Drones4Good: Supporting Disaster Relief Through Remote Sensing and AI

Nina Merkle[1], Reza Bahmanyar[1], Corentin Henry[1], Seyed Majid Azimi[1], Xiangtian Yuan[1], Simon Schopferer[2], Veronika Gstaiger[1], Stefan Auer[1], Anne Schneibel[3], Marc Wieland[3], and Thomas Kraft[4]

German Aerospace Center (DLR)
[1]Remote Sensing Technology Institute, [2]Institute of Flight Systems
[3]Remote Sensing Data Center, [4]Institute of Optical Sensor Systems
contact: firstname.lastname@dlr.de

Abstract

In order to respond effectively in the aftermath of a disaster, emergency services and relief organizations rely on timely and accurate information about the affected areas. Remote sensing has the potential to significantly reduce the time and effort required to collect such information by enabling a rapid survey of large areas. To achieve this, the main challenge is the automatic extraction of relevant information from remotely sensed data. In this work, we show how the combination of drone-based data with deep learning methods enables automated and large-scale situation assessment. In addition, we demonstrate the integration of onboard image processing techniques for the deployment of autonomous drone-based aid delivery. The results show the feasibility of a rapid and large-scale image analysis in the field, and that onboard image processing can increase the safety of drone-based aid deliveries.

1. Introduction

Every year, millions of people around the world are affected by natural and man-made disasters [5]. In order to respond effectively to such crises, emergency services and relief organizations rely on timely, comprehensive, and accurate information about the disaster's extent. For years, emergency mapping has been based on remote sensing data to support rescue operations, gathering information on affected areas by comparing images acquired before and after the event by satellites, aircrafts, or drones [9]. However, the automatic extraction of such information and its rapid, scalable, and reliable delivery is still a challenge. Recent developments in computer vision and the rapid evolution of graphics processing units have led to optimized, fast-running algorithms, opening up new possibilities in disaster and humanitarian relief [2].

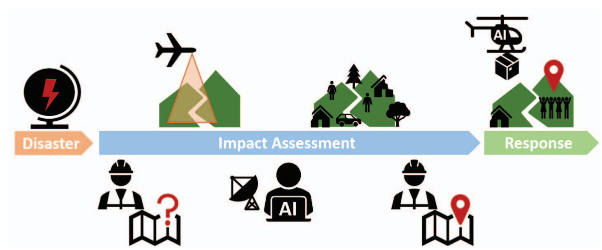

Figure 1: Illustration of the presented workflow.

In this paper, we explore the potential of state-of-the-art deep learning techniques for image analysis in combination with remote sensing data acquired by drones. The workflow, ranging from real-time and large-scale image acquisition and mapping, to automatic and fast image analysis, over to an onboard image processing method to support automatic aid delivery by drones is presented in Figure 1. The extracted information can be seamlessly integrated into an organization's operations control center to effectively support emergency services and aid organizations in their operations. The methodological focus is on: (1) automated, near real-time extraction of roads, buildings and people for initial impact assessment to help prepare missions in terms of logistics and routing of relief forces and supplies, and (2) real-time detection of people from drones to increase the safety of drone-based aid delivery once target areas have been identified. Here we use the term "near real-time" to refer to processing within hours, and "real-time" to refer to processing within seconds. The performance of all methods has been thoroughly tested and evaluated, demonstrating their potential to rapidly analyze large amounts of data and increase the safety of drone-based aid deliveries.

Ethically, there is no recognition of individuals and no sensitive data stored in a cloud where privacy could be compromised. Our work has been carried out in close collaboration with organizations such as the World Food Programme (WFP), I.S.A.R. Germany, and the Bavarian Red Cross (BRK), who recognize the high potential of our methods and the urgent need to put them into practice.

2. Situation Assessment

In order to support relief efforts in rapidly assessing the impact of a disaster, we present an approach consisting of two-steps: 1) mapping the scene in real time with a camera system mounted on a drone in order to provide up-to-date image data over the area of interest, and 2) automatically extracting relevant information in the field and in near-real time to help humanitarian organizations assess the acquired data faster. For the mapping of the disaster area, the rapid mapping camera system MACS-Micro [6] is used, which is carried and integrated into a fast-flying drone. The image data are available on the ground in real time using a commercial radio link. If the range of the radio link is exceeded, the data will be available immediately after landing. The system consists of nadir-pointing cameras, a GNSS receiver combined with an industrial-grade inertial measurement unit, an embedded computing unit, and a radio link. During a typical campaign, the camera is operated at an altitude of 200m above ground, with a speed of 80km/h, and a frame rate of 2Hz. This gives an acquisition rate of around 3200m^2 per second with a GSD of 3cm. The resulting product is a scaled image mosaic showing the current situation of the disaster area, which can be used as an additional map layer for common geographic information systems.

According to feedback from relief organizations involved, a number of important questions arise in the first moments after a disaster, such as: which areas are most affected? How many people have been affected and where are they now? Which infrastructure can still be used for rescue and relief? Based on these questions, we identified the objects ”roads, ”buildings” and ”people” as the most important in helping emergency responders to give timely answers to these questions. To this end, we developed three algorithms running on a GPU laptop in the field.

Road segmentation: We use a Dense-U-Net-121 [7] based on the widely-used U-Net [13]. For the backbone of both the encoder and decoder, we use a DenseNet-121 as it offers the best compromise between the accuracy of the result and the computational resources required. The resulting road mask provides rescue teams with an up-to-date map of the captured area and can be used to identify cut-off regions. In addition, changes to the road network and severely affected areas can be quickly identified when compared to a pre-disaster scene.

Building segmentation: For the segmenting of buildings, the HRNet [16] consisting of four parallel, multi-resolution streams that maintain fine-grained features throughout the network is used. This feature allows for more precise localization, which is crucial for the task of building segmentation. The resulting building mask can be used to identify populated areas and, if compared to a pre-disaster scene, to estimate the number of people affected and damaged houses.

Figure 2: Vertical take-off and landing drone (left) with integrated camera system MACS-Micro (right).

Person detection: For the detection of people, we use an adapted YOLOv3 [12] object detection method that addresses challenges such as variations in scenes, poses, scales, and viewing angles posed by images during real humanitarian missions. As no publicly available person detection dataset was suitable for our case, we created a new dataset consisting of aerial and drone images covering different scenarios and countries (for more details see [1]). The output of the model is the location of each detected person in the form of bounding boxes, which is extremely valuable information for search and rescue missions or for the safe delivery of supplies to affected areas.

After the inference phase, the models’ predictions are assigned the same geo-referenced coordinate space as the input image. The output layers are saved as a GeoTIFF file and overlaid on the geo-referenced input image for further analysis in any software supporting this format. The acquired image data as well as the derived information layers can either be shared directly with interested humanitarian organizations or delivered to institutions such as the Center for Satellite Based Crisis Information ZKI [19], where earth observation data are analyzed and situational awareness is generated before, during and after a natural or man-made disaster in form of ISO standardized products.

3. Delivery of relief supplies

After assessing the impact of a disaster, the next phase is to response. We focus on a specific case where we assume that people are cut off and can only be reached by air. Relief supplies are delivered by a drone dropping a payload. To increase the safety of people on the ground during the process, we investigated how a camera system can be combined with an AI algorithm to detect people onboard the drone. Generally, various drone configurations could be envisioned to deliver goods in humanitarian aid scenarios. In this work, we focused on a superARTIS demonstrator platform which is equipped with a box drop mechanism as depicted in Figure 3. This box-dropping payload was integrated and flight tested in cooperation with Wings For Aid [3]. For assess-

Figure 3: Unmanned helicopter superARTIS equipped with payload for aerial delivery.

ing the safety of the drop zone, the drone is also equipped with a downward looking camera and an onboard processing unit. The capability of running AI-based people detection onboard in real time may serve different purposes depending on the automation desired for the operation and the availability of a data link. In case the aircraft is remotely piloted via a low bandwidth data link, the processing results can still be transmitted with a low bandwidth demand as opposed to transmitting the video stream. This enables the remote pilot to assess whether or not it is safe to release the box of supplies. Also, no personal data is transmitted or recorded in this scenario. When no data link is available, the onboard autopilot may use the onboard person detection to conduct the delivery autonomously in a safe manner.

In order to enable real-time person detection on drones, we optimize the processing of the YOLOv3 discussed in section 2. We reduce the data type precision to float16 and simplify the non-maximum suppression procedure which is one of the most computationally intensive steps. The images acquired by the camera system are transferred to and processed by the onboard GPU. The detection results are then passed to the CPU and geolocated. We use the cross-platform data format Protocol Buffers [15] to ensure efficient data transfer rates and communication within the system in order to minimize the overall processing time.

4. Experiments

Training parameters: For **road segmentation**, a Dense-U-Net-121 was trained on the DeepGlobe18 [4] dataset consisting of 1,632km^2 annotated images data at 50cm/px from southeast Asian regions. We trained it for 40 epochs with a patch size of 512×512px and a batch size of 12. For **building segmentation**, we used the Inria dataset [11] to train the HRNet. This dataset contains 405 km^2 of labeled image data at 20cm/px from the USA and Austria. The training was performed for 20 epochs with a patch size of 512×512 and a batch size of 16. For training the **person detection** network, we used our own training dataset [1] consisting of 10,050 annotated persons in 311 aerial and drone images (train: 259, validation: 25, test: 27), with GSDs ranging from 0.2 to 6cm/pixel, covering areas in Germany, the Netherlands, Switzerland, Spain, France, and Nepal.

Figure 4: View from the delivery drone and demonstration of the onboard person detection algorithm.

Test hardware: To use our models in humanitarian contexts, they must run on portable and affordable computers. Therefore, we chose an Alienware Area51m laptop with 32GB of RAM and an NVIDIA RTX 2080 Super with 8GB of VRAM to process the image patches. The processing time of the three models is summarized in Table 1. For the onboard processing, we used a Jetson AGX Xavier with 8GB of VRAM. Here we tiled each image into patches of 416×416px with 10% overlap to fit into the GPU memory.

Results & Discussion: When applying research methods to real-life applications, the overall framework and all its requirements must be carefully considered and taken into account during the development and implementation of algorithms. For our specific application case, some of the constraints are due to technical, computational, legal, and financial limitations, while others are due to the overall situation during a disaster. But also the algorithms used come with limitations. For our models to generalize well to new locations, the images must be as similar as possible to those of our training set. Ideally, they should be taken at nadir, i.e. not be side-looking, in clear weather, and with sufficient illumination. The resolution of the images acquired after the disaster might be higher than during training, in which case the images are downsampled if necessary.

For the road segmentation task, we selected 20 test scenes to evaluate the generalization capability of our model, each annotated by hand with vector lines: 1 from Epeisses, Switzerland in a disaster training area (0.2km^2), 10 from the Ahr Valley, Germany after the major flooding event in 2021 (10.5km^2), and 9 from Beira, Mozambique after the Cyclone Idai in 2019 (3.5km^2). While our model was already shown to perform well in many regions around the world [7], e.g. in Nepal (see appendix), it also achieved excellent results in the chosen test scenes as in Figure 5. Most roads were successfully detected with a completeness of 71%, and few false positives with a correctness of 76% (metrics from [17], cf. Table 1). The predicted roads are regular and continuous despite changes in color and material. Some sections were incorrectly detected for three rea-

Figure 5: Qualitative results: image mosaic, segmented roads, segmented buildings, and detected people (left to right).

Table 1: Processing speed (per megapixel and per area) and quantitative results obtained on an Alienware Area51m laptop.

Task	**Training data**	**GSD** *[cm]*	**Computational time** *per MP*	*per km^2*	**Complet.** *[%]*	**Correct.** *[%]*	**Quality** *[%]*	**Prec.** *[%]*	**Recall** *[%]*	**IoU** *[%]*	**AP** *[%]*
Road	DeepGlobe18 [4]	50	0.80s	3.30s	70.96	76.48	58.08	-	-	-	-
Building	Inria [11]	20	0.38s	9.50s	-	-	-	83.74	77.70	68.12	-
People	Ours [1]	3	0.44s	19min	-	-	-	54.13	65.87	-	60.36

sons: 1) the model was trained to detect roads, but not larger asphalted areas such as parking areas or certain dead ends, 2) after disasters, sand, mud, and debris may be present on the road, which our training dataset does not feature, and 3) drone image mosaics may contain irregular regions of background along their borders, which removes the necessary context to correctly identify short sections of roads. Despite these obstacles, our model achieved 58% quality.

For the building segmentation task, we used the same 11 scenes as for the road segmentation from Epeisses and the Ahr Valley, and 3 separate scenes from Beira (2.8km^2) for evaluation with manually annotated ground truth. The average precision, recall, and intersection over union (IoU) scores are reported in Table 1). In contrast to the Ahr Vally scene, the numbers for Epeisses and Beira are much lower. There are three main reasons why: 1) the data has a much higher resolution than the training data and therefore has different spectral and textural features, 2) in Beira, a large number of buildings are very small, which is not reflected in the training data, and 3) in Epeisses, the scene includes large tents that are mistakenly segmented as buildings.

For the people detection task, we evaluate the trained model on a test set of 27 images with 410 annotations. One of the challenges we faced was the altitude required for the delivery drone. In order to safely drop the supplies, the flight altitude must be around 80m, resulting in images with a ground sampling distance of 1-3cm. Therefore, the training set for the person detection algorithm had to be adjusted to include more images in this GSD range. In addition, images with a GSD of less than 6cm had to be removed from the training set, as the visual appearance of people varies too much between 1-10cm. Overall, we achieved a precision, recall, and average precision (AP) of around 54%, 66%, and 60%, respectively (see Table 1). Images with a lower GSD generally give better predictions, although the model struggles with complex backgrounds such as vegetation or disaster ruins, and is affected by changes in camera angle. Increasing the variety of the training data could help to overcome these limitations. Testing the optimized model on the Jetson board achieved a 2s processing time for each 16MP image without compromising accuracy and recall. Figure 4 shows a sample result with a GSD of 3cm/pixel.

5. Conclusion & Future Work

Our results show that the combination of computer vision and remote sensing technologies has great potential to significantly improve disaster management and humanitarian aid. Fast and large-scale image analysis becomes feasible, and onboard image processing can increase the safety of drone-based aid deliveries. However, to improve their generalization and performance, these methods need to be trained on larger datasets from around the world, tested in the field, and extended to include other relevant features such as road or building damage. End-user feedback and knowledge will play an important role in the future development and improvement of the technologies. On the other hand, limitations in the algorithms should be overcome to make the models more robust to changes in the images (e.g. viewing angle) and trainable with less labeled data.

References

[1] Reza Bahmanyar and Nina Merkle. Saving lives from above: Person detection in disaster response using deep neural networks. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Geospatial Week 2023, Cairo, Egypt, 02-07 Sep*, 2023.

[2] Ana Beduschi. Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks. *International Review of the Red Cross*, 2022.

[3] Johann C. Dauer. *Unmanned Aircraft for Transportation in Low-Level Altitudes: A Systems Perspective on Design and Operation*, pages 523–550. Springer International Publishing, Cham, 2022.

[4] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 172–17209, 2018.

[5] Center for Research Epidemiology of Disasters (CRED). 2022 disasters in numbers, 2023.

[6] Daniel Hein, Thomas Kraft, Jörg Brauchle, and Ralf Berger. Integrated uav-based real-time mapping for security applications. *ISPRS International Journal of Geo-Information*, 8(5), 2019.

[7] Corentin Henry, Friedrich Fraundorfer, and Eleonora Vig. Aerial road segmentation in the presence of topological label noise. In *Proceeding of the International Conference on Pattern Recognition (ICPR)*, 2021.

[8] F. Kurz, D. Rosenbaum, O. Meynberg, G. Mattyus, and P. Reinartz. Performance of a real-time sensor and processing system on a helicopter. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 189–193, Nov. 2014.

[9] Stefan Lang, Petra Füreder, Barbara Riedler, Lorenz Wendt, Andreas Braun, Dirk Tiede, Elisabeth Schoepfer, Peter Zeil, Kristin Spröhnle, Kerstin Kulessa, Edith Rogenhofer, Magdalena Bäuerl, Alexander Oeze, Gina Schwendemann, and Volker Hochschild. Earth observation tools and services to increase the effectiveness of humanitarian assistance. *European Journal of Remote Sensing*, 53:67–85, 10 2019.

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision (ECCV)*, pages 740–755, 2014.

[11] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *Proceedings og the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017.

[12] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2015.

[14] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016.

[15] Kenton Varda. Protocol buffers: Google's data interchange format. Technical report, Google, 2008.

[16] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[17] Christian Wiedemann, Christian Heipke, Helmut Mayer, and Olivier Jamet. *Empirical Evaluation Of Automatically Extracted Road Axes*. 1998.

[18] Zhaohui Zheng, Ping Wang, Dongwei Ren, Wei Liu, Rongguang Ye, Qinghua Hu, and Wangmeng Zuo. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Transactions on Cybernetics*, 52(8):8574–8586, 2022.

[19] ZKI. Center for Satellite Based Crisis Information (ZKI): `https://zki.dlr.de`, 2023.