a construction site. The study proposes a set of early warning protocols and disaster prevention and mitigation measures for typhoons on construction sites. The results indicate that unmanned aerial vehicles (UAVs) can play a crucial role in improving the ability of construction site management to assess typhoon risk and prevent disasters. The study provides a foundation for the examination and evaluation of the advancements in UAV and immersive technologies and their impact on construction projects. In this work[3], the authors present a data transfer algorithm for evaluating the impact of a single historical training sample on model performance. The goal of the algorithm is to select advantageous samples from historical data to aid in the calibration of a new model. The study compares and evaluates the performance of four models created using two datasets of earthquake-damaged buildings. The results demonstrate that the proposed data transfer algorithm improves the accuracy of the building damage assessment model by selecting samples from historical data that are relevant to the new task. When the new task has only 10% of the training data as compared to the historical data and involves classifying building damage into four categories, the model built using the data transfer method exhibits an 8% improvement in overall accuracy on the test set as compared to the model trained directly on the new earthquake samples. In a context where data is limited, the proposed data transfer algorithm enhances the precision of seismic building damage assessment, making it a valuable approach for evaluating building damage in future disasters. The utilization of drones in the context of natural disasters has been the subject of recent research, resulting in the identification of four principal categories: disaster management and mapping, which has demonstrated the greatest efficacy; search and rescue operations; transportation of essential supplies and equipment; and training and simulation exercises. This systematic categorization of drone applications sheds light on the potential of this technology to support and enhance disaster response efforts. The authors in [4] evaluate the damage caused by the 2016 earthquake in central Italy and the 2019 cyclone Idai in Mozambique. They assessed damage using DEEP (Digital Engine for Emergency Image Analysis), a deep learning tool designed for automatic footprint segmentation and classification of building damage. Using image-based survey techniques, such as UAV photogrammetry, as the primary method of data collection, the application is designed to generate emergency response-useable cartography rapidly. Greenwood et al. [5] The paper employed multiple data sources, including social media, direct observation, participant observation, and semi-directed interviews, to create a comprehensive picture of the post-hurricane condition of homes in Texas and Florida. By using images taken during the hurricanes, the authors were able to accurately assess the extent of damage and provide valuable insights for future disaster response efforts. Their multi-method approach enabled them to create a robust depiction of the post-hurricane conditions, showing the damage caused to homes and communities in the affected areas. An observational study was conducted using 72 stereoscopic orthographic images of a landslide, acquired using a UAV manufactured by DJI. The images were processed using Pix4Dmapper to accurately estimate the landslide and disaster information in the post-disaster assessment phase. This study demonstrates the potential of using UAV technology and image processing tools in the aftermath of natural disasters to assess the extent of damage and inform response efforts quickly and accurately. The results of the study provide valuable insights into the use of UAV technology and image processing in disaster assessment and highlight the importance of incorporating such methods into disaster response protocols. [6] Chang et al.in [7] utilized high-resolution aerial photographs of the Laishe River taken between 2009 and 2015, with additional data collected using a UAV in January and November of 2015. The photographs were analyzed using software such as Pix4Dmapper, DEMs, and ArcGIS to quantify the migration of landslide material and assess morphological changes in the mountainous river. The results of the study emphasize the feasibility of using UAVs to collect data and inform the understanding of river morphological changes, particularly in mountainous areas where traditional data collection methods may be limited. The findings have important implications for natural resource management, hazard assessment, and disaster response planning in mountainous regions. Another study aimed to develop a damage-map estimation system for disaster management that leverages unmanned aerial vehicle (UAV) images and deep learning algorithms [8]. The implementation utilized a Phantom 4 Pro V2 UAV flown at an altitude of 150 meters, incorporating deep learning-based image segmentation algorithms. The results of the study showed that the system was highly effective in identifying burnt areas from UAV images, thereby facilitating the accurate estimation of damage maps caused by forest fires. The successful implementation of this approach highlights the potential of UAV images and deep learning algorithms in providing efficient and effective damage assessments for disaster management purposes. Andreadakis et al. [9] aimed to assess the efficacy of using unmanned aerial systems (UAS) in estimating peak discharge in ephemeral streams following floods. The study utilized a DJI Phantom 4 Pro quadcopter controlled through DJI GO 4 Pro software on an Apple iPad Pro to collect data. The results of the study revealed that the UAS-aided approach was able to deliver accurate results comparable to those obtained through traditional methods, but with the added advantage of being more flexible. The findings of this study emphasize the potential of UAS as a valuable tool in post-flood discharge estimation in ephemeral streams and highlight its role in mitigating flood hazards. Nex et al. proposed a novel, low-cost method of mapping building damage using unmanned aerial vehicles (UAVs). They used a DJI Mavic Pro, which was seamlessly connected to a remote control, smartphone, and laptop via USB and Wi-Fi. This allowed for the creation of high-quality building damage maps in near-real time. Despite the lower flight efficiency, using low-cost, commercially available UAVs produced higher image quality, making it a viable option for building damage mapping. A search and rescue simulation scenario were used to validate the solution's efficacy [10].

III. PROPOSED METHOD

The current state-of-the-art computer vision systems are limited in their generality and practicality due to their training on a predetermined set of object categories. This fixed form of supervision necessitates the acquisition of additional labeled data to recognize other visual concepts. An alternative and more promising approach is learning directly from raw textual descriptions of images, which draws upon a much more extensive source of supervision. In recent years, pre-training methods that learn directly from the raw text have revolutionized NLP [11,12]. CLIP demonstrates that the simple pre-training task of predicting which caption goes with which image is a scalable and efficient method for teaching SOTA image representations from scratch on an image database of 400 million image-text pairs collected from the internet. The dataset was used to test CLIP's capabilities. After pre-training, natural language is used to refer to learned visual concepts, enabling zero-shot model transfer to subsequent tasks. CLIP's performance is tested by comparing it to over 30 distinct computer vision datasets comprising OCR, video action identification, geo-localization, and a variety of fine-grained object classification tasks. The performance of CLIP is evaluated by benchmarking [13]. As shown in Fig. 1, To predict labels, standard image models usually train an image feature extractor and a linear classifier together. However, in the case of CLIP, both an image encoder and a text encoder are trained jointly to predict the correct pairings of a batch of

IV. image and text training examples. During the testing phase, the trained text encoder generates a zero-shot linear classifier by embedding the names or descriptions of the classes in the target dataset. Benchmarking against different computer vision datasets has shown that this method is effective for getting high performance. The CLIP approach is designed to enable pre-trained models to learn a diverse set of tasks. By leveraging natural language prompting, this approach allows for the zero-shot transfer of the learned task knowledge to a wide range of existing datasets. Despite the need for sufficient scale, the performance achieved through this approach can be competitive with task-specific supervised models. However, further advancements are necessary to continue improving the efficacy of this technique.



1-Contrastive pre-training

2-Create dataset classifier from label text

3- Use for zero-shot predicting

Fig. 1. CLIP steps for building damage assessment with UAV images

IV.RESULTS AND DISCUSSION

In this section, we present the results of using Rescue-Net, a high-resolution UAV semantic segmentation benchmark dataset, for natural disaster damage assessment. We utilized Roboflow[1] for image annotations, categorizing the images into three classes: no damage, moderate damage, and destruction, as illustrated in Figure 2. Our image dataset consisted of 1412 raw images. To improve the accuracy of our method, we implemented pre-processing and augmentations. The results demonstrate the effectiveness of our approach in accurately assessing the level of damage caused by natural disasters. The use of Rescue-Net and Roboflow allowed for more efficient and accurate processing of large image datasets, making this method a promising tool for future disaster response efforts.

A. Image Pre-Processing

To improve the performance of the CLIP method on this dataset, we resized the original images from 3000 by 4000 pixels to 1280 by 1280 pixels. This was done to reduce the computational cost and increase the efficiency of the algorithm. Our experimental results show that this resizing did not significantly affect the accuracy of the CLIP method on the Rescue-Net dataset, while also reducing the training time and memory usage. Thus, we recommend using this resized version of the dataset for further research and development of disaster response systems.

B. Image Augmentation

Image augmentation is a popular technique used in the field of machine vision to improve the accuracy of computer vision models. This technique involves generating new images from existing ones by applying various transformations such as flipping, rotating, zooming, and changing the brightness and contrast levels. By generating these new images, we can significantly increase the size of the dataset, which can help to prevent overfitting and improve the generalization of the model. In addition, image augmentation can also help to improve the robustness of the model by making it more tolerant to changes in the input data. Overall, the use of image augmentation in machine vision projects can lead to better accuracy and more reliable results, making it an important tool for researchers and practitioners in this field. In this study, we employed image augmentation techniques to improve the accuracy of our machine vision project. Specifically, we applied a combination of 90° rotations, clockwise and counter-clockwise, rotations between -15° and +15°, horizontal and vertical shearing within 15°, and brightness adjustments between -25% and +25%. By applying these techniques, we generated a final dataset of 5352 images, which included variations of the original dataset to increase its diversity. Fig. 3, shows some sample images after augmentation. Our experimental results showed that the accuracy of our machine vision model improved significantly when trained on this augmented dataset, indicating that image augmentation is an important technique to enhance the performance of computer vision models.

[1] https://app.roboflow.com/

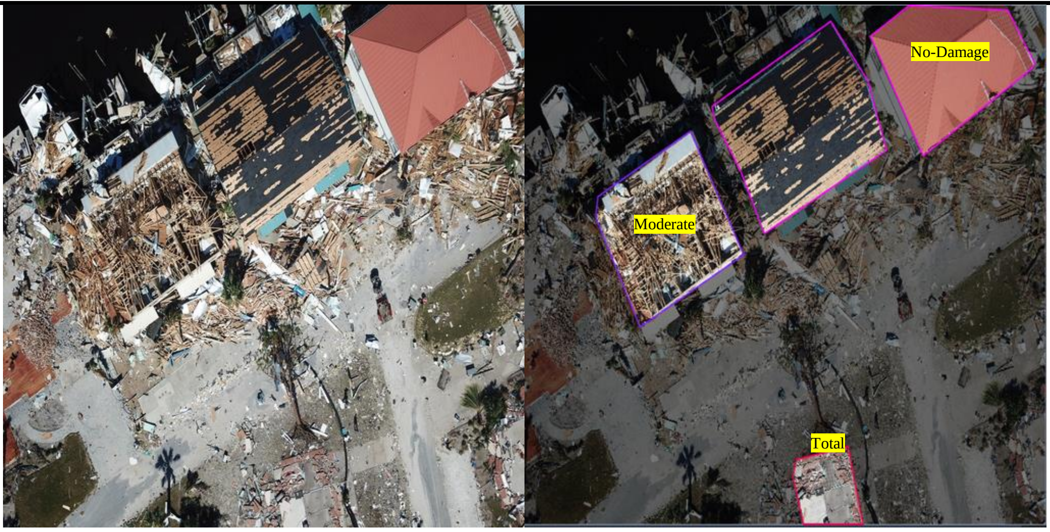


Fig. 2. Sample image and annotations

C. Training and Inferring the model.

After creating an image dataset and dividing it into training, testing, and validation sets, I exported the image dataset in CLIP format and used the Spyder platform for both training and inference. Our method was applied to the Rescue-Net dataset, resulting in an impressive accuracy of 92%. To demonstrate the effectiveness of our approach, we used Figs. 4–6 to present the results of deploying our method on some images from the Rescue-Net dataset. The images displayed in these figures show that our method successfully detects and classifies objects in various disaster scenarios, such as collapsed buildings, fires, and floods. Overall, our study shows that using the Spyder platform and CLIP format for training and inference can yield high-accuracy results in computer vision tasks, and the use of the Rescue-Net dataset can lead to more effective disaster response systems.

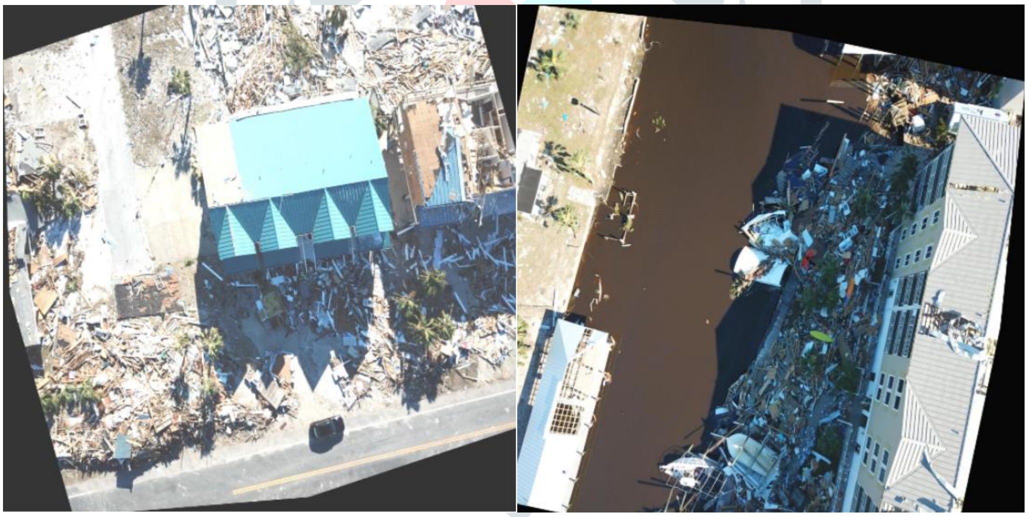


Fig 3.sample of images after augmentations

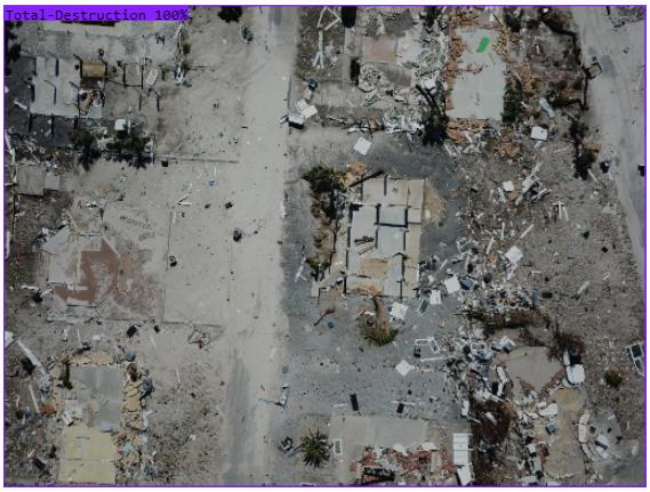


Fig.4.Total destruction detection (100%)

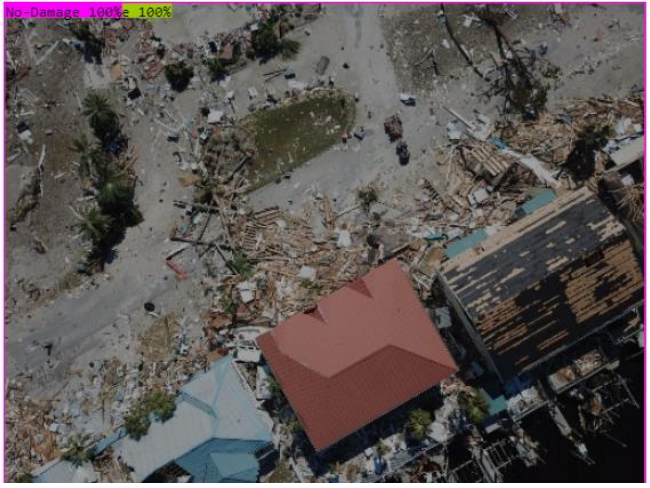


Fig.5.No damage and Moderate damage detection (100%)

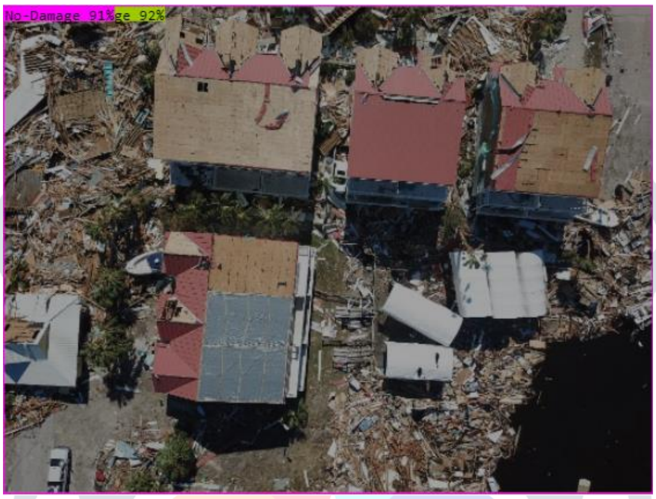


Fig.6.No damage (91%) and Moderate damage (92%) detection

Also, we evaluated our proposed method for multilabel building damage assessment using multiple datasets, including the ISBDA[2] image dataset and the ida-BD[3] image dataset. The results of our method were visually presented in Figs 7 and 8, showing good classification performance. Additionally, we obtained several images from the internet related to the recent earthquake in Turkey to further test the effectiveness of our proposed method. The segmentation results of these images were shown in Figs 9 and 10, demonstrating the robustness of our method in real-world scenarios. The experimental results suggest that our proposed method is a promising approach for accurate and efficient image classification, with potential applications in various fields, such as medical imaging, remote sensing, and computer vision.

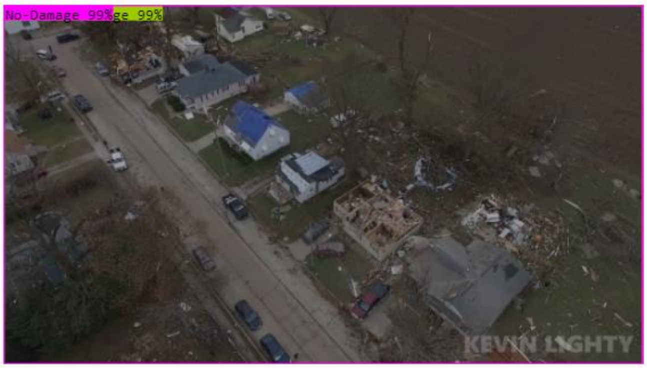


Fig.7. ISBDA dataset, No-damage (99%) and Moderate- damage (99%) detection

[2] https://github.com/zgzxy001/MSNET
[3] https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-3563

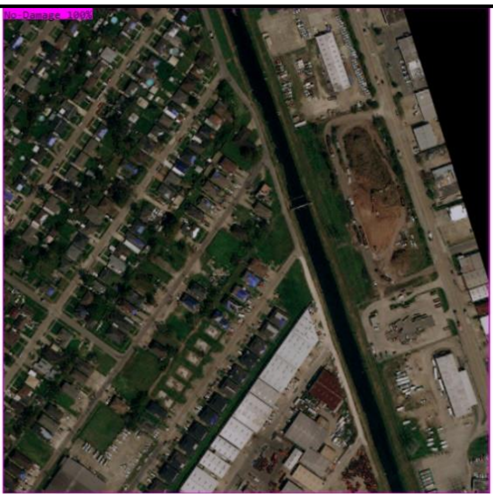


Fig.8. Ida-BD dataset, No-damage detection (100%)

To benchmark our proposed method's performance against well-known algorithms, we conducted comparative experiments on our image dataset, implementing YOLOV8, EfficientNet, and MobileNetV2 and thoroughly evaluating their outcomes. Figure 11 illustrates that although YOLOV8 achieves impressive results in terms of accuracy top 5, its accuracy top 1 rate remains limited to 66%, indicating that further optimization may be necessary to enhance its performance in this regard. Our experiments show that even with fine-tuning and re-training, MobileNetV2's highest accuracy rate remained at 29%, as demonstrated in Figures 12 and 13, indicating that this method may not be suitable for our particular image dataset. Our experimental results reveal that despite fine-tuning and training EfficientNet for approximately 500 epochs, the method's highest accuracy rate was only 25%, as shown in Figure 14&15, indicating that it may not be suitable for our image dataset. Table 1 presents a comprehensive comparison between our proposed method and other well-known algorithms, demonstrating the superiority of our approach in terms of accuracy.

V. Conclusion and Future works

In this paper, we propose a new method for multilabel building damage assessment based on the CLIP model, which was applied to an after-disaster image dataset, Rescue Net. Our proposed method used CLIP to classify the images into three classes, including No-Damage, Moderate-Damage, and Total-Destruction. The accuracy of the proposed method on Rescue Net was 92%, demonstrating the effectiveness of our approach. Additionally, we evaluated our method using other datasets and images from the internet, and the results were similarly promising, indicating the generalizability of our approach to different scenarios. Overall, the proposed method provides a valuable tool for efficient and accurate damage assessment in post-disaster situations, which can assist rescue teams and authorities in making critical decisions and prioritizing rescue efforts.

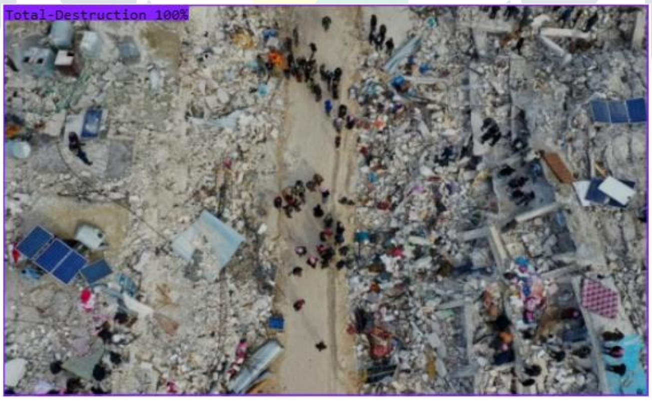


Fig.9. Turkey Earthquake (February 2023)- Total-Destruction (100%) detect.

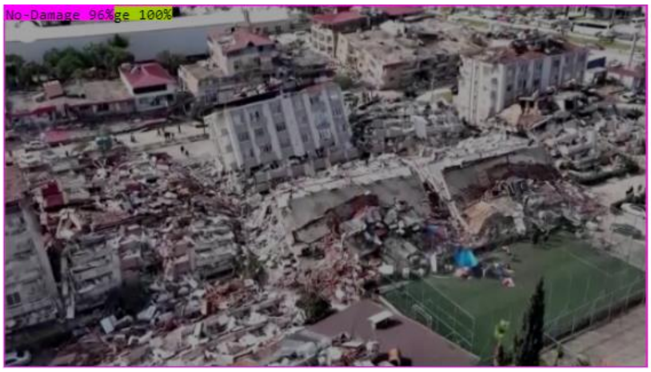


Fig.10. Turkey EarthquakeFebruaryy 2023)- No damage (96%) and Moderate- damage (100%) detection

Table .1: Comparison between CLIP and other well-known methods

Method	Accuracy (%)
YOLOV8(TOP 5)	98
YOLOV8(TOP 1)	66
EfficientNet	25
MobileNetV2	29
CLIP	**92**

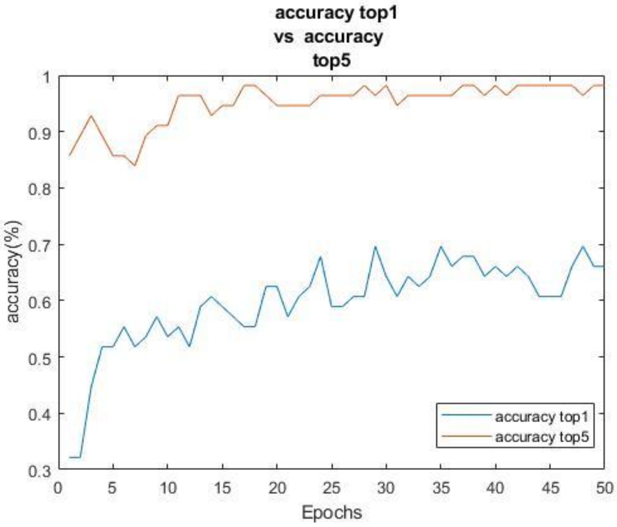


Fig.11. YOLOV8 accuracy top1 and top 5 results

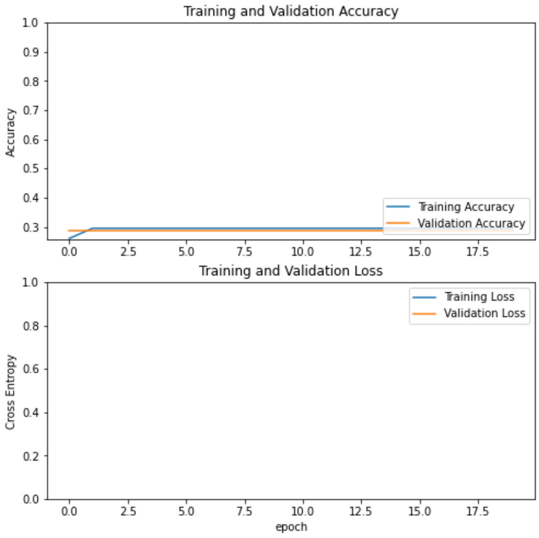


Fig.12. MobileNETV2 accuracy before finetuning

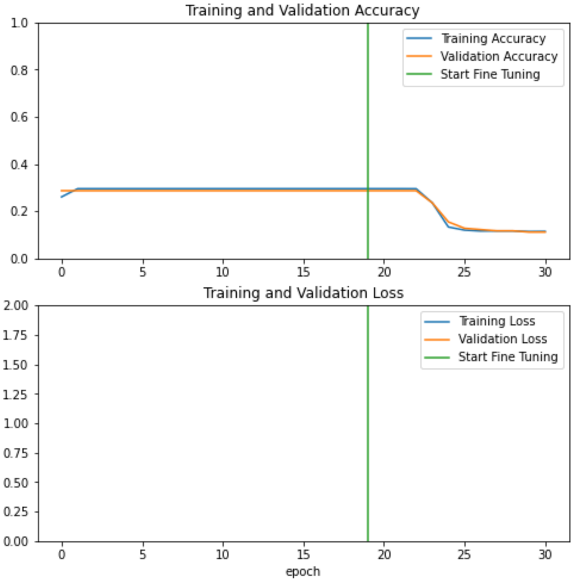


Fig.13. MobileNETV2 accuracy after finetuning

Despite the different angles and lighting of the images, the presented method performed well when tested on images other than those from the original data set. The proposed method provides a valuable tool for disaster management and response, enabling accurate and efficient damage assessment, which can facilitate decision-making and the prioritization of rescue efforts. We believe that our work can contribute to the development of more effective and comprehensive solutions for building damage assessment, ultimately helping to mitigate the impact of natural disasters on human lives and infrastructure.

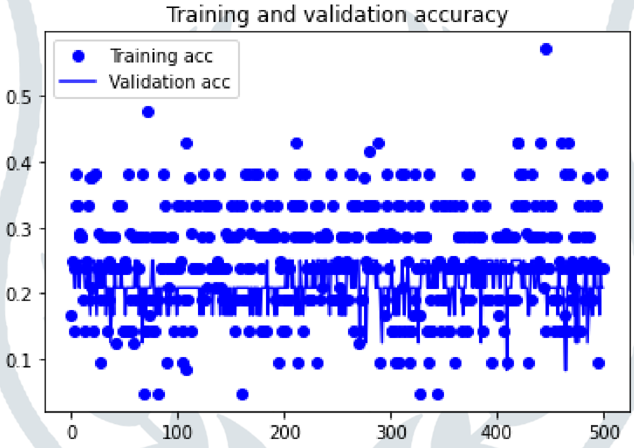


Fig.14. EfficientNet accuracy after finetuning

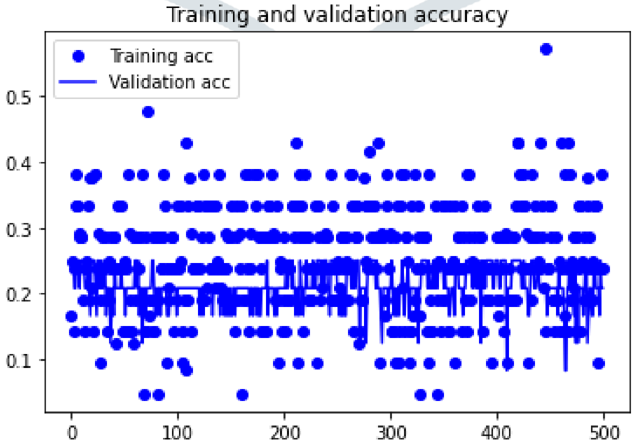


Fig.15. EfficientNet accuracy after finetuning

One promising avenue is to explore the use of additional pre-trained models, such as Vision Transformer (ViT), to improve the accuracy of the classification results. Another possibility is to investigate the impact of incorporating other sources of information, such as geographic and socioeconomic data, in the classification process to better capture the complexity of building damage assessment. Furthermore, there is a need to conduct a more comprehensive evaluation of the proposed method using larger and more diverse datasets to ensure its robustness in different scenarios. Finally, it may be valuable to develop a more user-friendly interface and integrate the proposed method into a software package to facilitate its use by non-experts in disaster management and response. it is possible to combine CLIP with object detection models such as YOLO or R-CNN for the detection and localization of damaged buildings in images. One approach could be to use the CLIP model to classify the image into damage categories (e.g., No-Damage, Moderate-Damage, Total-Destruction), and then use an object detection model to locate and extract regions of interest (ROIs) that correspond to the damaged buildings. The ROIs could then be further analyzed and quantified to provide more detailed information about the extent and severity of the damage.

References

[1] S.M.S.M. Daud, M.Y.P.M.Yusof, C.C.Heo, L.S.Khoo, M.K.C.Singh, M.S.Mahmood, and, H. Nawawi. Applications of drone in disaster management: A scoping review. *Science & Justice*, *62*(1), pp.30-42, 2022.

[2] C.Wang, Y.Tang, M.A.Kassem, and Z.Chen. UAV Application for Typhoon Damage Assessment in Construction Sites. *Applied Sciences*, *12*(13), p.6293, 2022.

[3] O.Lin, T.Ci, L.Wang, S.K.Mondal, H.Yin, and Y.Wang. Transfer Learning for Improving Seismic Building Damage Assessment. *Remote Sensing*, *14*(1), p.201, 2022.

[4] A.Calantropio, F.Chiabrando, M.Codastefano, and E. Bourke. Deep learning for automatic building damage assessment: application in post-disaster scenarios using UAV data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, pp.113-120,2021.

[5] [5]F. Greenwood, E.L. Nelson, P.G.and Greenough. Flying into the hurricane: A case study of UAV use in damage assessment during the 2017 hurricanes in Texas and Florida. *PLoS one*, *15*(2), p.e0227808, 2020.

[6] Y.Luo, W.Jiang, B.Li, O.Jiao, Y.Li, O.Li, and J.Zhang. Analyzing the formation mechanism of the Xuyong landslide, Sichuan province,China, and emergency monitoring based on multiple remote sensing platform techniques. *Geomatics, Natural Hazards, and Risk*, *11*(1), pp.654-677, 2020.

[7] K.J.Chang, C.W.Tseng, C.M.Tseng, T.C.Liao, and C.J.Yang,. Application of unmanned aerial vehicle (UAV)-acquired topography for quantifying typhoon-driven landslide volume and its potential topographic impact on rivers in mountainous catchments. *Applied Sciences*, *10*(17), p.6102, 2020.

[8] D.Q. Tran, M. Park, D. Jung, S. Park, Damage-map estimation using UAV images and deep learning algorithms for disaster management system, Remote Sensing. 12, pp. 1–17 (2020).

[9] E. Andreadakis, M. Diakakis, E. Vassilakis, G. Deligiannakis, A. Antoniadis, P. Andriopoulos, N.I. Spyrou, E.I. Nikolopoulos, Unmanned aerial systems-aided post-flood peak discharge estimation in ephemeral streams, Remote Sensing. 12, pp 1–27, (2020).

[10] F. Nex, D. Duarte, A. Steenbeek, N. Kerle, Towards real-time building damage mapping with low-cost UAV solutions, Remote Sensing. 11, 2019.

[11] C.Raffel, N.Shazeer, A.Roberts, K.Lee, S.Narang, M.Matena , Y.Zhou, , W.Li, & P. J.Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *ArXiv*. https://doi.org/10.48550/arXiv.1910.10683, 2019.

[12] J.Devlin, M.Chang, K.Lee, and K.Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*. https://doi.org/10.48550/arXiv.1810.04805, (2018).

[13] A. Radford, J.W.Kim, C.Hallacy, A.Ramesh, G.Goh, S.Agarwal, G.Sastry, A.Askell, P.Mishkin, J.Clark, and G.Krueger. Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR, 2021.