UAV-based GAN-aided Post Disaster 3D-Scene Reconstruction for Efficient Survivor Detection

Gunasekaran Raja, *Senior Member, IEEE,* Abhishek Manoharan, Nishanthini S, Vamsi Raju M

***Abstract*—Post-disaster scene understanding frameworks are becoming increasingly crucial in search and rescue operations and damage assessment initiatives. The use of Unmanned Aerial Vehicles (UAVs) provides an efficient method to complete the task of scene understanding. However, complex environments in post-disaster scenarios make it difficult for UAVs to accurately detect humans or objects. Moreover, inefficient object detection mechanisms lead to low accuracy and a long time for object detection tasks. Hence, to mitigate these issues, we propose a UAV-based scene understanding scheme involving a GAN-aided 3D reconstruction mechanism. This approach deploys a Generative Adversarial Network (GAN)-based model to denoise and remove occlusion in the images obtained from the UAVs. The framework classifies objects present in the visual scope of the UAV using a 3D reconstruction of the images obtained from the UAV, followed by semantic segmentation, resulting in pixel-level prediction and classification of entities present in the 3D model. Furthermore, an ensemble network consisting of a combination of single-stage and multi-stage detectors is to be used to improve the performance of the survivor detection model. This will help reduce the false negative rate and improve the system's overall accuracy.**

***Index Terms*—Unmanned Aerial Vehicles, Generative Adversarial Networks, Semantic Segmentation, Convolutional Neural Network, Computer Vision**

I. INTRODUCTION

Post-disaster scene understanding frameworks are becoming increasingly crucial in search and rescue operations and damage assessment initiatives. As the number of natural disasters continues to rise, the importance of efficient and accurate disaster response has become paramount. The use of unmanned aerial vehicles (UAVs) provides an efficient and cost-effective method to complete the task of scene understanding. However, the complex environments in post-disaster scenarios make it difficult for UAVs to accurately detect humans or objects. Additionally, inefficient object detection mechanisms lead to low accuracy and a long time for object detection tasks, which can be particularly problematic in urgent search and rescue situations. Survivors being small objects in post-disaster UAV images makes the task of survivor detection using traditional techniques daunting. Furthermore, survivors who are occluded in the image due to debris or damaged buildings covering them will make the survivor detection task challenging.

To mitigate these issues, we propose a UAV-based scene understanding scheme involving a GAN-aided semantic segmentation mechanism. This approach classifies objects present in the visual scope of the UAV using a 3D reconstruction from thermal images of the scene and pixel-level prediction. By leveraging the power of GANs, our method can better handle the challenges of post-disaster environments and improve the accuracy of object detection. A GAN-based denoiser results in images having lower occlusion and optimal brightness, thereby highlighting the important features of the object. Furthermore, using GAN improves the detection of small and dense objects, which is the case of survivors in images obtained from a UAV. The proposed system implements a GAN-infused SFM-enabled 3D reconstruction mechanism to detect survivors present in a post-disaster scene. A 3D reconstruction of the scene using the enhanced images obtained from the GAN-based model will be used to map and extract useful information. To overcome the challenge of detecting occluded survivors, we propose a GAN-based generative SFM framework, wherein occluded survivors will be visually generated before the point cloud is triangulated. The Bundle Adjustment technique is to be deployed to estimate the UAV camera poses with minimal error or drift while generating the 3D model. Semantic segmentation on the 3D model leads to a pixel-level prediction of various entities or objects present in the image, thereby generating a corresponding color-coding to each entity. The ensemble model, a hybrid architecture consisting of single-stage and multi-stage detectors, is to be implemented to detect the presence of survivors. The network will overcome the disadvantages of both frameworks. Deploying an ensemble network comprising the CenterNet and Cascade R-CNN frameworks improves the performance of survivor detection while decreasing the false negative rate. The overall framework will increase the accuracy and efficiency of the survivor detection task, thereby resulting in successful SAR operations. With these improvements, our approach has the potential to significantly enhance the effectiveness of post-disaster scene understanding frameworks and aid in the critical tasks of search and rescue and damage assessment initiatives.

The key contributions of this paper include:

1) A GAN-based denoiser and occlusion remover mechanism will be implemented to improve the detection of survivors using post-disaster UAV images.
2) A 3D scene-reconstruction mechanism based on the SfM and Bundle Adjustment algorithms will be executed to map and extract the most useful information present in the scene
3) A semantic segmentation mechanism will be deployed on the 3D model to classify various entities in the scene, thereby improving survivor detection.
4) A hybrid ensemble network comprising single-stage and multi-stage detectors will be developed for survivor detection using the color coding and classification of semantic entities. This will result in the decrease of the high false negative rate of the multi-stage mechanism and

the improvement in the performance of the single-stage detector.

II. Related Works

The authors of [1] propose a new thermal image dataset consisting of 6447 thermal images designed for survivor detection using UAVs in post-disaster scenarios. The paper also describes optimal values to prune survivor detection models in order to reduce the complexity of the models. The model applies knowledge distillation techniques to fine-tune them and improve accuracy. The performance of several survivor detection models based on YOLOv3 and YOLOv3-MobileNetV1 were compared with and without pruning and fine-tuning. However, Older and inferior detection models have been used for survivor detection, thereby resulting in models with high mean average precision (mAP) loss and low accuracy. Furthermore, [2] implements a 3D imaging mechanism for 2D images obtained from a swarm of UAVs. The proposed work involves the 3D imaging of a scene by the usage of 2D images obtained from several UAVs present in the UAV Swarm at different perspectives with a few points of overlap. The point cloud obtained is then triangulated, and Bundle Adjustment is used to create the 3D rendering of the image. But a considerable amount of data must be transmitted from the UAV swarm, as images obtained from each node in the swarm are used to produce the 3D rendering. Multiple UAVs also need to exchange information in order to efficiently collect data on the scenario.

In order to create greater fidelity terrain models, [3] describes a bundle adjustment technique for aerial texel images. The model enables relatively low-accuracy navigation systems to be employed with inexpensive LiDAR and camera data. On the contrary, outliers in the point cloud are not identified and mitigated, leading to lower accuracy. Furthermore, With the goal of lowering the high false negative rate of multi-stage detectors and improving the quality of the single-stage detector proposals, the authors of [4] propose an ensemble network called SyNet that combines a multi-stage method with a single-stage one. But according to the investigation, detecting objects in drone images is more challenging than detecting them in images that were taken from the ground, even with the most advanced object detection algorithms. Hence, the accuracy of the model trained on UAV images is still low compared to models trained on ground images.

[5] provides a review of vehicle detection from UAV imagery using deep learning techniques. It begins by outlining the various deep learning architectures, including generative adversarial networks, autoencoders, recurrent neural networks, and convolutional neural networks and their contributions to the challenge of improving vehicle detection. The paper then focuses on examining various vehicle detection techniques and presents different benchmark datasets and problems that have been discovered, along with possible remedies. However, videos captured in the UAVs are sent to on-ground workstations or to the cloud for processing rather than being implemented on the UAV itself, thereby leading to the absence of a lightweight system for vehicle detection Furthermore, [6] proposes a global-local feature-enhanced network (GLF-Net) to alleviate issues when detecting small and dense objects using UAVs. A feature-fusion module has been proposed to tackle the presence of numerous small objects. GLF-Net achieves 86.52% mean Average Precision (mAP) on the RO-UAV dataset. The scalability of the framework however is poor, and the application of GLF-Net on post-disaster UAV images leads to lower mAP, thereby requiring better frameworks.

The authors of [7] execute and compare various UAV detection mechanisms using air-borne UAVs that deploy deep neural networks. 4 datasets have been used and performance has been compared namely MAV-VID, Drone-vs-Bird, Anti-UAV RGB, and Anti-UAV IR. The performance of 4 models was compared using the datasets mentioned, namely Faster RCNN, SSD512, YOLOv3, and DETR (Detection Transformer). Overall, Faster RCNN performed best. But long-distance detection of small UAVs was not taken into consideration. Deep neural networks for the re-identification of UAVs were not considered as well. Furthermore, [8] introduces a high-resolution post-disaster UAV dataset named RescueNet, which contains comprehensive pixel-level annotation of 11 classes for semantic segmentation to assess damage after a natural disaster. The dataset collection and annotation process are discussed, along with the challenges it poses. However, RescueNet contains a small number of classes. As a result, smaller objects like "vehicles" and "pools" make it difficult to get a good segmentation compared to larger objects like buildings and roads. Besides that, since UAV images include only the top view of a scene, it is difficult to assess the actual damage since the horizontal view also brings information regarding all sides of a building.

[9] proposes a UAV-Human dataset for understanding human action, pose, and behavior. The proposed UAV-Human contains 67,428 multi-modal video sequences, 119 subjects for action recognition, 22,476 frames for pose estimation, 41,290 frames, 1,144 identities for person re-identification, and 22,263 frames for attribute recognition which encourages the exploration and deployment of various data-intensive learning models for UAV-based human behavior understanding. However, The UAV-Human dataset poses a limitation for attribute recognition because the dataset is captured over a relatively long period of time. As a result, the subjects have been diversified with different dressing types and large variations of viewpoints caused by multiple UAV altitudes. Furthermore, the authors of [10] propose and evaluate a novel self-attention segmentation model named ReDNet on a new high-resolution UAV natural disaster dataset named HRUD. The challenges of semantic segmentation on the HRUD dataset are discussed, along with the excellent performance of the proposed model. On the contrary, HRUD is a very challenging dataset due to its variable-sized classes along with similar textures among different classes. Debris, textures of debris, sand, and building with destruction damage make a great impact on the segmentation performance of the evaluated network models.

Hence, to mitigate the aforementioned limitations of currently existing systems for survivor detection, we propose an efficient post-disaster scene understanding framework using UAVs for survivor detection and SAR operations that will encompass a GAN-based denoiser and occlusion remover mechanism to improve the detection of survivors using post-disaster UAV images, a 3D scene-reconstruction mechanism based on the SfM and Bundle Adjustment algorithms to map and extract the most useful information present in the scene, a semantic segmentation mechanism on the 3D model to classify various entities in the scene to improve survivor detection, and a hybrid ensemble network comprising single-stage and multi-stage detectors for survivor detection using the color coding and classification of semantic entities. This will result in the decrease of the high false negative rate of the multi-stage mechanism and the improvement in the performance of the single-stage detector, which in turn leads to an efficient survivor detection model for Search and Rescue operations.

III. SYSTEM MODEL

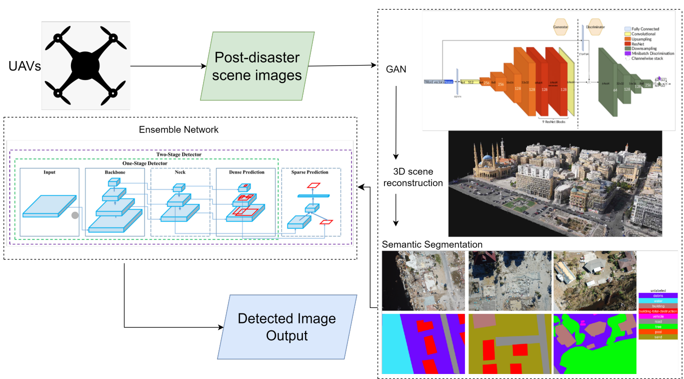

Fig. 1: High-Level Architecture

Algorithm 1 Image Denoising and Entity Separation

Input: Images (λ) of post-disaster scene obtained from UAV
Output: Post-disaster scene entities (Φ)

procedure GAN_DENOISING(λ)
  Pre-processing of UAV images λ
  $\lambda_{\text{epochs}} \leftarrow$ Number of iterations to train GAN
  $\lambda_{\text{batchsize}} \leftarrow$ Number of images to train per epoch
  for β `in` λ **do**
    $\vartheta \leftarrow TrainGAN(\beta, \lambda_{\text{epochs}}, \lambda_{\text{batchsize}})$
  end for
  $\varpi \leftarrow GAN\_denoiser(\vartheta)$
  return ϖ
end procedure
procedure SEMANTIC_SEGMENTATION(ϖ)
  $\Phi[] \leftarrow$ entities present in ϖ
  $\nu \leftarrow 0$
  for α `in` ϖ **do**
    $\Phi[\nu] = Classify(\alpha)$
    $\nu = \nu + 1$
  end for
  return Φ
end procedure

Algorithm 2 Survivor Detection Ensemble Network

Input: Enhanced Images containing entities(ϕ) of post-disaster scene obtained from UAV
Output: Minimized Loss of Ensemble Network (L_N)

procedure ENSEMBLE_NETWORK(Φ)
  $L_C \leftarrow$ Loss of Cascade RCNN
  $L_N \leftarrow$ Loss of CenterNet
  $y_n \leftarrow$ actual label of entity x_n
  $y \leftarrow$ predicted label of entity x_n
  $p(\psi) \leftarrow$ probability of occurrence of ψ
  $t_c \leftarrow$ true width of entity x_n
  for x_n `in` Φ **do**
    $L_C \leftarrow -log(p(y = y_n|x_n))$
    Applying Ensemble Loss Minimization:
    $L_N(L_C) \leftarrow -log(|y - y_n t_c|)$
  end for
  return L_N
end procedure

IV. PROPOSED WORK

V. RESULTS AND DISCUSSIONS

VI. CONCLUSION

REFERENCES

[1] J. Dong, K. Ota and M. Dong, "UAV-Based Real-Time Survivor Detection System in Post-Disaster Search and Rescue Operations" in IEEE Journal on Miniaturization for Air and Space Systems, vol. 2, no. 4, pp. 209-219, 2021

[2] H. Ren et al., "Swarm UAV SAR for 3-D Imaging: System Analysis and Sensing Matrix Design" in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022

[3] T. C. Bybee and S. E. Budge, "Method for 3-D Scene Reconstruction Using Fused LiDAR and Imagery From a Texel Camera" in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 8879-8889, Nov. 2019

[4] Albaba, Berat Mert, and Sedat Ozer, "SyNet: An ensemble network for object detection in UAV images" in 25th IEEE International Conference on Pattern Recognition (ICPR), pp. 10227-10234, 2021

[5] A. Bouguettaya, H. Zarzour, A. Kechida and A. M. Taberkit, "Vehicle Detection From UAV Imagery With Deep Learning: A Review" in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 11, pp. 6047-6067, Nov. 2022

[6] T. Ye, W. Qin, Y. Li, S. Wang, J. Zhang and Z. Zhao, "Dense and Small Object Detection in UAV-Vision Based on a Global-Local Feature Enhanced Network," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-13, 2022

[7] Isaac-Medina, Brian KS, Matt Poyser, Daniel Organisciak, Chris G. Willcocks, Toby P. Breckon, and Hubert PH Shum. "Unmanned aerial vehicle visual detection and tracking using deep neural networks: A performance benchmark" In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1223-1232, 2021.

[8] Rahnemoonfar, Maryam, Tashnim Chowdhury, and Robin Murphy. "RescueNet: A High-Resolution Post Disaster UAV Dataset for Semantic Segmentation." UMBC Student Collection, 2021

[9] Li, Tianjiao, Jun Liu, Wei Zhang, Yun Ni, Wenqian Wang, and Zhiheng Li. "Uav-human: A large benchmark for human behavior understanding with unmanned aerial vehicles" In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16266-16275, 2021

[10] T. Chowdhury and M. Rahnemoonfar, "Attention Based Semantic Segmentation on UAV Dataset for Natural Disaster Damage Assessment" 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 2325-2328, 2021